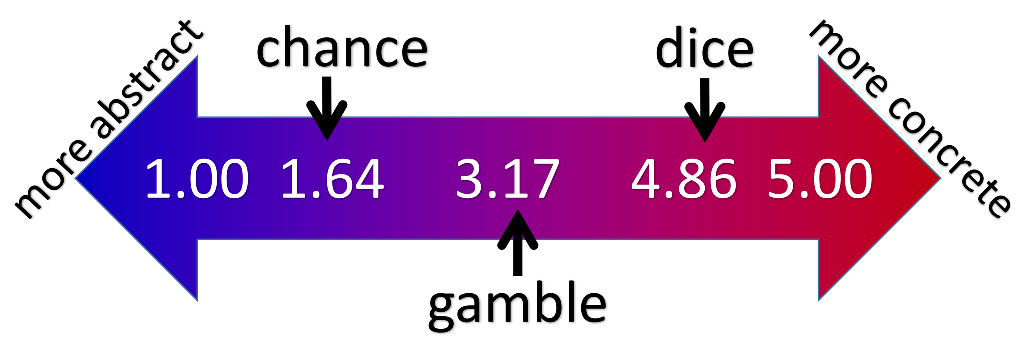
Discrimination and Prediction of Concreteness from Neuroimaging, Behavioral, and Corpus Data

Dominick DiMercurio II, MS; Chaleece Sandberg, PhD CCC-SLP
Department of Communication Sciences and Disorders, Pennsylvania State University

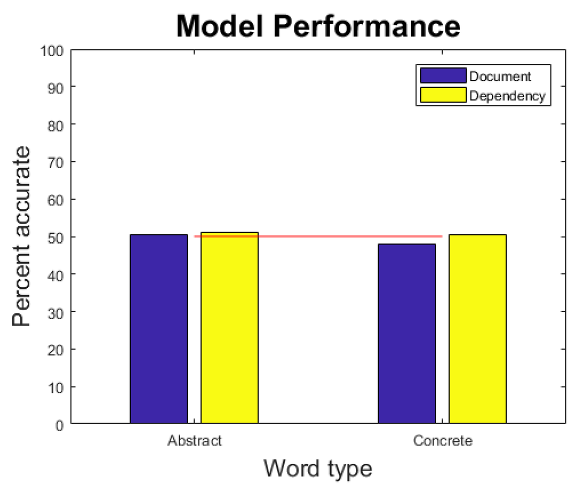
Background

The semantic system remains a mystery in cognitive neuroscience.

- Previous research suggest dissociation between processing of:
  - **Concrete words** (referring to objects detectable by senses)
  - **Abstract words** (referring to more intangible notions)

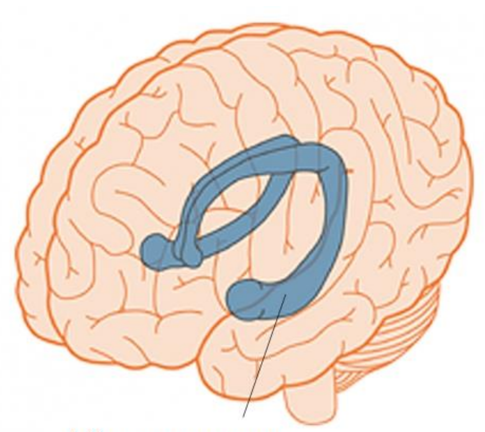


- Neuroinformatics is an emerging field to help understand neural systems.
- Predicting words with semantic space models (Mitchell et al., 2008)
  - *Semantic space models performed near chance with our fMRI data.*



Our fMRI Data	Mitchell et al. 2008
Older adults	Young adults
60 concrete words	60 concrete words
60 abstract words	No abstract words
Two repetitions	Six repetitions
Fyshe vectors	Custom vectors
Shallow task	Deep task

- Our data might be unsuited to semantic space models due to the shallow task, but might remain compatible with less sophisticated learning.
- Previous research by Wang et al. (2013) suggests that accuracy in learning concreteness from fMRI data is region-dependent:
  - Inferior frontal gyrus (IFG) ~ 75% accurate
  - Middle temporal gyrus (MTG) ~ 75% accurate
  - Hippocampus ~ 52% accurate
- Wang et al. notice left-lateralization in their study.


Hippocampus

Neuroimaging Methods

- 11 Neurotypical Adults
- Word judgment in MRI scanner
- 60 abstract, 60 concrete, 120 nonwords
  - Vowels or consonants (e.g. *auuai* or *dmsdf*)

	01	02	03	04	05	06	07	08	09	10	11
Sex	F	M	M	F	F	M	M	M	F	F	M
Age	64	66	47	55	72	74	72	57	59	61	56

	ACC	RT	AOA	CNC	FAM	IMG	NLET	NSYL
Concrete	0.95	1175.83	336.43	570.52	519.96	575.93	6.94	2.06
Abstract	0.89	1485.43	333.65	455.60	540.56	393.76	7.00	2.31
p-value	<.001	<.001	.015	<.001	.091	<.001	.874	.217

General linear models of BOLD signal
- Performed per stimulus per participant
- Included motion, word length, scan, and other regressors in the GLMs
- Resulting beta values per voxel of ROI:
  - L IFG, R IFG, L MTG, R MTG, B Hip
- Resulting vectors were used in support vector machines
  - Recursive feature elimination (RFE)
  - Correlation bias reduction (CBR)
  - As described in Yan & Zhang (2015)

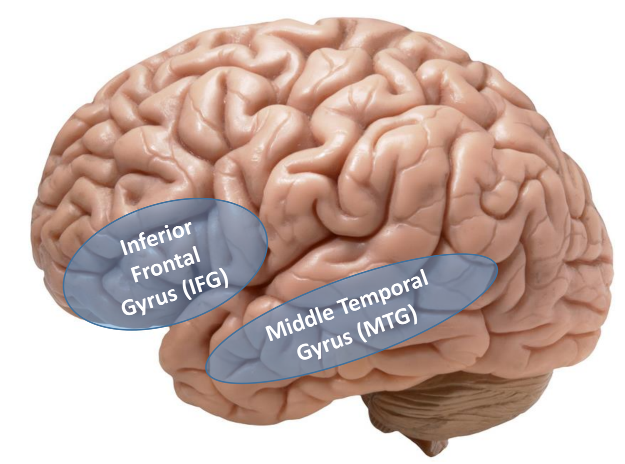
Neuroimaging Results

Support vector machine models of concreteness were developed with fMRI data.

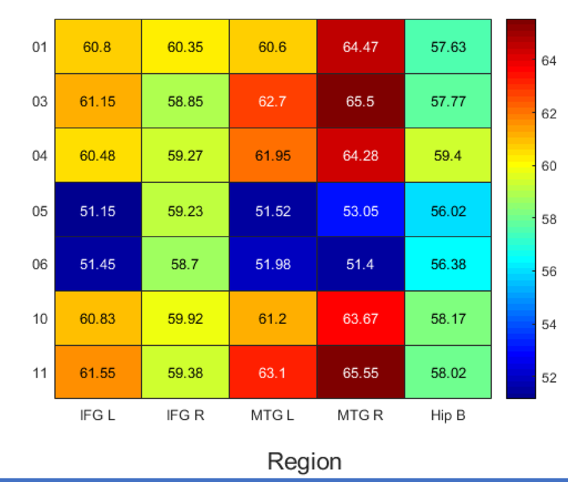
- Performed SVM RFE CBR on fMRI
- 100 trials on 200 voxels
- Trained 12, predicted 12 words
- Words (Abstract vs. Concrete)
  - μ = 50.81%
- Nonwords (Consonants vs. Vowels)
  - μ = 46.23%
- Two-tailed paired t-test: p < 0.001

	Words	Nonwords
IFG L	54.75%	45.67%
IFG R	49.67%	48.50%
MTG L	51.67%	45.58%
MTG R	50.25%	44.08%
Hip B	49.67%	45.42%

Sample results from support vector machines on data from participant 03.



- Trained 60, predicted 40 words
- 02, 07, 08, 09 excluded for too few words
- Balanced for category and concreteness
- Words (Abstract vs. Concrete)
  - μ = 59.07%
- Two-way ANOVA:
  - Effect of participant: p < 0.001
  - Effect of region: p < 0.001
  - Interaction: p < 0.001

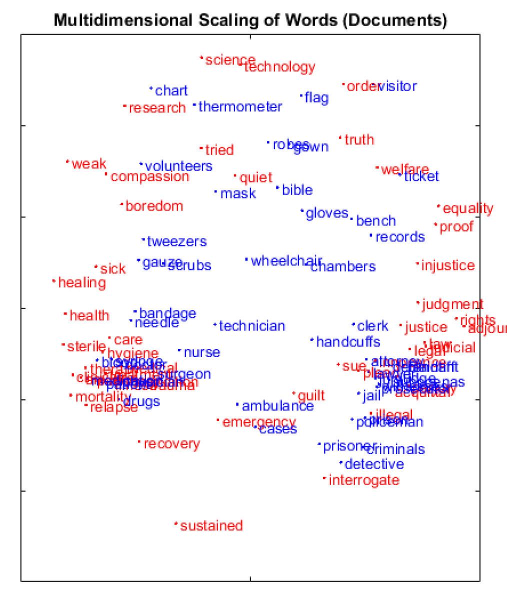


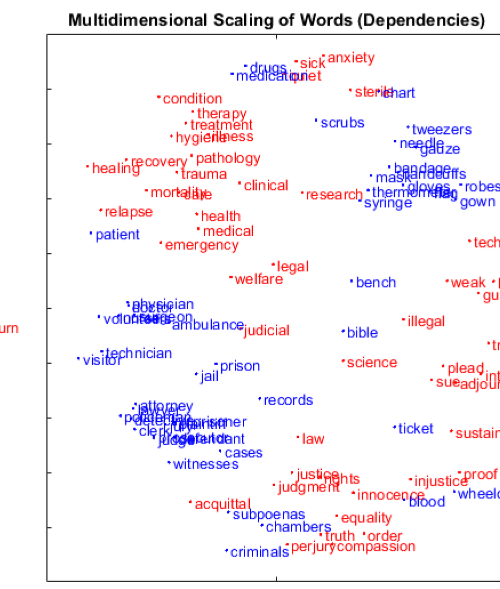
Corpus Results

Semantic vectors for words from our study provided in a database by Fyshe et al. (2013).

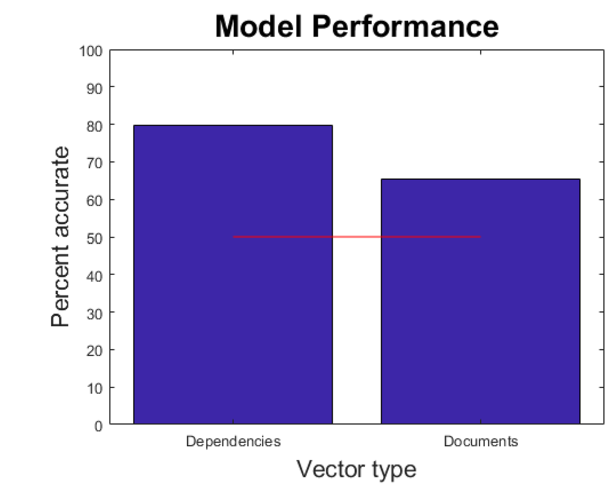
Documents:
- Co-occurrence in documents

Dependencies:
- Co-occurrence of syntactic relationships

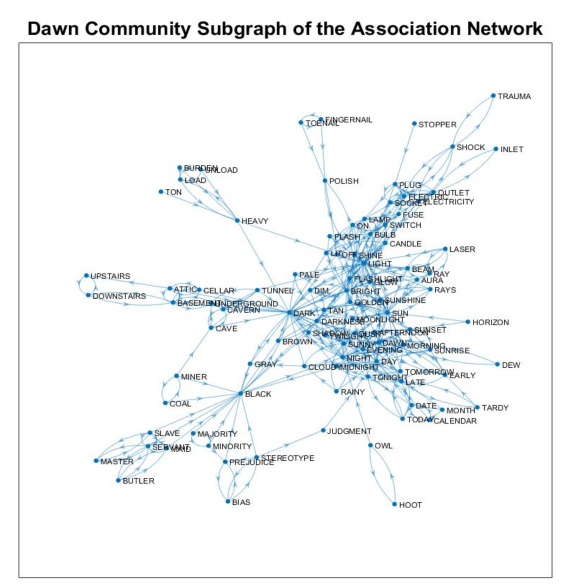




- Performed SVM on Fyshe vectors
- 100 trials on first 300 features
- Trained 48 words, predicted 24 words
- Documents (Abstract vs. Concrete),
  - μ = 65.50%
- Dependencies (Abstract vs. Concrete)
  - μ = 79.78%
- Two-tailed t-test: p < 0.001



Further Analysis



Association network created for SVM study:
- Free association norms collected from a public database from the University of South Florida.
- A digraph of 4763 nodes (i.e., words), weighted by forward strengths, was created, with modules detected by Louvain method (Q = .6265).
- The within-module degree z-score of each node was calculated, and the node with the highest z-score was deemed the *community leader*.
  - **Left:** Example of community with leader *dawn*.
- Norms for 80 of our words (42 abstract) were available. Vectors created with each value being the distance to 104 selected nodes.

- SVM RFE CBR on association network
- 100 trials (Abstract vs. Concrete)
- Used 104 nodes for feature extraction:
  - *Randoms* – random nodes, μ = 50.15%
  - *Members* – within-module random, μ = 63.10%
  - *Leaders* – as described earlier, μ = 69.00%
- Trained 38 words, predicted 38 words
- One-way ANOVA: p < 0.001

Discussion

Interpretations:
- Support vector machines trained on fMRI data were successful.
- Individual variability is seen. Might be linked to age.
- Models respond differently to different regions of interest.
- Corpus data can also be used in support vector machines to discriminate concreteness.
- Association networks hint at a promising next step.

Future studies:
- Attempt modeling with other types of corpus methods
- Attempt modeling with other types of semantic networks (such as taxonomic or featural)
- Improve fMRI design to be more compatible with models

Broader impact:
- *Scientific relevance:*
  - Can help us to understand semantic system
  - Can be adapted for questions in other fields
- *Technological relevance:*
  - Used to develop recommender systems
  - Used to develop text-based diagnostic tools
- *Clinical relevance:*
  - Can explain findings in people with aphasia
  - Can expand to simulate aphasia and inform therapy

References:

Blondel, VD, Guillaume, J, Lambiotte, R, Lefebvre, E. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* (10), P10008.
Brysbaert, M, Warriner, AB, Kuperman V. 2014. Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods,* 46. 904-911.
Fyshe, A, Talukdar, P, Murphy, B, Mitchell, T. 2013. Documents and Dependencies: an Exploration of Vector Space Models for Semantic Composition. *Proceedings of Seventeenth Conference on Computational Natural Language Learning*, 84-93.
Mitchell, TM, Shinkareva, SV, Carlson, A, Chang K, Malave, VL, Mason, RA, Just, MA. 2008. Predicting Human Brain Activity Associated with the Meanings of Nouns. *Science*, 320. 1191-1195.
Nelson, DL, McEvoy, CL, Schreiber, TA. 1998. The University of South Florida word association, rhyme, and word fragment norms. http://www.usf.edu/FreeAssociation/.
Wang, J, Baucom, LB, Shinkareva, SV. 2013. Decoding Abstract and Concrete Concept Representations Based on Single-Trial fMRI Data. *Human Brain Mapping,* 34. 1133-1147
Yan, K, Zhang, D. 2015. Feature selection and analysis on correlated gas sensor data with recursive feature elimination. *Sensors and Actuators B: Chemical*, 212. 353-363.