Chapter NLP:VII

VII. Semantics

- Semantic Structures
- Natural Language Understanding

Semantic Structures

Semantics

“The man sighed.
It’s raining cats and dogs, he felt.”

Semantic Structures
Semantics

Definition 1 (Semantics)

The study of how meaning is communicated via language. In computer science, how to use algorithms to models, understand, and react on text, or generate text that communicates a “meaning” to humans.

Psychology: Meaning is how the mind makes sense of the world.

Philosophy: Meaning is the truth, equivalence, presupposition, and inferences of statements.

Natural language	Logic
“All men are mortal.”	$\forall x : man(x) \rightarrow mortal(x)$
“Socrates is a man.”	$man(Socrates)$
Socrates is mortal.	$mortal(Socrates)$

Linguistics: Meaning is how humans share information and express feelings or intentions via verbal, paralinguistic (e.g. intonation) and non-verbal (e.g. gestures) language.

Remarks

- Semantics stems from the ancient Greek *semantikos* (relating to signs as in symptoms of a disease).
- Today: The study of signs (road signs, disease symptoms, animal tracks) is called semiotics or semiology, while the study of *linguistic signs* (words, phrases, sentences, utterances) is called semantics.

Semantic Structures

Lexical Semantics[OxfordRE Linguistics]

Lexical semantics describes the relation between meaning and form of words:

Semasiology Which meaning can be assumed from word form. Semantic relations describe difference in meaning with identical form.

Polysemy:	`mouse`	animal *vs.* input device
Specialization:	`corn`	wheat *vs.* oats
Generalization:	`moon`	of earth *vs.* any satellite
Metaphor:	`desktop`	of a desk *vs.* on screen

Onomasiology Which relations exist between the concepts. Which forms exists for the concepts.

Lexical Relations:	`mouse` *vs.* `rodent`
Lexical Fields	
Frames	
Distributional relations	
...	

Semantic Structures

Lexical Semantics: Word Senses

Meaning is traditionally represented with symbols called word sense.

- Senses are often denominated by superscript^1.
- The informal description of a sense is it's gloss,
- Senses can be modeled explixitly with their lexical relations or by componential analysis, or intrinsically via distributional semantics.

Sense description by gloss:

Sense	Gloss
Bank^1:	financial institution that accepts, deposits, and lends money.
Bank^2:	sloping land (especially the slope beside a body of water).
Red^1:	the color of blood or a ruby.
Blood^1:	the red liquid that circulates in the vains of animals.

Semantic Structures

Lexical Semantics: Lexical Relations (selection)

Polysemy Same lexeme, different sense.

The semantic relations are subtypes of polysemy.

Synonym Different sense but similar meaning.

`couch` $\longleftrightarrow$ `sofa` `big` $\longleftrightarrow$ `large`

Antonymy Opposite meaning.

`long` $\longleftrightarrow$ `short`

Hyponymy/Hypernomy One sense is less/more specific. Also called IS-A

`car` $\xrightarrow{\text{Hyponym}}$ `vehicle`

`car` $\xrightarrow{\text{Hypernym}}$ `ID.3`

Meronym/Holonym The part-whole relation.

`wheel` $\xrightarrow{\text{Meronym}}$ `car`

`car` $\xrightarrow{\text{Holonym}}$ `wheel`

Relations are defined over senses, not lexemes:

Synonymous `big`^{1} `plane` $\longleftrightarrow$ `large`^{1} `plane`

Not synonymous `big`^{2} `sister` $\longleftrightarrow$ `large`^{1} `sister`

Semantic Structures

Lexical Semantics: WordNet

The largest English database of word senses is WordNet.[Fellbaum, 1998]

- WordNet has entries for lemmas.
- An entry has 1 or more synsets: sets of near-synonymous **senses**. Synsets represent concepts of meaning.
- Synsets and lemmas are split by word class: Noun, Verb, Adjective, Adveb.
- Each synset has a topic (supersense), assigned from a per word class fixed set.

Entry for the lemma `bass (Noun)`:

Synset	**Topic**	**Gloss**
$bass^{1}$	attribute	the lowest part of the musical range
$bass^{2}$	animal	edible marine and freshwater spiny-finned fishes
sea bass, $bass^{2}$	food	the lean flesh of a saltwater fish of the family Serranidae
$bass,^{1}$ bass part	communication	part the lowest part in polyphonic music
$bass,^{1}$ basso	person	an adult male singer with the lowest voice

Semantic Structures

Lexical Semantics: WordNet

The largest English database of word senses is WordNet.[Fellbaum, 1998]

- WordNet has entries for lemmas.
- An entry has 1 or more synsets: sets of near-synonymous **senses**. Synsets represent concepts of meaning.
- Synsets and lemmas are split by word class: Noun, Verb, Adjective, Adveb.
- Each synset has a topic (supersense), assigned from a per word class fixed set.

Entry for the lemma `ride (Verb)`:

Synset	**Supersense**	**Gloss**
sit, ride	motion	sit and travel on the back of animal
sit, ride	motion	be carried or travel on or in a vehicle
tease, ride, rally, bait, rag, twit, tantalize, razz, taunt, cod	communication	harass with persistent criticism or carping
ride	stative	continue undisturbed and without interference ?Let it ride?

Remarks:

- WordNet synsets are separated by word class and do not overlap:
  - Nouns: 117,798 lemmas (avg. 1.23 senses)
  - Verbs: 11,529 lemmas (avg. 2.16 senses)
  - Adjectives: 22,479 lemmas
  - Adverbs: 4,481 lemmas
- In WordNet, semantic relations are encoded as one lexical relation (Polysemy) with 3 additional subrelations:
  - Constructional/structured polysemy: Same sense entry refers to different entitires
    `(The) Times` printed paper *vs.* the news contained in it *vs.* the organization
  - Sense extension polysemy: Derives a new synset from an old sense
    `chicken` animal *vs.* meat of animal
  - Homonymy: Same sense, very different meaning
    `bank` river bank *vs.* financial bank

Semantic Structures
Word Sense Disambiguation

Word Sense Disambiguation (WSD) is the task of assigning to each word in a text the correct sense from a sense lexicon.

- WSD is similar to tagging, but more difficult.
  There are several classes for each word.
- To disambiguate a small set of words, classification works.
- Using the most frequent sense every time is a stong baseline. Disambiguate medical terms in lab reports
- There are several datasets (semantic concordance) where each word is annotated with its sense.
  - SEMCOR (Brown Corpus, English), SENSEVAL-3 (English), [Vossen et al., 2011] (Dutch), [Heinrich et al., 2012] (German)
  - Annotated are word class and sense id for open-class words.

Example annotations from SEMCOR:

You will find$^{9}_{v}$ that avocado$^{1}_{n}$ is$^{1}_{v}$ unlike$^{1}_{j}$ other$^{1}_{j}$ fruit$^{1}_{n}$ you have ever$^{1}_{r}$ tasted$^{2}_{v}$

Semantic Structures

Word Sense Disambiguation: Lesk

Idea: The context of a word should overlap with the words in the gloss of its sense.

- Does not need training data.
- Easy to apply to new or low-resource languages.
- Glosses can easily be extended with (annotated) examples.

Sense	**Gloss**
bank[1]	A financial institution that accepts deposits and channels the money into lending activities
bank[2]	sloping land (especially the slope beside a body of water)

w_{i-4}	w_{i-3}	w_{i-2}	w_{i-1}	w_i	w_{i+1}	w_{i+2}	w_{i+3}	w_{i+4}	w_{i+5}
`We`	`guarantee`	`that`	`your`	`bank`	`deposits`	`will`	`cover`	`future`	`costs.`

Semantic Structures

Word Sense Disambiguation: Lesk

SIMPLIFIED LESK:

For the word w_i in a sequence $w_1, \ldots, w_{i-k}, \ldots, w_i, \ldots, w_{i+k}, \ldots, w_n$ with window size k and glosses $G_{w_i} = \{g_{w_i,1}, \ldots, g_{w_i,j}\}$:

1. Remove stopwords from and lemmatize the context window $v_i := (w_{i-k}, \ldots, w_{i-1}, w_{i+1}, \ldots, w_{i+k})$ and all glosses $g_j \in G_{w_i}$.
2. Vectorize v_i and all $g_j \in G_{w_i}$.
3. Disambiguate w_i by the lowest cosine between context and gloss vectors.

- LESK can be improved using tf·idf-weighted vectors or any other (semantic) similarity measure.
- Gloss vectors can be pre-computed.

Semantic Structures

Word Sense Disambiguation: Classification

Idea: Classify the sense with sliding-window features (cf. sequence tagging).

Example features:

1. Words (lemmas/stems) in the context window
2. Part-of-speech tags for each word in the window
3. n-grams
4. Weighted average of the word embeddings

Example:

w_{i-4}	w_{i-3}	w_{i-2}	w_{i-1}	w_i	w_{i+1}	w_{i+2}	w_{i+3}	w_{i+4}	w_{i+5}
`We`	`guarantee`	`that`	`your`	`bank`	`deposits`	`will`	`cover`	`future`	`costs.`

Features for w_i with $k = 2$:

w_{i-2}	POS_{i-2}	w_{i-1}	POS_{i-1}	w_{i+1}	POS_{i+1}	w_{i+2}	POS_{i+2}	w_{i-2}^{i-1}	w_{i+1}^{i+2}
that	IN	your	PRP	deposits	NN	will	MD	that your	deposits will

Remarks:

- The state of the art for WSD uses contextualized word embeddings.
- Word embeddings are identical for each sense of the same word form. They are not intrinsically useful for WSD.
- However, since word vector spaces embed semantic similarity, they can solve tasks like lexical substitution with a simple nearest neighbor search.
- A large language model like BERT produces contextualized word embeddings more or less as a by-product. Here, the same lexeme can have different vectors, depending on its context words. BERT solves WSD extremely well if there are vectors for each sense.
- `Word Sense Induction` tries to create lexicons like WordNet automatically by clustering the embedding space. This produces a synset collection with context vectors for each sense (the mean vector of each cluster) in an unsupervised fashion.

Semantic Structures

Lexical Substitution

Lexical substitution tasks are subtask of WSD.

- Classic lexical subsitution looks for one or more semantically similar replacement for certain words.

  `My favorite thing about her is her straightforward honesty`

  → `My favorite thing about her is her sincere/genuine/frank honesty`

- Lexical simplification looks for a easier to understand byt semantically similar replacement.

  `John composed these verses` → `John wrote these poems`

- Lexical substitution often uses the same techiniques as WSD but does not require a lexicon.

Semantic Structures

Multi-Word Expressions

Multi-word expressions (MWE) can function as singular lexical units.

MWE semantics can be

1. compositional,

   `driving instructor`

   `argumentation quality assessment`

2. idiomatic,

   `vice versa`

   `kick the bucket`

3. or in-between.

   `Long time no see`

Semantic Structures

Limitations of Lexical Semantics [NLP:III 64 ff.]

Lexical semantics faces two essential problems:

1. Lexemes do not encode semantic similarity or dissimilarity, only identity.

   Same difference: `cat` $\leftrightarrow$ `feline` and `plane` $\leftrightarrow$ `pony`

2. Synsets are (computationally) unpractical.
   Synset relations are semantically shallow.

   Same difference: `fake` $\leftrightarrow$ `falsify` $\leftrightarrow$ `misrepresent`

   Same relationship: `long` $\leftrightarrow$ `short` and `Heaven` $\leftrightarrow$ `Hell`

$\rightarrow$ A semantically rich encoding represents words by their dimensions of meaning.

Semantic Structures

Compositional Semantics

Compositional semantics describes the meaning of phrases or sentences as a composition of words.

Collocations. Words that frequently appear together have a combined meaning.

Relations. Semantic Relations between entities from the world.

Temporal Relations describing courses of events.

Operators. Quantifiers: Indicating the quantity of objects.

Hedges: Lessening the impact of a proposition.

Negation: Inverting an adjective, predicate, or similar.

Semantic Structures

Compositional Semantics: Semantic Relations

Semantic relations are word compositions that capture relational predicates using arguments. Who, what, where, when, with what, why, and how?

Common relation types:

- **Binary relations.** Relations with two arguments.

  Relation: `founded(organization, date)`

  Google was established in 1998 → `founded(Google, 1998)`

- **Events.** Relations with multiple arguments, possibly nested relations.

  Event: `reads(agent, theme, date, time, location)`
  Relation: `origin(theme, author)`
  Max reads a book in the garden on Monday at midnight. It is from Shakespeare.
  → `reads(Max, book, Monday, midnight, garden)`
  $\wedge$ `origin(book, Shakespeare)`

Semantic Structures

Compositional Semantics: Operators

Linguistic operator modify the meaning of a span of text (the scope).

Common operators:

- Quantifiers. Scope (mostly) resolved by syntax.

  `Every student reads some book.`

  $\forall x(student(x) \wedge \exists y(book(y) \wedge read(x,y)))$ vs. $\exists y(book(y) \wedge \forall x(student(x) \wedge read(x,y)))$

- Hedges. Scope resolved by syntax.

  `I worked only tonight.` *vs.* `I only worked tonight.`

- Pronouns. Complex resolution of scope.

  `A person got run over. He really got angry about it.`

- Negations. Complex resolution of scope.

  `It’s not good manners I don’t care about.`

Semantic Structures

Compositional Semantics: Collocation

Collocation are sequences of words that co-occur (uncommonly) often.

`do homework`	As frequent as `homework` alone.
`in my opinion`	Most common phrase including `opinion`.
`vice versa`	More common together than individually (idioms).

- Collocations are often less ambiguous than the words taken in isolation.

  `heavy smoker` `smoker` clarifies which of the 30 senses of `heavy` is meant

- Multi-word expressions are a particular type of collocations.

Remarks:

- A statistical approach to extract collocation from a corpus is cooccurrence significance on left/right neighbor windows.
- Additionally one could use n-grams to capture collocation within a corpus as feature for machine learning, which covers phrase or compositional semantics in a simple feature.

Semantic Structures

Compositional Semantics: Componential Analysis

Idea: Define the dimensions of meaning and score the words in each dimension. (COMPONENTIAL ANALYSIS)

Dim	**Component**	**Words**		
		chair	*couch*	*bean bag*
s_1	for seating	1	1	1
s_2	one person	1	0	1
s_3	with legs	1	1	0
s_4	with back	1	1	0
s_5	with armrest	0	1	0
s_6	of rigid material	1	1	0

Computational Problems:

- Dimensions are hand crafted.
- Dimensions are different for each lexical field. semantically similar words
- Vectors become very sparse when combining many fields.

Semantic Structures

Frame Semantics: Semantic Roles

Semantic roles describe the roles of the arguments of a predicate.

`She saw Max.`

Deep Role. $\exists e, y$ Observe(e) $\wedge$ Observer(e, `She`) $\wedge$ ObservedObject(e,y) $\wedge$ Max(y)

Thematic Role. `[`AGENT `She] saw [`THEME `Max].`

Semantic Structures

Frame Semantics: Semantic Roles

Semantic roles describe the roles of the arguments of a predicate.

`She saw Max.`

Deep Role. $\exists e, y$ Observe(e) $\wedge$ Observer(e, $\texttt{She}$) $\wedge$ ObservedObject(e,y) $\wedge$ Max(y)

Thematic Role. `[`AGENT `She] saw [`THEME `Max].`

Common Thematic Roles:

AGENT	Who causes the event
EXPERIENCER	Who experiences an event
FORCE	What causes the event
THEME	Who is (directly) affected by the event
INSTRUMENT	What instrument is used in the event
BENEFICIARY	Who benefits from the event
...	