Cognitive Models of Semantic Representation

Lecture 14: How Humans Represent Meaning

PSYC 51.07: Models of Language and Communication - Week 4

Winter 2026

Today's Lecture 📋

1. 🧠 **Human vs. Computational Semantics**
2. 🔬 **Cognitive Theories of Meaning**
3. 🌐 **Embodied & Grounded Cognition**
4. 📊 **Semantic Similarity: What Does It Mean?**
5. 🔍 **Empirical Evidence from Cognitive Science**
6. 🤖 **Bridging the Gap: Models Meet Minds**

Goal: Understand what computational models are really learning

The Fundamental Question 🤔

What does "meaning" actually mean?

Computational View:

- Vectors in high-dimensional space
- Learned from text co-occurrence
- Distributional patterns
- Statistical relationships
- "You shall know a word by the company it keeps"

Human View:

- Sensory experiences
- Emotional associations

How Do Humans Represent Meaning? 🧠

Example: The word "coffee"

What comes to YOUR mind?

- **Visual:** brown liquid, mug, steam, beans
- **Olfactory:** aroma, roasted smell
- **Gustatory:** bitter taste, smooth texture
- **Tactile:** hot, warm cup, liquid
- **Auditory:** brewing sounds, pouring
- **Motor:** lifting cup, drinking motion
- **Contextual:** morning routine, work, café
- **Emotional:** comfort, alertness, pleasure
- **Social:** conversations, meetings

The Symbol Grounding Problem 🔗

Harnad (1990): Can symbols have intrinsic meaning?

The Chinese Room (Searle, 1980):

```
Input:
         ↓
[Rule Book: If see
             output      ]
         ↓
Output:

Correct response! But no
understanding of Chinese.
```

Analogy to LLMs:

Embodied Cognition Theory 🏃

Meaning arises from bodily experience and sensorimotor interaction

Key Principles:

1. **Embodiment:** Cognition shaped by body
2. **Situatedness:** Meaning context-dependent
3. **Enactivism:** Knowing through doing
4. **Grounding:** Concepts tied to perception/action

Neuroscience Evidence:

```
Reading "kick the ball":
→ Motor cortex activates
→ Leg area specifically!

```

The Distributional Hypothesis Revisited 📖

"You shall know a word by the company it keeps"

--- J.R. Firth (1957)

Strong version: Word meaning IS distributional patterns

Weak version: Distributional patterns REFLECT meaning

Supports:

- Works remarkably well in practice
- Captures semantic similarity
- Enables analogical reasoning
- Scales to huge vocabularies
- Unsupervised learning

What Does "Semantic Similarity" Really Mean? 🔍

Different types of similarity:

Taxonomic (IS-A):

```
dog ↔ cat: Both are animals
           Similarity: HIGH
```

Thematic (GOES-WITH):

```
dog ↔ leash: Co-occur in events
             Relatedness: HIGH
             Similarity: LOW!
```

Test Yourself:

Semantic Projection: Recovering Human Knowledge 🔬

Grand et al. (2022): Can we extract human-like features from embeddings?

The Experiment:

1. Collect human ratings on perceptual features
    - Is it edible?

- Is it heavy?
- Is it alive?
- Can you hold it?

2. Train linear projections from word embeddings
3. Test if embeddings predict human ratings

Multimodal Models: Bridging the Gap 🌉

Combining language with perception

Vision-Language Models:

- CLIP (OpenAI)
- ALIGN (Google)
- Flamingo (DeepMind)
- GPT-4V (OpenAI)

Key Idea:

- Learn joint embedding space
- Text and images map to same space
- Enables cross-modal understanding

Conceptual Spaces Theory 🌌

Gärdenfors (2000): Meaning as geometry

Key Ideas:

- Concepts represented in quality dimensions
- Dimensions: color, size, temperature, etc.
- Each dimension has a metric
- Concepts are regions in space
- Similarity = geometric proximity

Example: Colors

- Hue, saturation, brightness
- Natural categories (red, blue, green)

Lexical Semantic Theories 📚

How do linguists think about word meaning?

1. **Feature-Based:**
    - Words = bundles of features
    - bachelor = [+human, +male, +adult, -married]
    - Compositional
    - Logical
2. **Prototype Theory:**
    - Categories have best examples
    - Robin is a prototypical bird
    - Penguin is peripheral

Empirical Evidence from Neuroscience 🧬

What does the brain tell us about semantic representation?

fMRI Studies:

- Can predict brain activity from word embeddings
- Semantic information distributed across cortex
- Different regions for different features
- Temporal lobe: objects
- Motor cortex: actions
- Visual cortex: visual features

Findings:

- Word2Vec correlates with neural patterns

The "Stochastic Parrots" Debate 🦜

Bender et al. (2021): On the Dangers of Stochastic Parrots

The Argument:

- LLMs learn form, not meaning
- "Stochastic parrots" - repeating patterns
- No understanding of world
- No communicative intent
- Risk: Mistaking fluency for understanding

Evidence:

- Fail on simple reasoning
- Sensitive to phrasing

Common Sense Reasoning 🤯

What humans know but models don't

Physical Intuition Failures:

```
Q: "Can you fit an elephant
    in a refrigerator?"

GPT-3: "Yes, if you open the
       door wide enough..."
```

Winograd Schema (reasoning):

```
"The trophy doesn't fit in the
brown suitcase because it is
too [small/large]."

```

Compositionality: Phrases and Sentences 🧩

How do we combine word meanings?

The Problem:

```
# Vector math doesn't work!
vec("hot") + vec("dog") ≠ vec("hot dog")

# "hot dog" = food item
# "hot" + "dog" = warm canine

# Same issue:
vec("red") + vec("herring") ≠ vec("red herring")
# red herring = distraction, not a fish!
```

Non-compositional Phrases:

- Kick the bucket (= die)

Grand Discussion 💬

Do large language models "understand" language?

Arguments FOR:

- Solve complex tasks
- Generalize to new domains
- Show emergent capabilities
- Capture linguistic structure
- Pragmatic criterion: if it works...
- Maybe understanding = prediction
- Human understanding also imperfect

"The question is not whether machines think, but whether they behave intelligently" - Turing

Summary 🎯

What we learned today:

1. **Symbol Grounding:** Computational models lack perceptual grounding
2. **Embodied Cognition:** Human meaning tied to bodily experience
3. **Distributional Semantics:** Powerful but incomplete theory
4. **Semantic Similarity:** Multiple types, models capture some
5. **Empirical Evidence:** Models align with neural patterns but miss multimodality
6. **Common Sense:** Models struggle with physical and social reasoning
7. **Compositionality:** Non-literal language remains challenging

The gap between computation and cognition remains, but we're making progress!

Key References 📚

Foundational Papers:

- Searle, J. (1980). "Minds, Brains, and Programs"
- Harnad, S. (1990). "The Symbol Grounding Problem"
- Barsalou, L. (2008). "Grounded Cognition"
- Gärdenfors, P. (2000). "Conceptual Spaces: The Geometry of Thought"

Distributional Semantics:

- Firth, J.R. (1957). "A Synopsis of Linguistic Theory"
- Boleda, G. (2020). "Distributional Semantics and Linguistic Theory"
- Hill et al. (2015). "SimLex-999"

Critical Perspectives:

Questions? 🙋

Next Week:

Advanced Topics in Language Models

Scaling, emergent abilities, and the future of NLP