University of Amsterdam

AAAI 2020

Modelling Form-Meaning Systematicity with Linguistic and Visual Features

Arie Soeteman; Dario Gutierrez; Elia Bruni; Ekaterina Shutova

February 9, 2020

Linguistic Arbitrariness

Hockett and Hockett (1960)

“The word ‘salt’ is not salty nor granular; ‘dog’ is not canine; ‘whale’ is a small word for a large object; ‘microorganism’ is the reverse.”

Gasser (2004)

”Arbitrariness becomes necessary as the number of words increases... there are more ways to avoid ambiguity in an arbitrary language”

Non-Arbitrariness: Form-meaning Systematicity

- Phonaesthemes (Otis and Sagi, 2008; Hutchins, 1999)

GL-	VISION / LIGHT	*glimmer, glisten, glitter, gleam, glow, glare, glint*
SN-	MOUTH / NOSE	*snore, snack, snout, snort, sniff, sneeze*

- Recent research suggests systematic structures (Shillcock et al., 2001; Monaghan et al., 2014), but:

  1. How strong is this systematicity?

  2. To what extent is this systematicity clustered?

Why we should use visual information

- Past research into form-meaning systematicity used textual corpora to model meaning
- Semantic concepts have a visual component in human language processing (Zwaan and Madden, 2005)
- Semantic models can be improved by incorporating visual information (Bruni et al., 2014; Kiela and Bottou, 2014)

Alice went to see the apple **tree** in the garden

Our task:

Gain insight into the structure of form-meaning systematicity, using multimodal semantic models

1. Our methods for investigating systematicity and constructing semantic models
2. Experimental setup
3. Results
4. Conclusion and Future research

Kernel Regression

- Predictor variable: Word
- Target variable: Semantic representation

Tree

Measuring String-Distance

- Levenshtein edit-distance
- Optimize weighting for differential semantic relevance (Gutiérrez et al., 2016)

$$d(s_i, s_j) = \sum_{s=1}^{S}(W_s * V_{ijs}) = W^T V_{ij} \tag{1}$$

Text-based Semantic Model

Skip-gram with Negative Sampling (Mikolov et al., 2013)

- Learn word vectors from contextual use in a large text corpus
- Maximize corpus-probability by optimizing semantic vectors

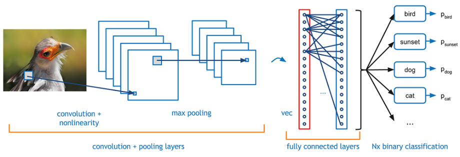
Climate Change: What is being done around the world to plant trees?

Image-based Semantic Model

Convolutional Neural Networks (Krizhevsky et al., 2012)

Train a network of 5 Convolutional layers, 3 max-pooling layers, and 3 fully connected layers on image-classification

1. Provide 10 images per word as input
2. Do a forward pass and extract pre-final layer
3. Aggregate



Multimodal Semantic Models

Scoring-Level Fusion

- Perform Kernel-Regression Separately
- Compute semantic distance as a weighted average
- Optimize weighting factor (0.75)

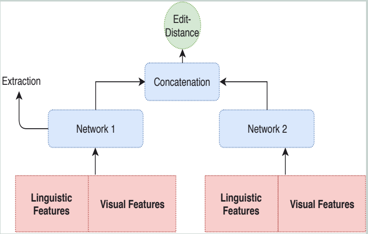
Multimodal Concatenation

- Normalize and concatenate semantic representations
- Optimize weights on the multimodal representations

Multimodal Semantic Models

Neural Network Fusion

- Predict Levenshtein distance
- Extract pre-final layer



Experimental Setup

Lexicon

4479 Monomorphemes

Measuring Corpus-wide Systematicity

Mantel Permutation test for pairwise distances

Identifying Systematic Clusters

Predictability as a measure for systematicity

Quantitative Results

Model	Correl.	p-value
Text-based	0.0362	< 0.001
Image-based	0.0198	0.025
Scoring-level fusion	**0.0401**	< 0.001
Multimodal concat	0.0351	< 0.001
Neural Network fusion	0.0175	< 0.001

Table: Correlations between unweighted Levenshtein distance and semantic distance

Model	Correl.	p-value
Text-based	0.0383	< 0.001
Image-based	0.0243	0.007
Scoring-level fusion	**0.0420**	< 0.001
Multimodal concat	0.0376	< 0.001
Neural Network fusion	0.0266	< 0.001

Table: Correlations between weighted Levenshtein distance and semantic distance

Qualitative Results

Ph	Systematic Words
sn-	sneeze, sniff, snore, snort, snout
pe-	pea, peach, pear, pearl, pebble

Table: Phonaesthemes (text-based model): 22 total

Ph	Systematic Words
cr-	crab, crawl, creep, crouch
si-	sight, sign, silhouette, simulate

Table: Phonaesthemes (multimodal concatenation): 10 total

Conclusion

Systematicity

Our findings corroborate the existence of form-meaning systematicity, and its clustered nature

The importance of a Multimodal approach

Incorporating visual information can increase the level of systematicity identified, and capture novel relations between form and meaning

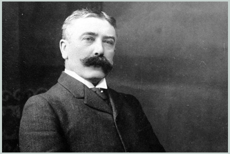
Future Research

How systematic is our language? (and why)

- Is systematicity universal accross languages?
- Why is language systematic?

Expanding the multimodal approach

- Advanced fusion methods
- Incorporating new modalities



(De Saussure, 1916)

"Everything that relates to language as a system must be approached from this viewpoint: the limiting of arbitrariness. This is the best possible basis for approaching the study of language"

References I

Bruni, E., Tran, N. K., and Baroni, M. (2014). Multimodal distributional semantics. *JAIR*, 49:1–47.

De Saussure, F. (1916). *Course in general linguistics*. Columbia University Press.

Gasser, M. (2004). The origins of arbitrariness in language. *Proceedings of CogSci, 26(26)*, 30(5):434–439.

Gutiérrez, E., Levy, R., and Bergen, B. K. (2016). Finding Non-Arbitrary Form-Meaning Systematicity Using String-Metric Learning for Kernel Regression. *Acl*, (1984):2379–2388.

Hockett, C. F. and Hockett, C. D. (1960). The origin of speech. *Scientific American*, 203(3):88–97.

References II

Hutchins, S. S. (1999). The psychological reality, variability, and compositionality of english phonesthemes.

Kiela, D. and Bottou, L. (2014). Learning Image Embeddings using Convolutional Neural Networks for Improved Multi-Modal Semantics. *In Proceedings of EMNLP 2014*, pages 36–45.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *NIPS*, pages 1–9.

References III

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *NIPS*, pages 3111–3119.

Monaghan, P., Shillcock, R. C., Christiansen, M. H., and Kirby, S. (2014). How arbitrary is language? *PTRS B*, 369.

Otis, K. and Sagi, E. (2008). Phonaesthemes: A corpus-based analysis. In *Proceedings of the 11th CogSci*, volume 30.

Shillcock, R., Kirby, S., McDonald, S., and Brew, C. (2001). Filled pauses and their status in the mental lexicon. In *ISCA ITRW on Disfluency in Spontaneous Speech*.

References IV

Zwaan, R. A. and Madden, C. J. (2005). Embodied sentence comprehension. *Grounding cognition: The role of perception and action in memory, language, and thinking*, pages 224–245.

Weight-optimization for Kernel Regression

Predict target variable based on local structure in predictor variable

$$\hat{y}(x_j) = \frac{\sum_{i \neq j} k_{ij} * y_j}{\sum_{i \neq j} k_{ij}}, \tag{2}$$

Kernel penalizing distance

$$k(x_i, x_j) = exp(-d(x_i, x_j)/h) \tag{3}$$

Weight-optimization for Kernel Regression

Mean squared error

$$\mathcal{L} = \sum_{i=1}^{N} ((y_i - \hat{y}_i)^T (y_i - \hat{y}_i)) \tag{4}$$

Gradient

$$\frac{\partial \mathcal{L}}{\partial W} = \frac{2}{N} * \sum_{i=1}^{N} (y_i - \hat{y}_i) * \frac{\sum_{j \neq i} (y_j - \hat{y}_i)^T k_{ij} v_{ij}}{\sum_{j \neq i} k_{ij}} \tag{5}$$

Skip-Gram with Negative Sampling

Objective

$$\sum_{(w,c)\in D} \log \frac{1}{1+e^{-v_c * v_w}} \tag{6}$$

Avoid trivial solution

$$\forall (w,c) \in D' [(w,c) \notin D] \tag{7}$$

Final objective

$$\sum_{(w,c)\in D} \log \frac{1}{1+e^{-v_c * v_w}} + \sum_{(w,c)\in D'} \log \frac{1}{1+e^{v_c * v_w}} \tag{8}$$

Convolutional Neural Network

- 5 Convolutional layers
- 3 Max Pooling layers
- 3 fully connected layers
- One Softmax layer
- ReLU