CSCI 544, Lecture 17: Semantics

Ron Artstein

2022-10-20

Administrative notes

Coding Assignment 3 will be graded soon

Coding Assignment 4 due October 25 **Submit early, submit often!**

Presentation:

Due Date	Task
October 18	Article selection
October 27	Presentation slides
November 1–10	Presentations

Project:

Due Date	Task
November 3	Project status report
Nov 29/Dec 1	Poster presentations (in class)
December 1	Final report
December 3	Self-evaluation and peer grading

Semantics

semantics *noun* **1** the study of meanings (Merriam-Webster)

Large field, multiple NLP tasks.

Generally fall under two broad categories:

Lexical semantics Meanings of words, word-like items (multi-word expressions)

Compositional semantics How word meanings are put together

Model-theoretic semantics

Generally associated with the work of Richard Montague

Map human language expressions to logical representation instantiated in a model

human language → logical representation ↔ model

Model-theoretic semantics

Generally associated with the work of Richard Montague

Map human language expressions to logical representation instantiated in a model

human language $\longrightarrow$ logical representation $\longleftrightarrow$ model

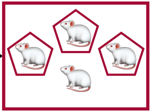
A mouse lives in every building

$\downarrow$

$\exists x(\text{mouse}(x) \wedge \forall y(\text{building}(y) \rightarrow x \text{ lives in } y))$

$\forall y(\text{building}(y) \rightarrow \exists x(\text{mouse}(x) \wedge x \text{ lives in } y))$

$\longleftrightarrow$



Scope ambiguity

Generalized quantifiers

From the meanings of the sentences, we can derive meanings for phrases and words.

A mouse lives in every building

$$\begin{cases} \exists x(\text{mouse}(x) \wedge \forall y(\text{building}(y) \rightarrow x \text{ lives in } y)) \\ \forall y(\text{building}(y) \rightarrow \exists x(\text{mouse}(x) \wedge x \text{ lives in } y)) \end{cases}$$

Generalized quantifiers

A mouse	$\lambda P.\exists x(\text{mouse}(x) \wedge P(x))$
Every building	$\lambda Q.\forall y(\text{building}(y) \rightarrow Q(y))$

Determiners

A	$\lambda P \lambda Q.\exists x(P(x) \wedge Q(x))$
Every	$\lambda P \lambda Q.\forall x(P(x) \rightarrow Q(x))$

More scope ambiguities

Negation

- All that glitters isn't gold
- All doors won't open

Intension

- John finds a unicorn (unambiguous)
- John seeks a unicorn
  - $\exists x(\text{unicorn}(x) \wedge \text{John seeks } x)$
  - John seeks ($\hat{}$unicorn)

Event semantics

Pat ate the soup with a spoon at the restaurant on Sunday

"Flat" representation (Davidson, Hobbs)

e: agent(e) = Pat
theme(e) = soup
manner(e) = with a spoon
location(e) = the restaurant
time(e) = Sunday
eat(e)

PropBank (proposition bank) roles

Argument roles: Arg0 (~agent), Arg1 (~theme), Arg2. . .
Adjunct roles: manner, instrument, . . .

Semantic role labeling

Identify the roles of each element in the sentence

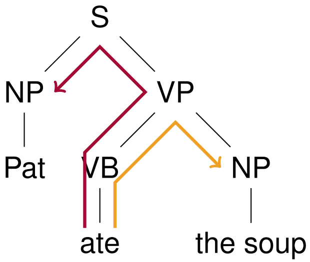
- Identify the arguments ("markables")
- Identify the roles

Gildea and Palmer use features and classifier

Phrase Type	Parse Tree Path	Position	Voice	Head Word

Main question: to what extent are parse features useful?

Parse tree paths



Path to subject: Ate $\rightarrow$ Pat $=$ VB $\uparrow$ VP $\uparrow$ S $\downarrow$ NP

Path to object: Ate $\rightarrow$ The Soup $=$ VB $\uparrow$ VP $\downarrow$ NP

Lexical semantics

What is a word? — Dictionary entry

Homonyms: two (or more) words with the same spelling and pronunciation

bat^1  **bat^2**  **pitcher^1**  **pitcher^2** 

Polysemes: two (or more) senses of the same word

newspaper 1.  **2.** 

Is the distinction between homonymy and polysemy psychologically real?

Frazier and Rayner, Journal of Memory and Language 29(2):181–200, 1990

Distributional properties of words

"If we consider **oculist** and **eye-doctor** we find that, as our corpus of actually-occurring utterances grows, these two occur in almost the same environments… In contrast, there are many sentence environments in which **oculist** occurs but **lawyer** does not…

Zellig S. Harris, Distributional Structure,
Word 10(2-3):146–162, 1954

Association between words and contexts

$$
\begin{array}{c|ccccccccc c}
 & c1 & c2 & c3 & c4 & c5 & c6 & c7 & c8 & c9 & \cdots \\
\hline
w1 & & & & & & & \bullet & & & \\
w2 & & & & & & & & & & \\
w3 & & \bullet & & & & & & & & \\
w4 & & & & & & & \bullet & & & \\
w5 & & & \bullet & & & & & & & \\
w6 & & & & & & & & & & \\
\vdots & & & & & & & & & &
\end{array}
$$

Word $\times$ context matrix

PPMI: positive pointwise mutual information

$$\max\left(0, \log \frac{\hat{P}(w,c)}{\hat{P}(w)\hat{P}(c)}\right)$$

Sparse

Dimensionality reduction

SVD (Singular Value Decomposition)

- Decompose sparse matrix
  - Reconstruct in lower dimension
- Used in Latent Semantic Analysis, etc.

Neural Language Models

- By-product: word embedding matrix

Omer Levy, Yoav Goldberg and Ido Dagan: Improving Distributional Similarity with Lessons Learned from Word Embeddings, TACL 2015

Why are neural models more successful?

Designing a low-dimension matrix

Preprocessing: What counts as a “context”?

- What constitutes the columns of the sparse matrix?

Association: How are words counted in contexts?

- What constitutes the cells of the sparse matrix?

Postprocessing:

- How are the values of the dense matrix optimized?

All of these are **hyperparameters** which determine the usefulness of the word embeddings.

Comparing traditional and neural embeddings

Multiple embeddings, varying:

- Hyperparameters
- Factorization algorithm
- Amount of data

Hyperparameters appear to have the largest impact.