The CoCo Effect: Quantifying the Conceptual Combination Effect on Word Meanings

Nora Aguirre-Celis[1,2], Risto Miikkulainen[1]
[1]The University of Texas at Austin [2]Monterrey Institute of Technology and Higher Education
{naguirre,risto}@cs.utexas.edu

Abstract

How does concept meaning change in different contexts? People construct meaning dynamically according to the combination of concepts that occur in a sentence, i.e., *olive oil*, *baby oil*, and *oil lamp*. The hypothesis that the meaning of a concept depends critically on the concepts around it has been for a long time, but only recently it has become possible to test it based on neuroimaging and quantify it using computational modeling. In this research, a neural network is trained with backpropagation to map feature-based semantic representations to fMRI images of subjects reading everyday sentences. Backpropagation is then extended to the features, demonstrating how word meanings change in different contexts. Across a large corpus of sentences, the new features are more similar to the features of other words in the sentence than to the original features, validating the CoCo effect. Such dynamic conceptual combination effect could be included in natural language processing systems to encode richer contextual embeddings .

Introduction

How does Concept Meaning Change?

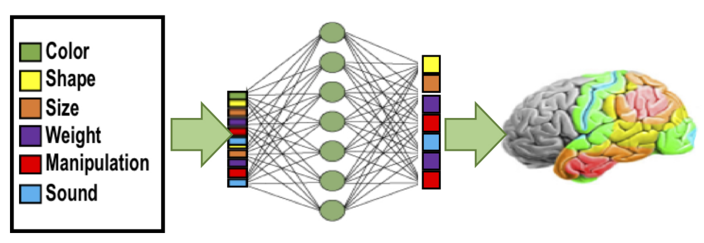
- A concept consists of a collection of features that describe the basic components of meaning
- Grounded on different brain systems (Perceptual Grounding approach, Binder et. al., 2016)
  - sensorimotor, spatiotemporal, social, affective, cognitive, emotional, etc
- People dynamically construct meaning according to the combination of concepts that occur in the stce.
  - "playing the *piano*": the function, sound, and fine hand movements are more relevant
  - "lifting the *piano*": the weight, size and shape matters

- Features are weighted differently according to how important each feature is in that context
- The weights are affected by the sentence, other words in the sentence and context
- It is possible to track the meanings of words by tracking how the weighting changes across contexts

Modeling Approach

Quantifying the Context Changes

- CARs: brain-based semantic representations relate semantic content to brain networks (Anderson et al. 2016)
- Mapping CARs to fMRI patterns as an approach to understand how concepts are encoded in the brain
- Using NN to quantify how meaning change across sentences
  - Predicting Sentence fMRI
  - Changing CARs to account for context



Representation Scheme

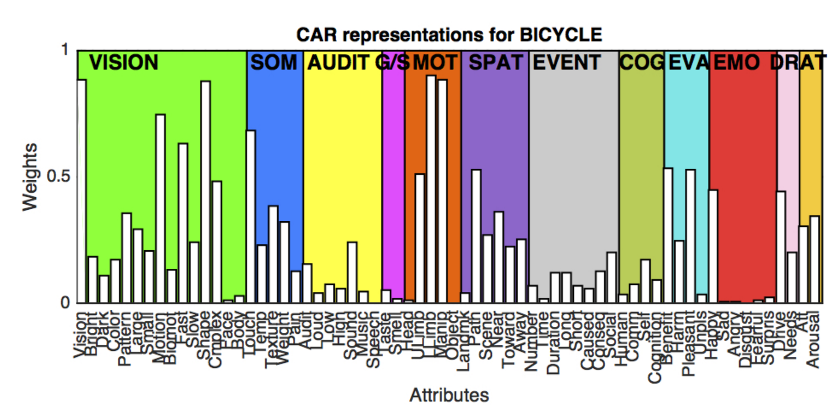
Concept Attribute Representations (CARs)

- CARs represent the basic components of meaning
- Provide a direct correspondence between conceptual content and brain systems
- CARs are composed of a list of modalities that correspond to specialized sensory, motor and affective functions and are not limited to the classical sensory-motor dimensions
- These aspects of mental experience represent each word as a collection of a high-dimensional feature vector
- The attributes were selected after an extensive body of physiological evidence (Binder, et al., 2016)

Data Collection and Preprocessing

Sentences, Words, fMRI Images, and CARs Collection

- A collection of 240 everyday sentences and 242 abstract and concrete words
- Prepared as part of the Program for Knowledge Representation in Neural Systems
- Sentences are composed of two to five content words (141 nouns, 39 adjectives and 62 verbs)
  - *Family survived the powerful hurricane*
- Words include objects, actions, settings, roles, state and emotions, and events
  - *couple, author, theater, hospital, desk, red, damaged, gave, happy, chicken*
- Sentence fMRI data was acquired by the the Medical College of Wisconsin
- Eleven subjects took part in the experiment producing 12 repetitions each
- Participants viewed the sentences word by word while in the scanner
- It was preprocessed and averaged with a final selection of 396 voxels per sentence for each subject
- fMRI images for words were synthetized by averaging all fMRI images for the sentence where each word occurred (Anderson et al., 2016)
- Final collection includes 237 sentences and 236 words (138 nouns, 38 adjectives and 60 verbs)
- Semantic vectors of word representations were collected via Amazon Mechanical Turk
- Words were represented by CARs in a high-dimensional set of 66 attributes
- Approximately 30 ratings were collected

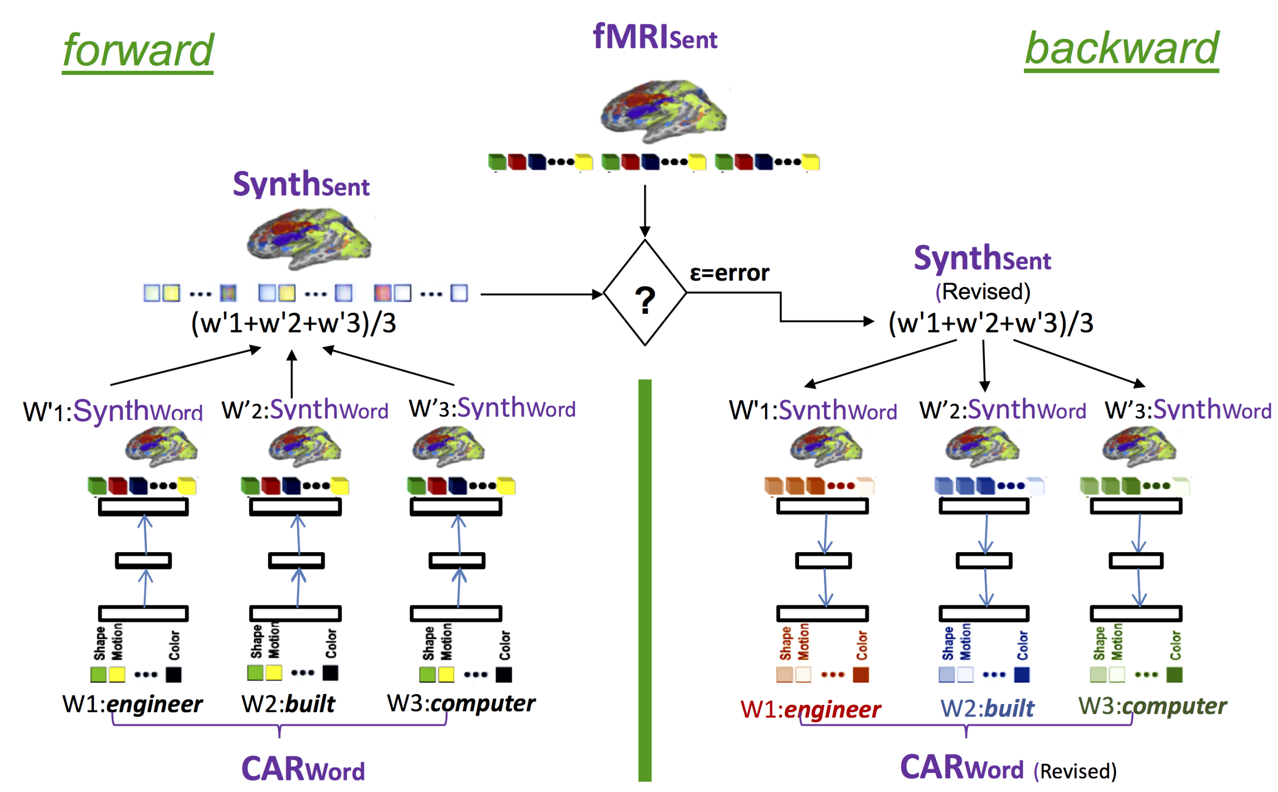


BRAIN AREAS	ATTRIBUTES	EXPLANATION
VISION	Vision	is visible
	Bright	is bright
	Dark	is dark
	Color	has a defining color
	Pattern	has texture or pattern
	Large	is large
	Small	is small
	Motion	moves
	Biomotion	living thing that moves
	Fast	moves fast
	Slow	moves slow
	Shape	has a defining shape
	Complexity	is visually complex
	Face	has a face
	Body	has body parts

Computational Model

Predicting fMRI Sentences with NN and Changing CARs with FGREP

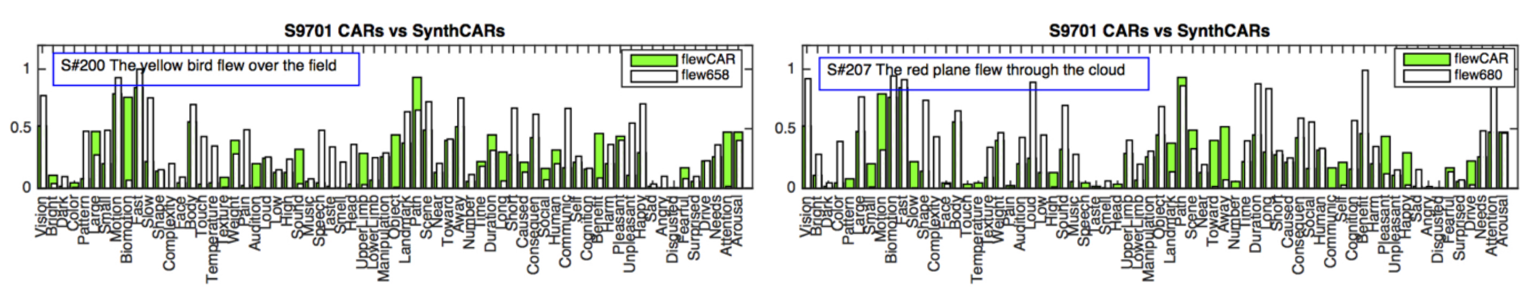
- Train a NN to map CARWord to SynthWord for all sentences
- Assemble SyntSent for the predicted sentence by averaging SynthWords
- SynthSent is then compared to the observed fMRISent to form a new error signal
- Backpropagate the error from fMRISent separately for each sentence
- Freeze weights; only change CARs using FGREP
  (Forming Global Representations with Extended BP, Miikkulainen, 1991)
- Using the new CARs, the process is repeated several times until zero error
- Trained 20 times with different seeds for each of the 11 fMRI subjects
- Mean of the new CAR representations was used to analyze the experiments



Results

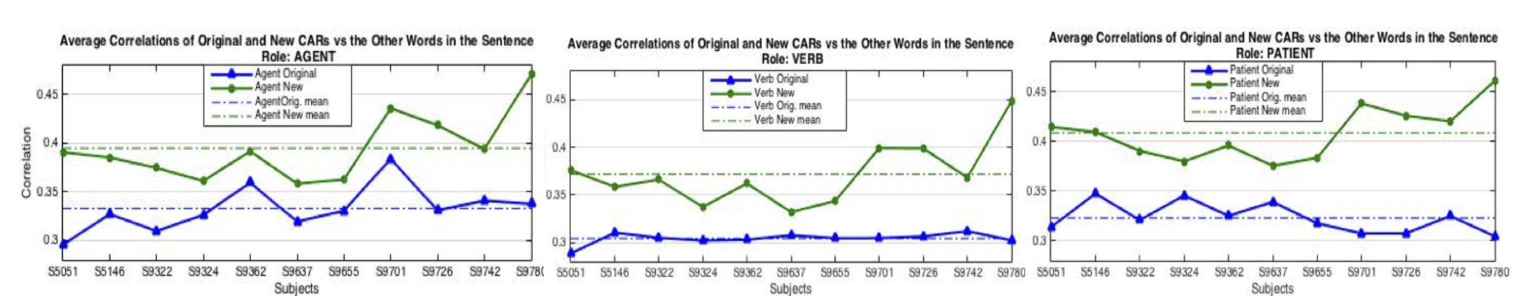
Previous Work: Individual Cases

- An individual example of noun-verb interactions for the verb *flew*
- FGREP exposed the features to represent animacy
  - Stce. 200: High values on animate features like Pain, Small, Smell and Taste, Speech and Communication
  - Stce. 207: High values on inanimate object features like a Loud, Large, Heavy, Duration, Social, Benefit and Attention
- The noun-verb combinations show how some noun properties were transferred to the verb; sharing similar features



Current Work: Aggregation

- Quantifying the CoCo effect through a statistical analysis across an entire corpus of stces. and subjects
  - The CoCo effect use one or more features of one concept to describe other concept (attribute combination)
  - In CAR theory, concepts interactions arise within multiple brain networks, activating similar brain zones for both concepts
  - These interactions determine the meaning of the concept combination (CoCo Effect)
- Similar sentences have a similar effect, and this effect is consistent across all words in the sentence
- Correlate the original CAR and new CAR representations vs. the other words in the sentence (OWinS)
- The correlation analysis measures whether the word meaning changes towards the context meaning
- Correlations are significantly higher for new CARs than orig. CARs across subjects, sentences, and roles
- Orig CARs average corr. 0.3201 (STDEV 0.020), new CARs 0.3918 average corr. (STDEV 0.034)
- Results confirm the CoCo effect: new CARs are more similar to the OWinS. than to orig. CARs



Conclusions

- Meaning can be characterized using fMRI images, CARs and Neural Networks can be used to map meaning to fMRI, and FGREP can be used to quantify how the meaning of words change in context
- Across an entire corpus of sentences, new CARs are more similar to the OWinS than to the orig. CARs
- Evaluates the soundness of the CARs approach as a measure of semantics by capturing the fMRI sentence representations grounded in specific neural systems
- This work differs from text-based semantic models by using brain-based semantic representations (CARs) to directly associate semantic content to brain networks
- In the future, it may be possible to use such dynamic representations to build artificial natural language processing systems by dynamically adapting the vector representations to fit context

References

1. Anderson, A. J., Binder, J. R., et. al. (2016). Perdicting Neural activity patterns associated with sentences using neurobiologically motivated model of semantic representation. Cer. Cortex. 1-17
2. Binder, J. R., Conant L. L., et. al. (2016). Toward a brain-based componential semantic representation. Cog. Neuropsychology, 33:3-4, 130-174
3. Miikkulainen, R., Dyer, M., G. (1991). Natural Language Processing with Modular PDP Networks and Distributed Lexicon. Cognitive Science, 15, 343-399

Acknowledgements

Special thanks to Jeffery Binder (Medical College of Wisconsin), Rajeev Raizada and Andrew Anderson (University of Rochester), Mario Aguilar and Patrick Connolly (Teledyne Scientific Company) for their for their work and valuable help regarding this research. This research was supported in part by IARPA-BAA-12-05 and NIH 1U01DC014922 grants.