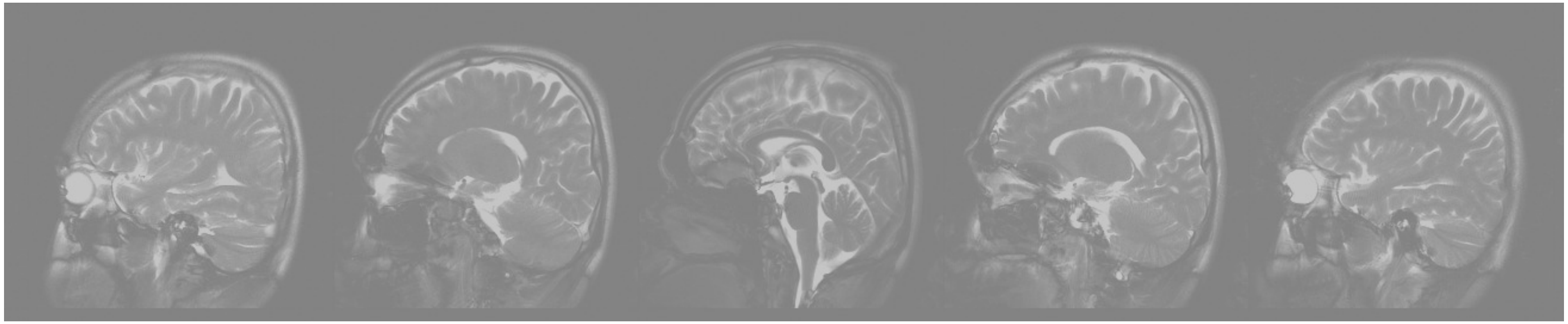


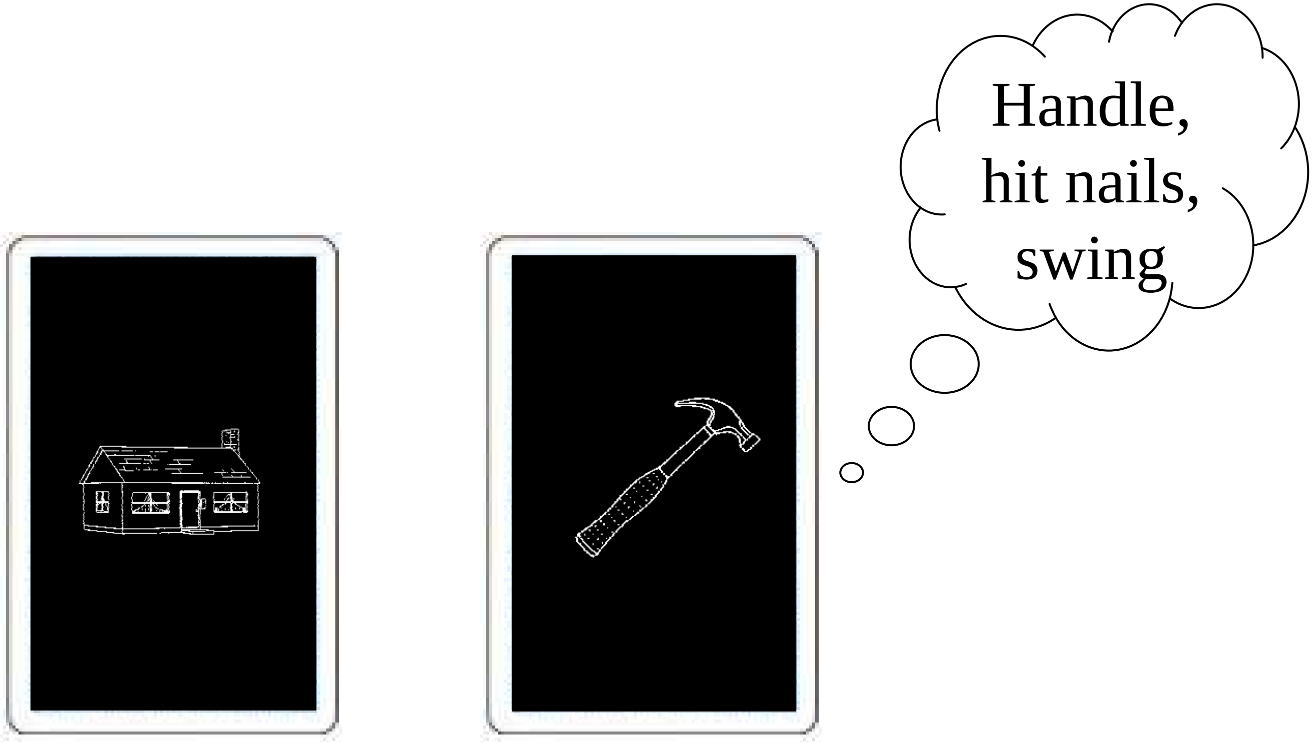
Quantitative modeling of the neural representation of semantic compositions

Student:	Kai-min Kevin Chang
Committee Members:	Marcel Adam Just (chair)
	Tom Mitchell (co-chair)
	Charles Kemp
	Brian Murphy (University of Trento)

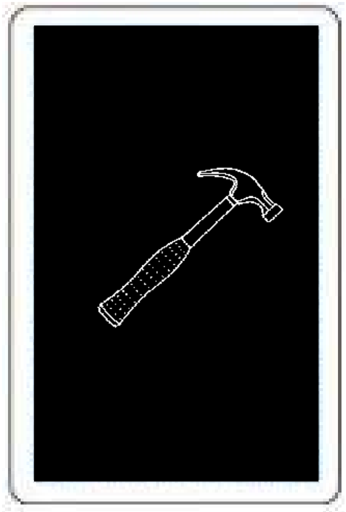
Feb 2, 2010 LTI Thesis Proposal

Magic Trick…
(well, a hypothetical one)


Pick a card and think consistently about properties of the object shown in that card

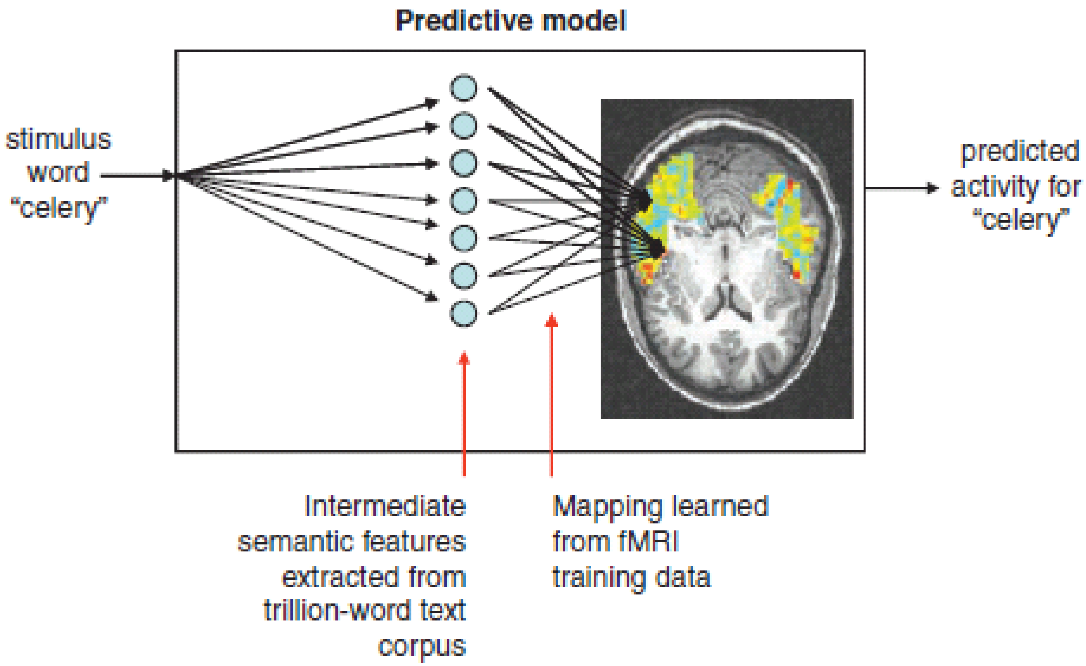


We can correctly predict which card you picked 79% of the time and there is no trick, we did it by reading your mind!



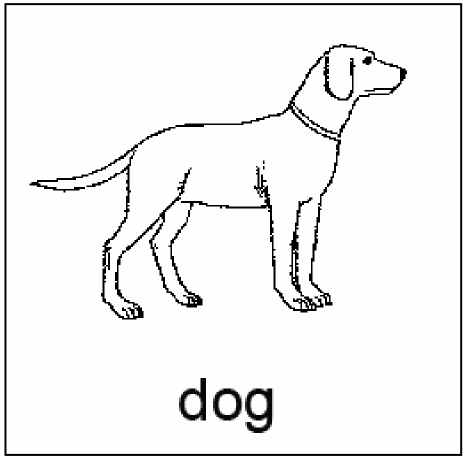
Sixty Words Experiment

- We developed a generative model that is capable of predicting fMRI neural activity well enough that it can successfully match words it has not yet encountered, with accuracies close to 79% (Mitchell et al., 2008).



From Nouns to Phrases

1. Can we decode which noun or adjective-noun phrase a participant is thinking?
2. How does the brain compose the meaning of words or phrases?



strong dog

Thesis Statement

- The thesis of this research is that the distributed pattern of neural activity can be used to model how brain composes the meaning of words or phrases in terms of more primitive semantic features.

Three Major Advancements

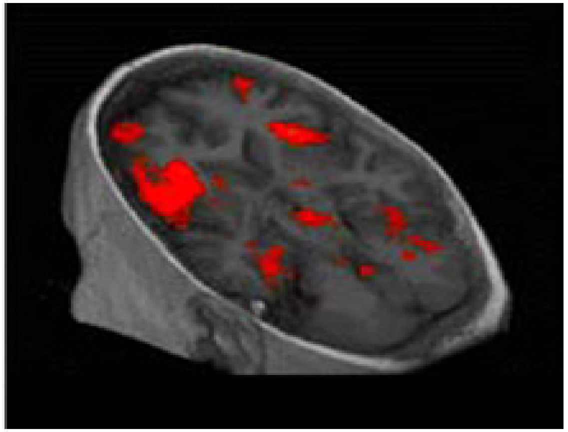
- Brain imaging technology allows us to directly observe and model neural activity when people read words or phrases.
- Machine learning methods can automatically learn to recognize complex patterns.
- Linguistic corpora allow word meanings to be computed from the distribution of word co-occurrence in a trillion-token text corpus.

Overview

1. Thesis statement
2. Brain imaging experiment
3. Methodology
4. Results to date
5. Proposed work

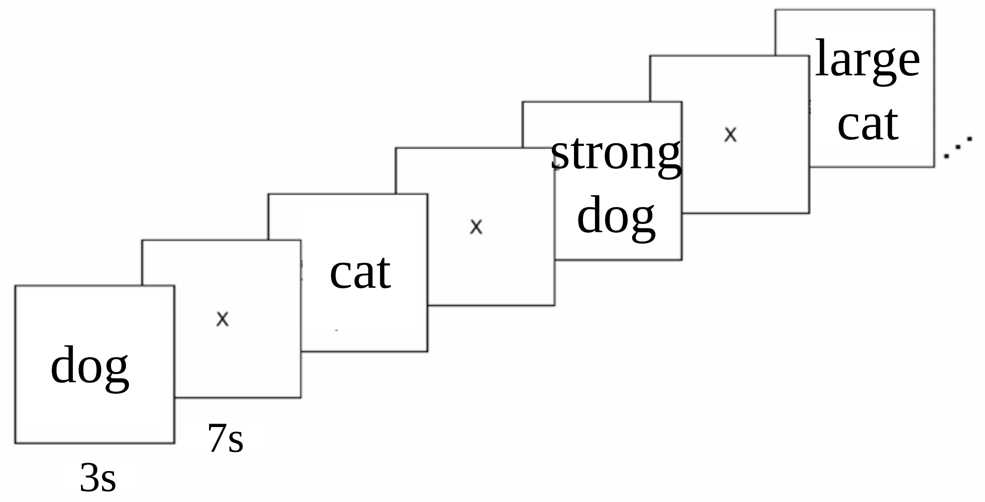
Functional Magnetic Resonance Imaging (fMRI)

- Measures the hemodynamic response (changes in blood flow and blood oxygenation) related to neural activity in the human brain.
- The activity level of 15,000 - 20,000 brain volume elements (voxels) of about 50 mm^3 each can be measured every second.



Brain Imaging Experiment

- Human participants were presented with line drawings and/or text labels of nouns (e.g. *dog*) and phrases (e.g. *strong dog*).
- Instructed to think of the same properties of the stimulus object consistently during multiple presentations.
- Each object is presented 6 times with randomized order.



fMRI Data Processing

- Data processing and statistical analysis were performed with Statistical Parametric Mapping (SPM) software.
- The data were corrected for slice timing, motion, linear trend, and were temporally smoothed with a high-pass filter using 190s cutoff.
- The data were normalized to the MNI template brain image using 12-parameter affine transformation and resampled to 3x3x6 mm^3 voxels.

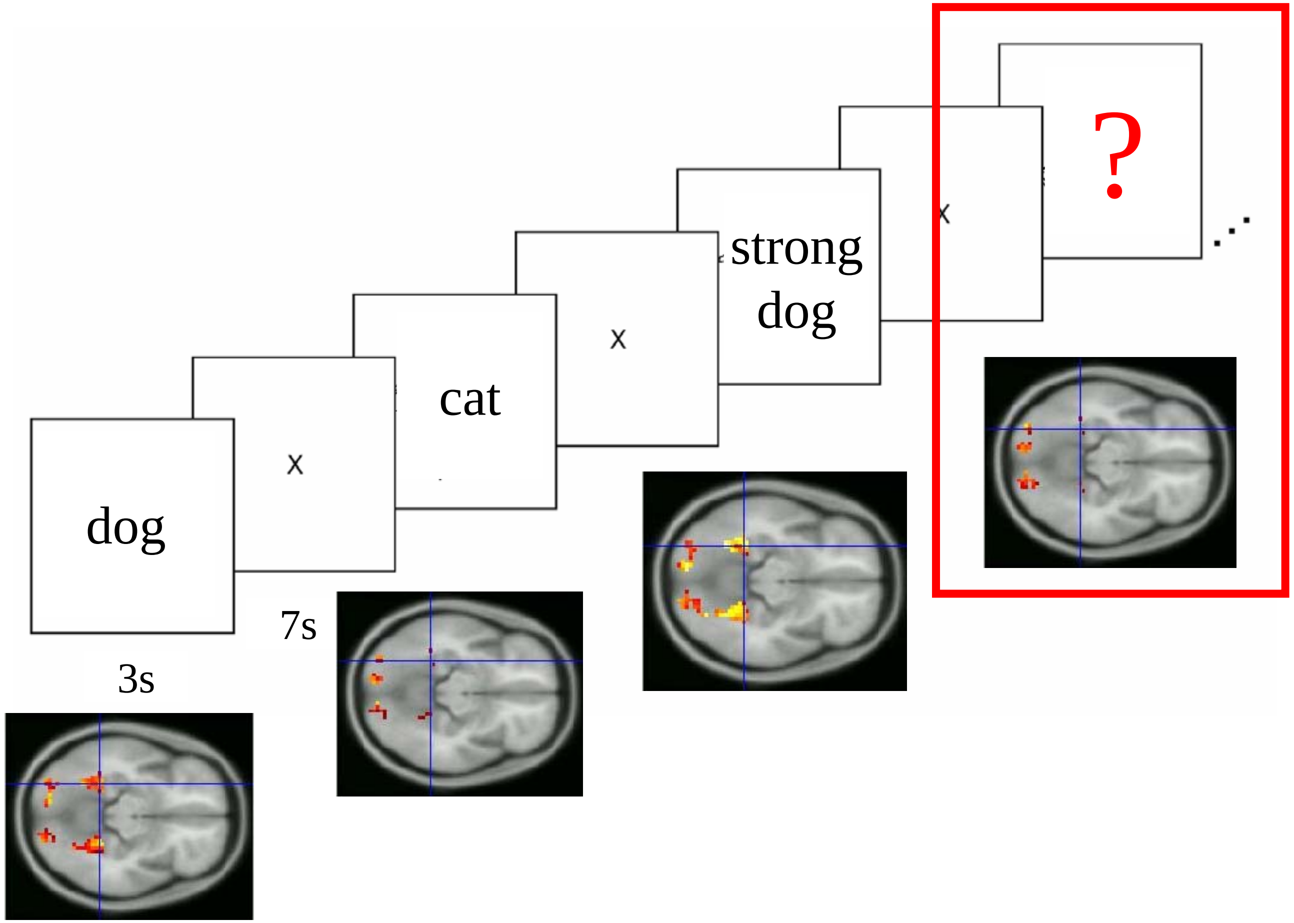
fMRI Data Processing

- Consider only the spatial distribution of the neural activity.
- Select voxels whose responses are most stable across presentations.
- The percent signal change (PSC) relative to the fixation condition was computed.

Overview

1. Thesis statement
2. Brain imaging experiment
3. Methodology
    - Decode mental state
    - Predict neural activity
4. Results to date
5. Proposed work

Decode Mental State



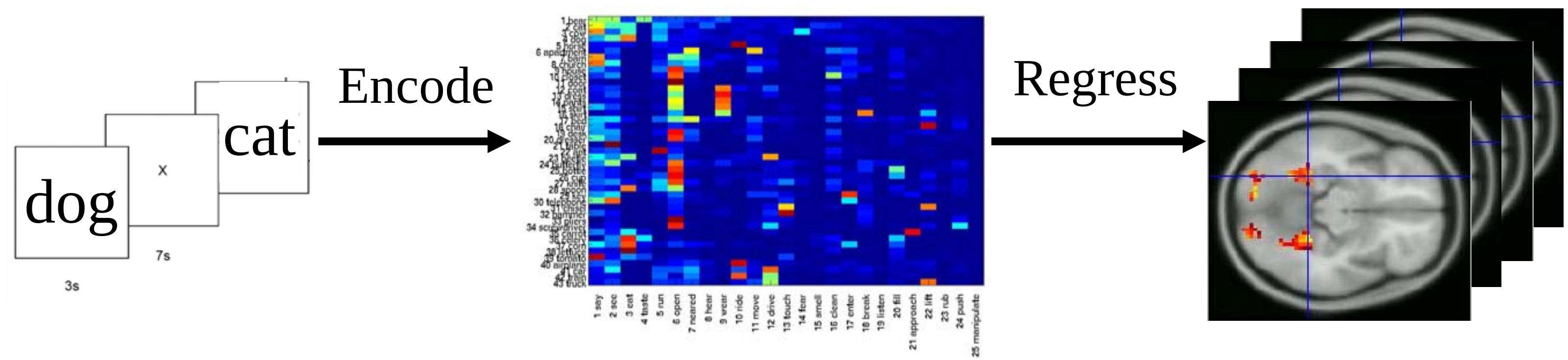
Which noun or adjective-noun phrase is the participant thinking?

Classifier Analysis

- Classifiers were trained to identify cognitive states associated with viewing stimuli.
- Gaussian Naïve Bayes (GNB), Support Vector Machine (SVM), Logistic Regression.
- 6-fold cross validation.
- Rank accuracy was used as a measure of classifier performance (Mitchell et al., 2004).

Predict Neural Activity

- Discriminative classification provides a characterization of only a particular dataset.
- We want to predict neural activity for previously unseen words.



Stimulus

Semantic Representation

Observed Activation

Vector-based Semantic Representation

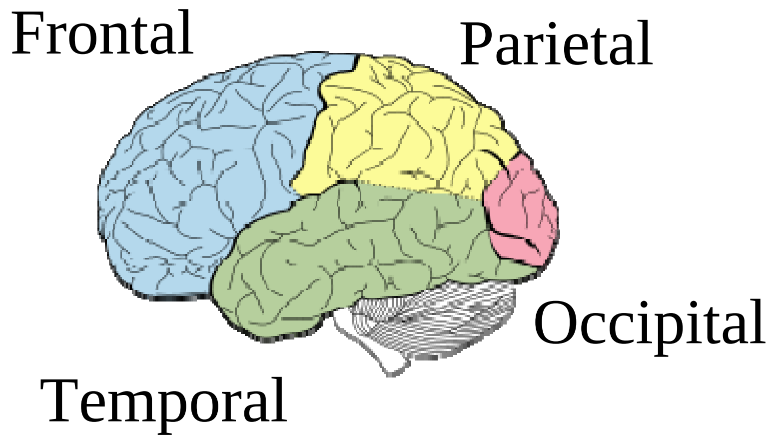
- Words with similar meaning often occur in similar contexts
  - Word meanings can be computed from the distribution of word co-occurrence in a text corpus (Lund & Burgess, 1996; Landauer & Dumais, 1997).
- Google trillion-tokens text corpus, with co-occurrence counts in a window of 5 words.
- Sensory-motor features.

	See	Hear	Smell	Eat	Touch
Strong	0.63	0.06	0.26	0.03	0.03
Dog	0.34	0.06	0.05	0.54	0.02

Linear Regression Model

- Learn the mapping between semantic features and voxel activations with regression.
  - "Touch" feature predicts activation in prefrontal cortex.
  - "Eat" feature predicts activation in gustatory cortex.
- The regression fit, R^2, measures the amount of systematic variance in neural activity explained by the model.

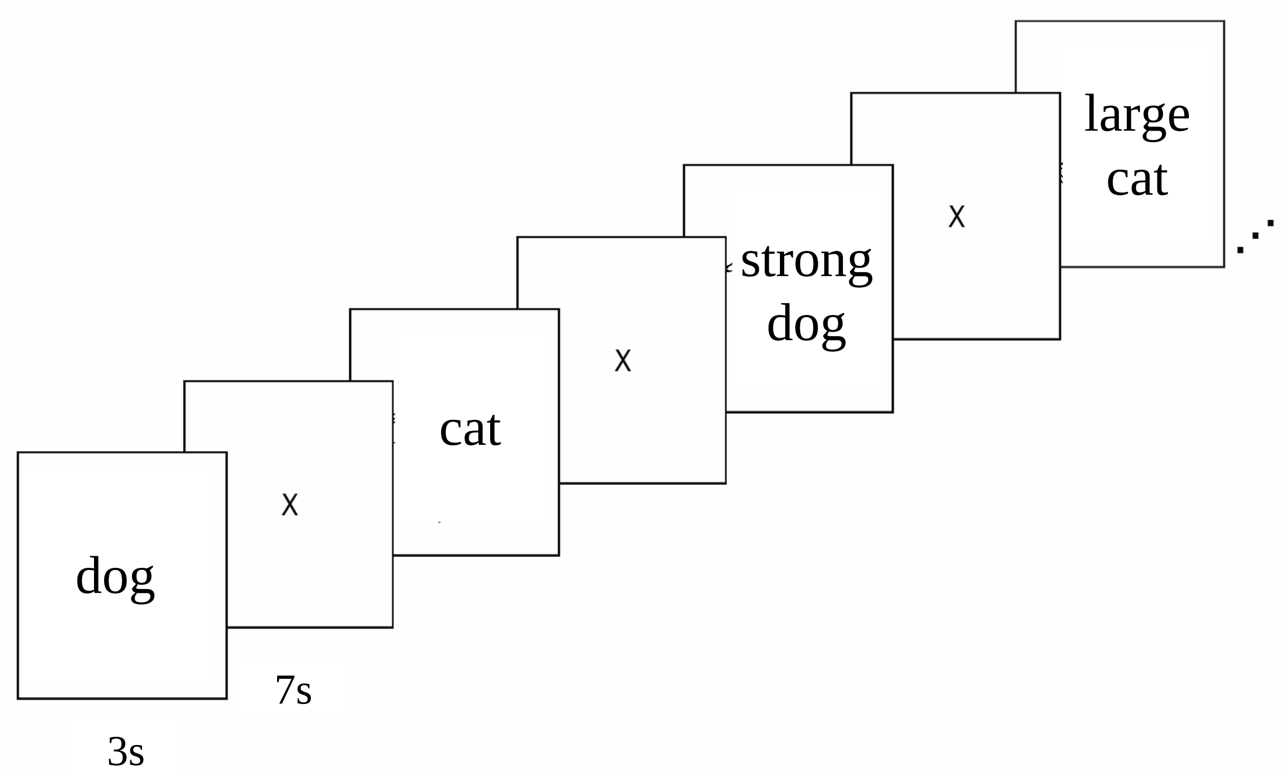
$$a_v = \sum_{i=1}^{n} \beta_{vi} f_i(w) + \varepsilon_v$$



Overview

1. Thesis statement
2. Brain imaging experiment
3. Methodology
4. Results to date
    - Adjective-noun experiment
    - Decode mental state
    - Predict neural activity
5. Proposed work

Adjective-Noun Experiment (Chang et al., 2009)



Word Stimuli

Adjective	**Noun**	**Category**
Soft	Bear	Animal
Large	Cat	Animal
Strong	Dog	Animal
Plastic	Bottle	Utensil
Small	Cup	Utensil
Sharp	Knife	Utensil
Hard	Carrot	Vegetable
Cut	Corn	Vegetable
Firm	Tomato	Vegetable
Paper*	Airplane	Vehicle
Model*	Train	Vehicle
Toy*	Truck	Vehicle

Decode Mental State

- All rank accuracies were significantly higher from chance levels computed by permutation tests.
- Classifier performed significantly better on the nouns than the phrases.

Classifying	**Rank Accuracy**
All 24 exemplars	0.69
12 nouns only	0.71
12 phrases only	0.64

Predict Neural Activation

- Need to represent the meaning of phrases.
- Mitchell & Lapata (2008) presented a framework for representing the meaning of phrases in the vector space.

Strong Dog	**See**	**Hear**	**Smell**	**Eat**	**Touch**
Adjective	0.63	0.06	0.26	0.03	0.03
Noun	0.34	0.06	0.05	0.54	0.02
Additive	0.97	0.12	0.31	0.57	0.05
Multiplicative	0.21	0.00	0.01	0.01	0.00

Semantic Composition Models

- The adjective and the noun model assume people focus exclusively on one of the two words.
- The additive model assumes that people concatenate the meanings of the two words.
- The multiplicative model assumes that the contribution of the modifier word is scaled to its relevance to the head word, or vice versa.

Strong Dog	**See**	**Hear**	**Smell**	**Eat**	**Touch**
Adjective	0.63	0.06	0.26	0.03	0.03
Noun	0.34	0.06	0.05	0.54	0.02
Additive	0.97	0.12	0.31	0.57	0.05
Multiplicative	0.21	0.00	0.01	0.01	0.00

Comparing Semantic Composition Models

- The noun in the adjective-noun phrase is usually the linguistic head.
  - Noun > Adjective.
- Adjective is used to modify the meaning of the noun.
  - Multiplicative > Additive.

Composition Model	**R^2**
Adjective	0.34
Noun	0.36
Additive	0.35
Multiplicative	0.42

Comparing Two Types of Adjectives

- Attribute-specifying adjectives (e.g., *strong*, *large*)
  - Simply specifies an attribute of the noun (e.g., *strong dog* emphasizes the strength of a dog).
- Object-modifying adjectives (e.g., *paper*, *model*)
  - These modifiers combine with the noun to denote a very different object from the noun in isolation (e.g. *paper airplane* is a toy used for entertainment, whereas *airplane* is a vehicle used for transportation).

Decode Mental State

- Harder to discriminate between *dog* and *strong dog* (attribute-specifying).
- Easier to discriminate between *airplane* and *paper airplane* (object-modifying).

	Accuracy
Attribute-specifying	0.68
Object-modifying	0.76

Predict Neural Activity

- For the object-modifying adjectives, the adjective and additive model now perform better.
  - Suggests that when interpreting phrases like *paper airplane*, it is more important to consider contributions from the adjectives, compare to when interpreting phrases like *strong dog*.

Overview

1. Thesis statement
2. Brain imaging experiment
3. Methodology
4. Results to date
5. Proposed work

Proposed Work

1. Noun-noun concept combination experiment.
2. Extend the semantic composition model.
   A. Feature norming features.
   B. Infinite latent feature model.
3. Explore the time series data.

1. Noun-noun Concept Combination

- To study semantic composition:
  - Record activation for the individual words.
  - Work with nouns.
  - Avoid lexicalized phrases (e.g. *paper airplane*).
  - Investigate specific combination rules
    - Concept combination can be polysemous.

Two Types of Interpretations

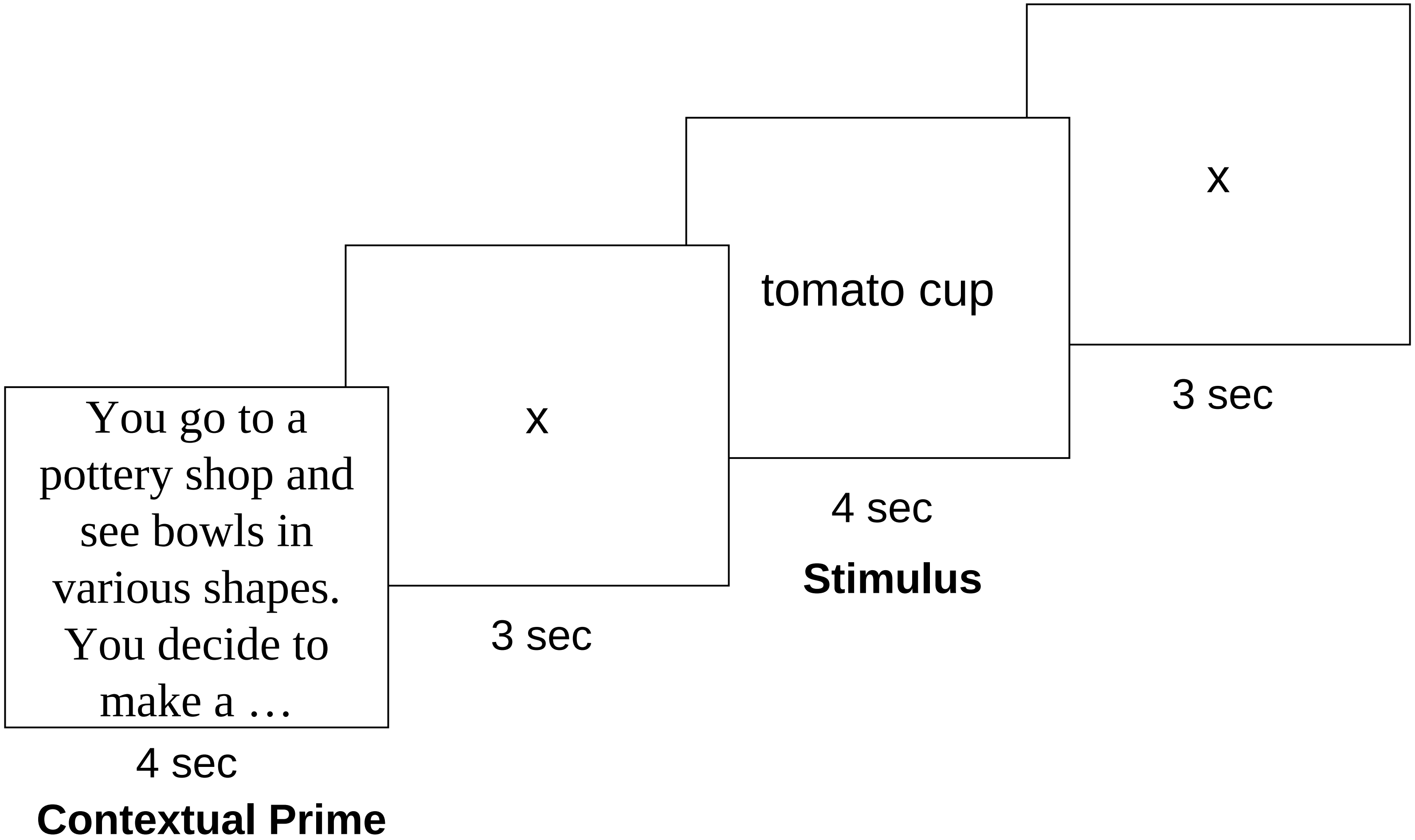
- Property-based interpretation, one property (e.g., shape, color, size) of the modifier object is extracted to modify the head object.
  - For example, *tomato cup* is a cup that is in the shape of a tomato.

- Relation-based interpretation, the modifier object is realized in its entirety and related to the head object as a whole.
  - For example, *tomato cup* is a cup that is used to scoop (cherry) tomatoes.

Noun-noun Concept Combination

- Contexts are used to bias toward certain interpretations:
  - Property-based: "*You go to a pottery shop and see bowls in various shapes. You decide to make a …*" will lead the participant to interpret a *tomato cup* that is in the shape of a tomato.
  - Relation-based:"*You go to a farmer's market to buy some fruits. You scoop with a …*" will lead the participant to interpret a *tomato cup* as a cup that is used to scoop tomatoes.

1. Noun-Noun Experiment

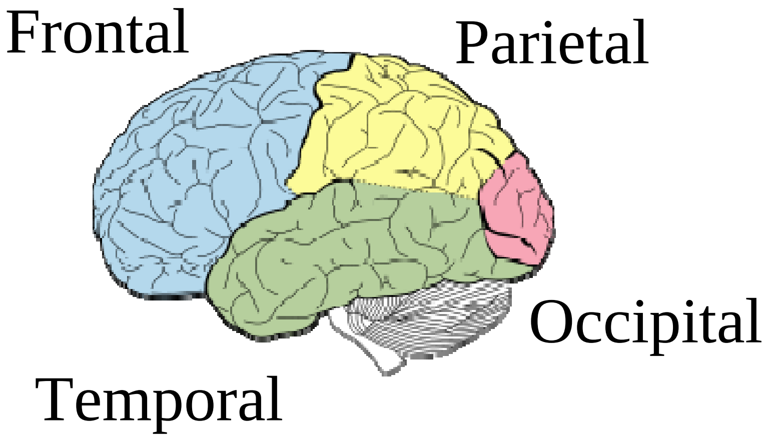


Word Stimuli

window cup
cow chair
corn coat
bell dress
bee airplane
pliers hand
dog beetle
refrigerator house
celery table
tomato ant

Stable Voxels from Different Areas (Preliminary Result)

- For nouns
  - Occipital, Postcentral
- For contextual primes
  - Frontal
- For phrases
  - Fusiform, Temporal



Exemplar Classification (Preliminary Result)

- Classify individual exemplars (rank accuracies).
- Classification rank accuracies significantly higher than chance.

	AVG	P1	P2	P3	P4
20 Noun	0.73	0.70	0.66	0.74	0.81
20 Phrase	0.72	0.75	0.69	0.64	0.78

Category Classification (Preliminary Result)

- Classify property-based or relation-based (accuracies).
- Can discriminate between two types of stimuli interpretations, but not contextual sentences.

	AVG	P1	P2	P3	P4
Context	0.50	0.49	0.48	0.51	0.53
Stimuli	0.62	0.64	0.58	0.61	0.63

Comparing Neural Activity for Phrases to Individual Words (Preliminary Result)

- Correlate the neural activity for phrases to individual words (correlations).
- Property-based: more similar to modifier word.
- Relation-based: more similar to head word.

	Modifier	Head
Property-based	0.48	0.12
Relation-based	0.29	0.42

2. Extend Semantic Composition Models

- Current semantic composition models are overly simplistic:
  - Do not differentiate between different types of interpretation of the same stimulus.
  - Do not reflect the asymmetry between the head and modifier noun.

2A. Feature Norming Features

- Cree and McRae’s (2003)
  - Asked participants to list features of 541 words.
  - The features that participants produce are a verbalization of actively recalled semantic knowledge.
  - Eg. *House* is used for living, is warm, is made of brick, etc.

Example of Features

Concept	**Feature**	**BR Encoding**	**WB Encoding**
House	Made by humans	Encyclopedic	Origin
	Used for living in	Function	Function
	Is warm	Tactile	Internal surface property
	Is large	Visual-form and surface properties	External surface property
	Made of brick	Visual-form and surface properties	Made of
	Has rooms	Visual-form and surface properties	Internal component
	Has windows	Visual-form and surface properties	External component
Cow	Lives on farms	Encyclopedic	Location
	Eaten as meat	Function	Function
	Is smelly	Smell	External surface property
	Moos	Sound	Entity behavior
	An animal	Taxonomic	Superordinate
	Is white	Visual-color	External surface property
	Has 4 legs	Visual-form and surface properties	External component
	Eats grass	Visual-motion	Entity behavior
	Produces manure	Visual-motion	Entity behavior

2A. Feature Norming Features

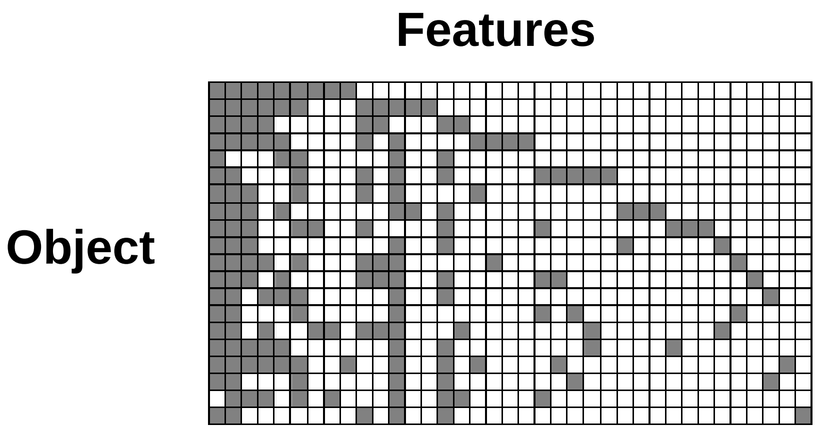
- Code participants' behavioral response for the modifier noun, the head noun, and the compound noun.
- Then, we could check
  - If the compound noun inherits features more from the modifier or head noun?
  - If the pattern differs for the two types of interpretations?

2B. Infinite Latent Semantic Models

- Model the semantic representation as a hidden variable in a generative probabilistic model.
- The basic proposition of the model is that
  - There can be an infinite list of features (or semantic components) associated with a concept.
  - Only a subset is actively recalled during any given task (context-dependent).
  - A set of latent indicator variables is introduced to indicate whether a feature is actively recalled.

Griffiths & Ghahramani (2005)

- Infinite latent semantic feature model (ILFM; Griffiths & Ghahramani, 2005)
  - Assumes a non-parametric Indian Buffet prior to the binary feature vector and models neural activation with a linear Gaussian model.

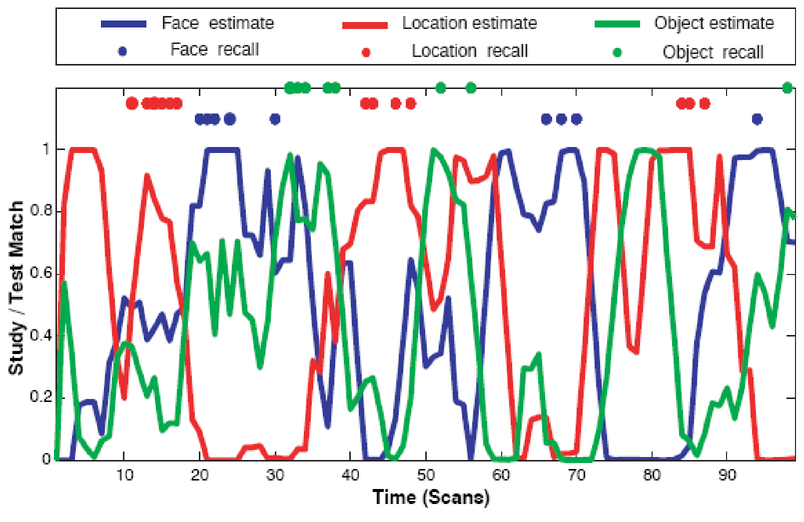


2B. Infinite Latent Feature Models

- Learn the infinite latent feature models for both noun and phrases.
- Then, we can check
  - If the compound noun share more latent feature with the modifier or head noun?
  - If the pattern differs for the two types of interpretations?

3. Explore Time-Series Data

- Polyn et al. (2005) analyzed the time-series data of fMRI. They showed that category-specific brain activity during a free-recall period correlated more with brain activity of matching categories during a prior study period.



3. Explore Time-Series Data

- We can adopt an approach similar to Polyn et al. (2005) and correlate the brain activity of the noun phrases to the brain activity of each word in the phrase.
  - Do this for each time slice and see if the pattern changes across time.

Timetable

Task	**Time**
Thesis Proposal	Jan, 2010
60 words experiment	Complete
Adjective-noun experiment	Complete
Noun-noun experiment	Dec 2009 - Feb, 2010
Explore feature norms	Feb, 2010 (already started)
Explore latent feature models	Mar, 2010 (already started)
Explore time series data	Apr, 2010 (already started)
Thesis Writing	May, 2010
Thesis Defense	June, 2010

Questions?

- Kai-min Kevin Chang
  - kaimin.chang@gmail.com
  - http://www.cs.cmu.edu/~kkchang
  - Carnegie Mellon University
  - Center for Cognitive Brain Imaging