Word Senses and WordNet and Semantic Role Labeling

Natalie Parde

UIC CS 421

This Week's Topics

Tuesday

- Word Senses
- WordNet
- Word Sense Disambiguation

Thursday

- Semantic Roles
- Semantic Role Labeling
- Selectional Preferences

This Week's Topics

Tuesday

- Word Senses
- WordNet
- Word Sense Disambiguation

Thursday

- Semantic Roles
- Semantic Role Labeling
- Selectional Preferences

Words can carry many meanings.

- The different possible meanings for a word are its **senses**
- For example:
  - Book_1: To reserve something
  - Book_2: A large written source of fiction or non-fiction text
  - Book_3: To move quickly
- Word senses can be represented in numerous ways

Glosses

- Dictionaries or thesauruses often provide definitions for each sense of a word, referred to as **glosses**
    - Not a formal meaning representation!
    - Written to facilitate human understanding of the senses a word may take
    - May be circular
        - Direct self-reference (e.g., “Right: Located nearer the right hand”)
        - Implicit self-reference (e.g., “Left: Located nearer to the side opposite the right”)
        - Complementary external reference (e.g., “Red: The color of a ruby” and “Ruby: A red gemstone”)

Glosses

- Even if glosses aren't meaning representations themselves, they can still be useful for computationally modeling word senses
  - Glosses are sentences
    - Convenient input for representation learning
  - Glosses are often accompanied by example sentences
    - Additional useful data

Dictionary-Based Sense Definitions

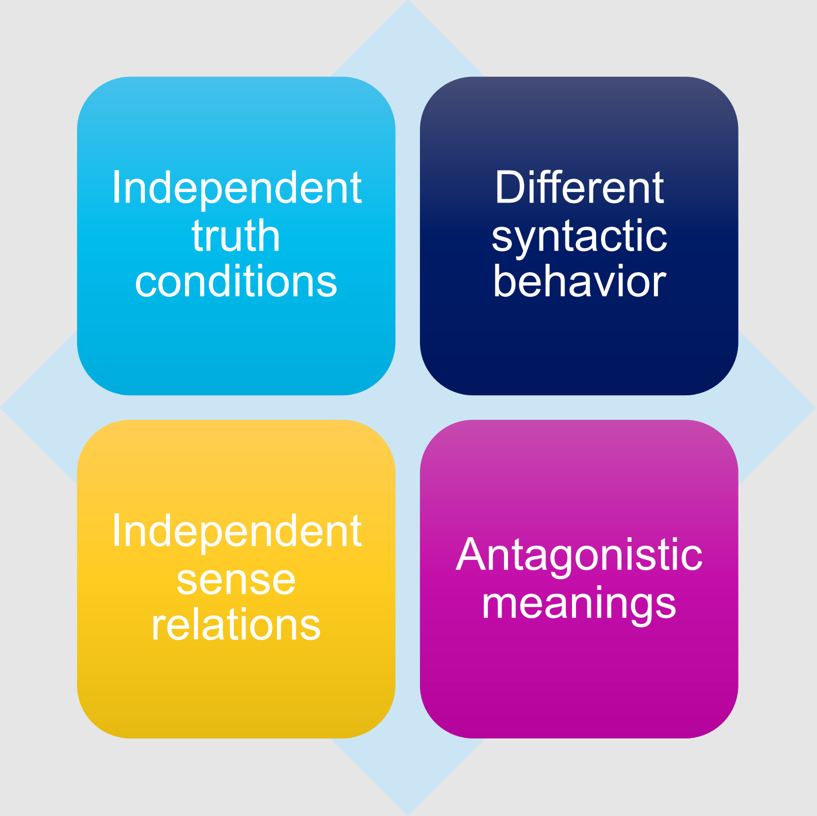
- Senses can be defined through their relationship with other senses
- Given a large database of senses and the relations between them, we can leverage these associations to perform semantic tasks

Words with numerous senses are **polysemous**.

- **Polysemy:** The phenomenon in which a single word is associated with two or more distinct senses
- There is no limit to how many senses a word can have!
- Sense distinctions vary depending on the dictionary:
  - Some dictionaries represent very fine-grained distinctions as different senses
  - Computational resources usually focus on broader, more coarse-grained sense categories

How can we distinguish between senses?

- Word embeddings offer continuous, high-dimensional word representations that aren’t easily discretized into sentences
  - Contextual word embeddings produce a different representation for each unique use of a word
- Dictionaries separate words into senses based on predetermined criteria



Practical Technique for Determining Sense Distinction

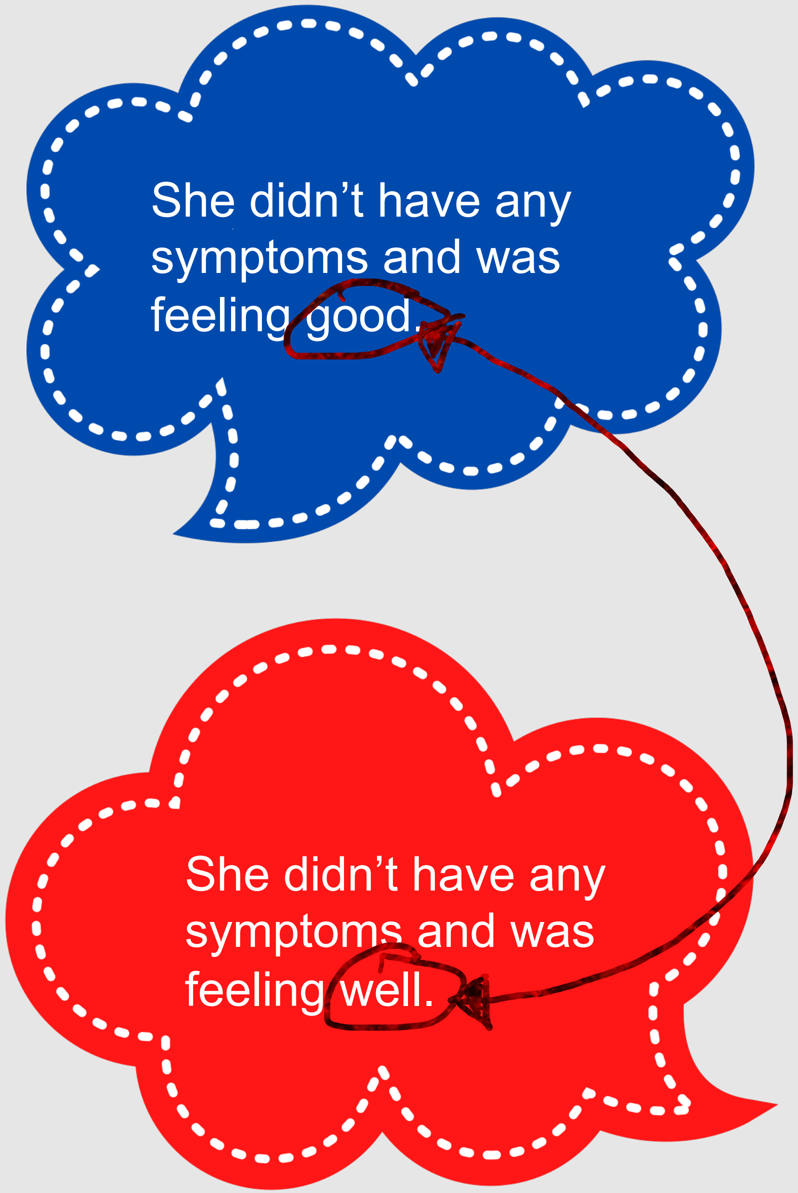
- Conjoin two uses of a word in a sentence
- For example:
  - Which of those flights serve ice cream?
  - Does American Airlines serve Chicago?
  - Does American Airlines serve ice cream and Chicago?
- If you observe that this creates a **zeugma** (a conjunction of antagonistic uses of the same word), consider these as distinct senses

How do word senses relate to one another?

- Many types of relations can exist between word senses
- Particularly useful for NLP purposes:
  - Synonymy
  - Antonymy
  - Hypernymy

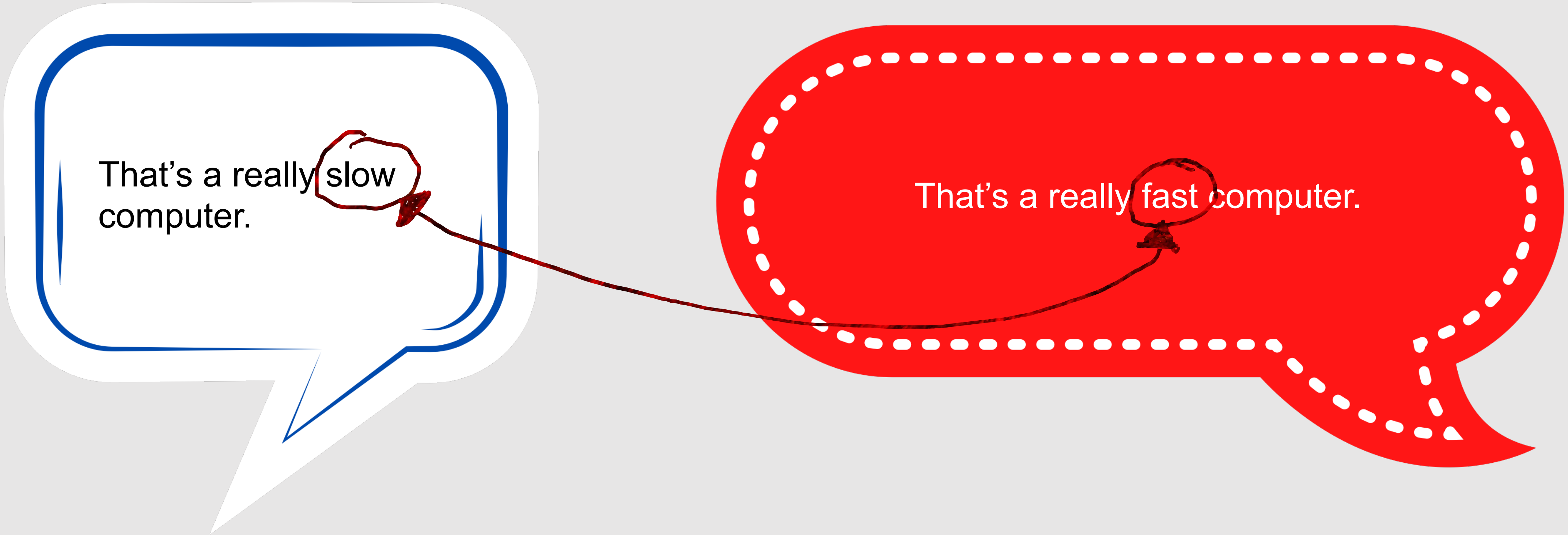
Synonymy

- Occurs when two word senses are highly similar to one another
  - Substituting one for another should convey essentially the same meaning
- *All* senses for both words do not need to be highly similar



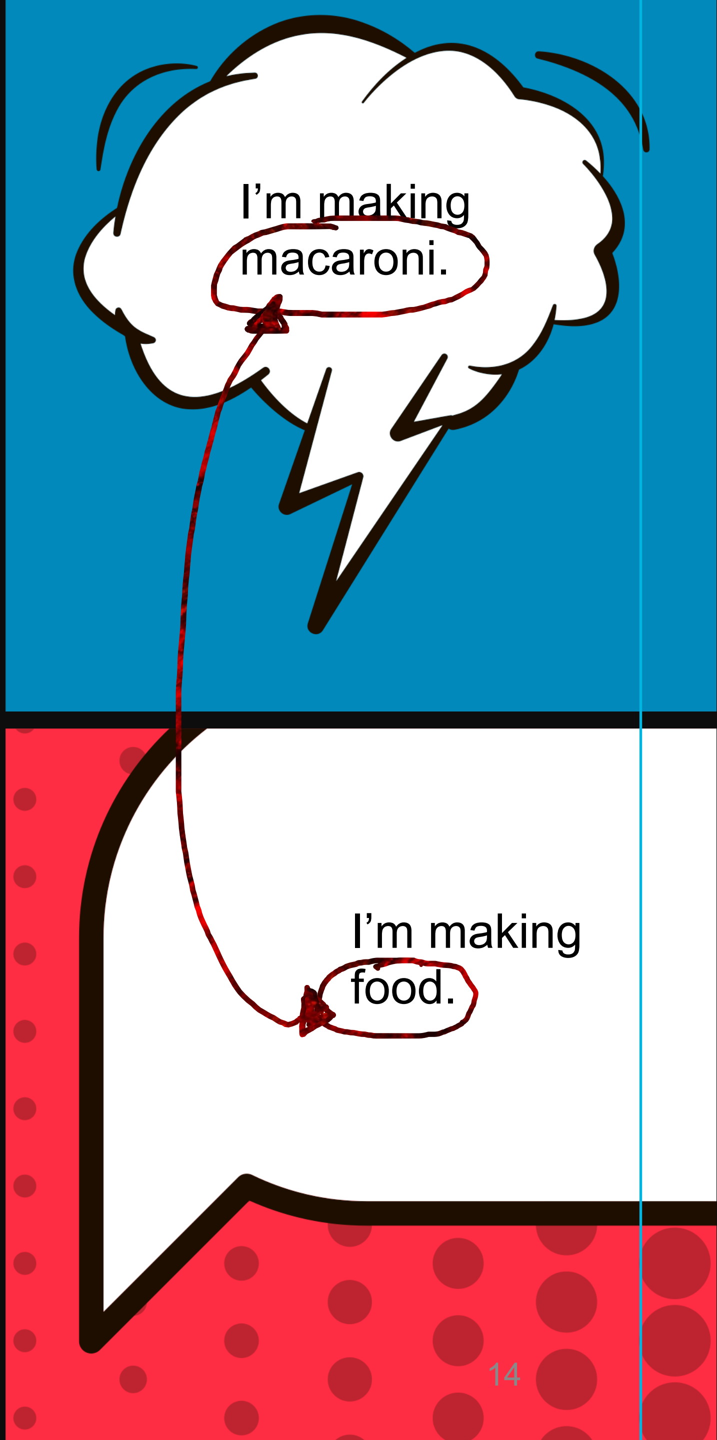
Antonymy

- Occurs when two word senses convey opposite meaning to one another
- The word senses should otherwise be interchangeable in similar contexts



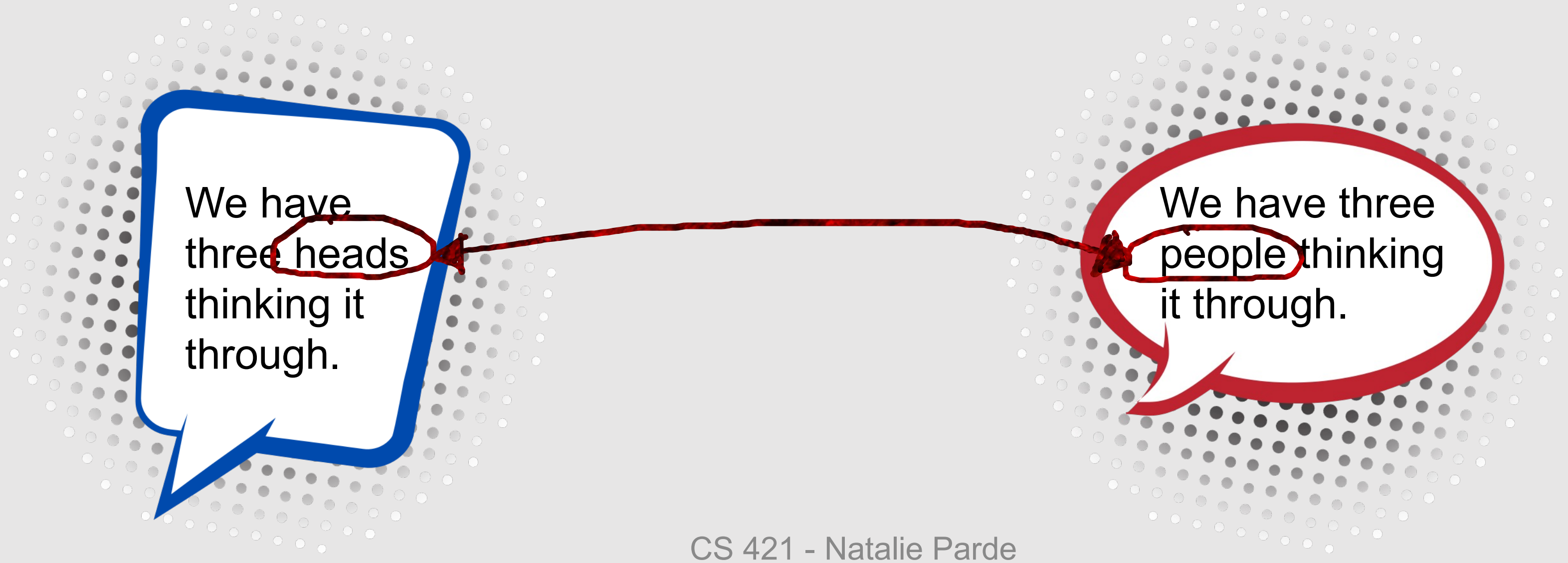
Hypernymy

- Occurs when one word sense is a generalization, or broader category, of another
- The word sense that is more general is the **hypernym**
- The word sense that is the more specific subclass of the broader word sense is the **hyponym**



Meronymy

- Closely related to hypernymy
- Occurs when one word sense refers to a part of another word sense
- The word sense that is the more general whole is the **holonym**



Structured Polysemy

- Semantically related senses associated with the same word
- Often seen when one word sense refers to an organization, and another sense refers to the building housing that organization

Metonymy

- Structured polysemy for which one aspect of a concept or entity is used to refer to other aspects of the entity or the entity itself
- Common examples are also found in:
  - Pairings between authors or artists and their works
  - Pairings between plants and their respective foods

Did you see the new Van Gogh at the art institute?

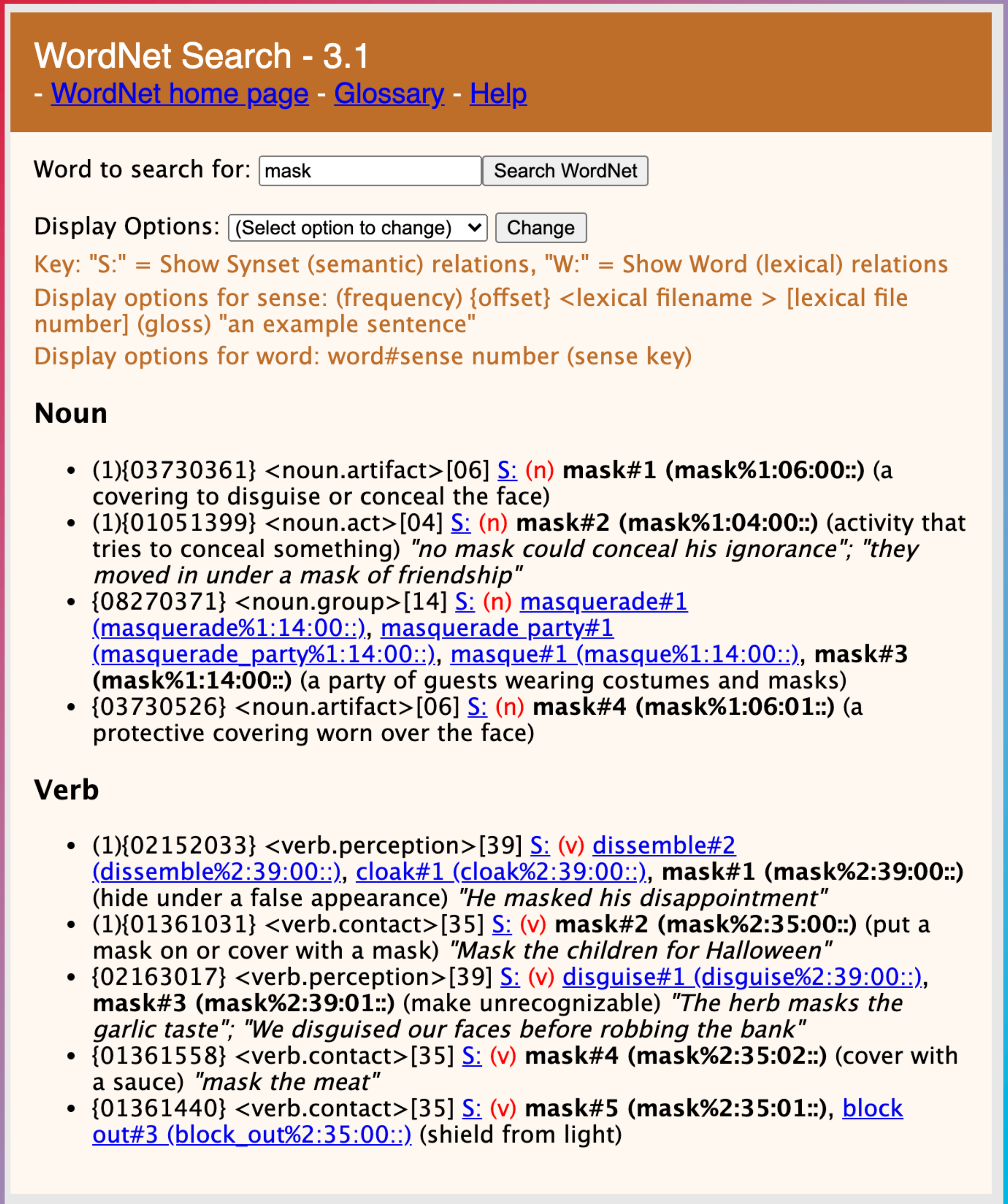
This Week's Topics

Tuesday

- Word Senses
- WordNet
- Word Sense Disambiguation

Thursday

- Semantic Roles
- Semantic Role Labeling
- Selectional Preferences


WordNet Search - 3.1
- WordNet home page - Glossary - Help

Word to search for: mask [Search WordNet]

Display Options: (Select option to change) [Change]

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
Display options for sense: (frequency) {offset} <lexical filename > [lexical file number] (gloss) "an example sentence"
Display options for word: word#sense number (sense key)

Noun

- (1){03730361} <noun.artifact>[06] S: (n) **mask#1 (mask%1:06:00::)** (a covering to disguise or conceal the face)
- (1){01051399} <noun.act>[04] S: (n) **mask#2 (mask%1:04:00::)** (activity that tries to conceal something) *"no mask could conceal his ignorance"; "they moved in under a mask of friendship"*
- {08270371} <noun.group>[14] S: (n) masquerade#1 (masquerade%1:14:00::), masquerade party#1 (masquerade_party%1:14:00::), masque#1 (masque%1:14:00::), **mask#3 (mask%1:14:00::)** (a party of guests wearing costumes and masks)
- {03730526} <noun.artifact>[06] S: (n) **mask#4 (mask%1:06:01::)** (a protective covering worn over the face)

Verb

- (1){02152033} <verb.perception>[39] S: (v) dissemble#2 (dissemble%2:39:00::), cloak#1 (cloak%2:39:00::), **mask#1 (mask%2:39:00::)** (hide under a false appearance) *"He masked his disappointment"*
- (1){01361031} <verb.contact>[35] S: (v) **mask#2 (mask%2:35:00::)** (put a mask on or cover with a mask) *"Mask the children for Halloween"*
- {02163017} <verb.perception>[39] S: (v) disguise#1 (disguise%2:39:00::), **mask#3 (mask%2:39:01::)** (make unrecognizable) *"The herb masks the garlic taste"; "We disguised our faces before robbing the bank"*
- {01361558} <verb.contact>[35] S: (v) **mask#4 (mask%2:35:02::)** (cover with a sauce) *"mask the meat"*
- {01361440} <verb.contact>[35] S: (v) **mask#5 (mask%2:35:01::)**, block out#3 (block_out%2:35:00::) (shield from light)

WordNet

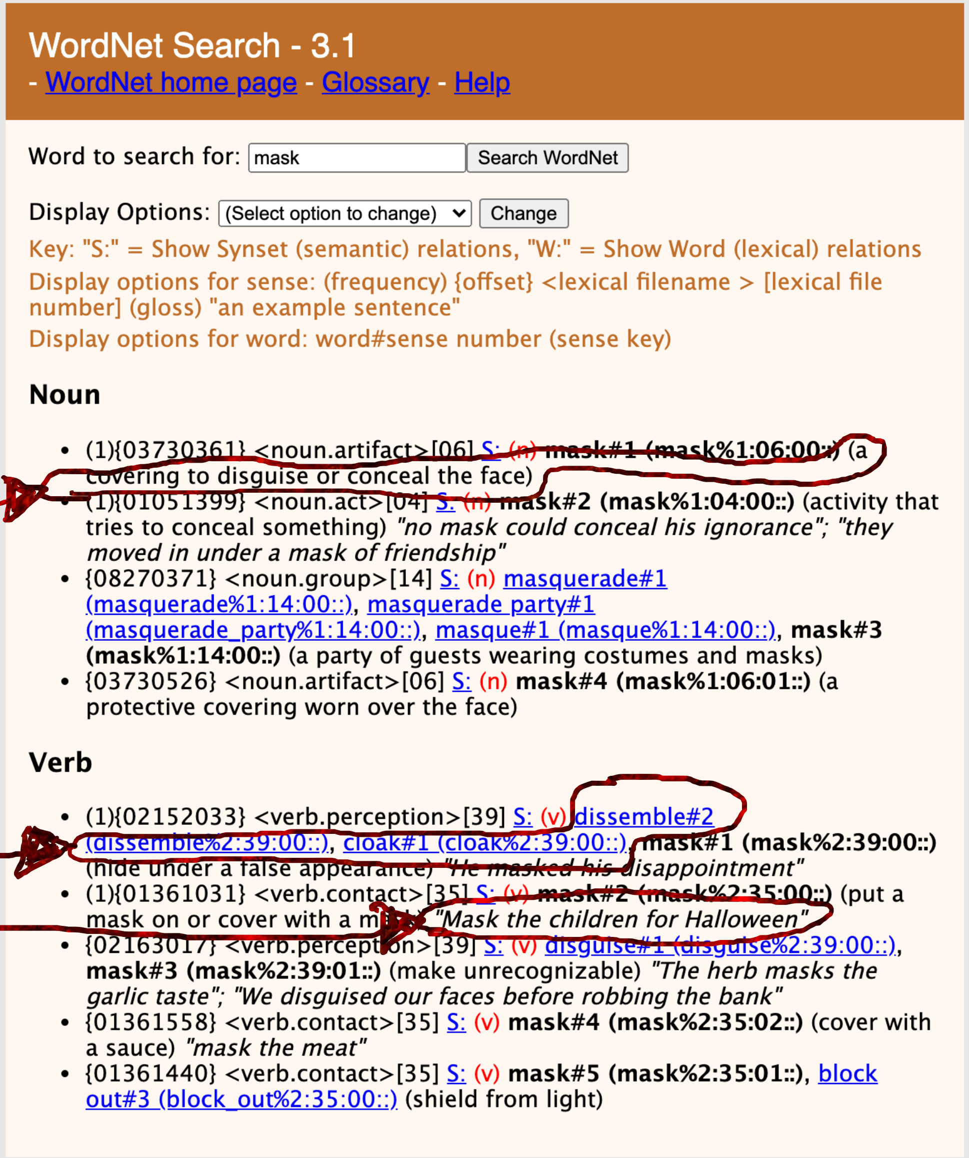
- Large lexical resource with information about:
  - Nouns
  - Verbs
  - Adjectives and adverbs
- Each entry is annotated with one or more **senses**
- Each sense provides a variety of information

WordNet

- Statistics for English WordNet 3.0:
  - 117,798 nouns
  - 11,529 verbs
  - 22,479 adjectives
  - 4,481 adverbs
- Average noun has 1.23 senses
- Average verb has 2.16 senses

WordNet Entries

- Senses contain:
  - **Gloss**
    - A definition of the sense
  - (Often) list of synonyms
    - Commonly referred to as a **synset**
  - (Sometimes) example sentence



WordNet Search - 3.1
- WordNet home page - Glossary - Help

Word to search for: mask Search WordNet

Display Options: (Select option to change) Change

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (frequency) {offset} <lexical filename > [lexical file number] (gloss) "an example sentence"

Display options for word: word#sense number (sense key)

Noun

- (1){03730361} <noun.artifact>[06] S: (n) **mask#1 (mask%1:06:00::)** (a covering to disguise or conceal the face)
- (1){01051399} <noun.act>[04] S: (n) **mask#2 (mask%1:04:00::)** (activity that tries to conceal something) *"no mask could conceal his ignorance"; "they moved in under a mask of friendship"*
- {08270371} <noun.group>[14] S: (n) masquerade#1 (masquerade%1:14:00::), masquerade party#1 (masquerade_party%1:14:00::), masque#1 (masque%1:14:00::), **mask#3 (mask%1:14:00::)** (a party of guests wearing costumes and masks)
- {03730526} <noun.artifact>[06] S: (n) **mask#4 (mask%1:06:01::)** (a protective covering worn over the face)

Verb

- (1){02152033} <verb.perception>[39] S: (v) dissemble#2 (dissemble%2:39:00::), cloak#1 (cloak%2:39:00::), **mask#1 (mask%2:39:00::)** (hide under a false appearance) *"He masked his disappointment"*
- (1){01361031} <verb.contact>[35] S: (v) **mask#2 (mask%2:35:00::)** (put a mask on or cover with a mask) *"Mask the children for Halloween"*
- {02163017} <verb.perception>[39] S: (v) disguise#1 (disguise%2:39:00::), **mask#3 (mask%2:39:01::)** (make unrecognizable) *"The herb masks the garlic taste"; "We disguised our faces before robbing the bank"*
- {01361558} <verb.contact>[35] S: (v) **mask#4 (mask%2:35:02::)** (cover with a sauce) *"mask the meat"*
- {01361440} <verb.contact>[35] S: (v) **mask#5 (mask%2:35:01::)**, block out#3 (block_out%2:35:00::) (shield from light)

Synsets

- Fundamental unit associated with WordNet entries
- Participate in lexical sense relations
- Facilitate relational navigation through the WordNet hierarchy

Sense Relations

- **Hypernym:** Relation between a concept and its superordinate
  - *Food* is a hypernym of *cake*
- **Hyponym:** Relation between a concept and its subordinate
  - *Corgi* is a hyponym of *dog*
- **Meronym:** Relation between a part and its whole
  - *Wheel* is a meronym of *car*
- **Holonym:** Relation between a whole and its parts
  - *Car* is a holonym of *wheel*
- **Antonym:** Relation between two semantically opposite concepts
  - *Leader* is an antonym of *follower*

Taxonomic Entities in WordNet

- Two kinds of taxonomic entities
  - **Classes**
  - **Instances**
- Instances: Individual proper nouns that represent unique entities
  - Chicago
- Classes: Generalized groups of instances
  - city

Additional Sense Relations

- Noun relations have a few additional distinctions:
  - **Instance hypernyms** are relationships from instances to their concepts (e.g., “Austen → author” rather than “breakfast → meal”)
  - **Derivations** are lemmas with the same morphological root (e.g., “destruction ↔ destroy”)
- So do verbs:
  - **Troponyms** are relationships from events to subordinate events (e.g., “stroll” is a troponym of “walk”)
  - **Entailments** are relationships from verbs to the verbs they entail (e.g., “borrow” entails “obtain”)

Lexicographic Categories

- Coarse-grained semantic categories
  - Often referred to as **supersenses**
- 26 categories for nouns
- 15 categories for verbs

Category	Example	Category	Example	Category	Example
ACT	service	GROUP	place	PLANT	tree
ANIMAL	dog	LOCATION	area	POSSESSION	price
ARTIFACT	car	MOTIVE	reason	PROCESS	process
ATTRIBUTE	quality	NATURAL EVENT	experience	QUANTITY	amount
BODY	hair	NATURAL OBJECT	flower	RELATION	portion
COGNITION	way	OTHER	stuff	SHAPE	square
COMMUNICATION	review	PERSON	people	STATE	pain
FEELING	discomfort	PHENOMENON	result	SUBSTANCE	oil
FOOD	food			TIME	day

- {03211439} <noun.artifact>[06] S: (n) disguise#2 (disguise%1:06:00::) (any attire that modifies the appearance in order to conceal the wearer's identity)
  - {02759103} <noun.artifact>[06] S: (n) attire#1 (attire%1:06:00::), garb#1 (garb%1:06:00::), dress#2 (dress%1:06:01::) (clothing of a distinctive style or for a particular occasion) *"formal attire"; "battle dress"*
    - {03055525} <noun.artifact>[06] S: (n) clothing#1 (clothing%1:06:00::), article of clothing#1 (article_of_clothing%1:06:00::), vesture#2 (vesture%1:06:00::), wear#2 (wear%1:06:00::), wearable#1 (wearable%1:06:00::), habiliment#1 (habiliment%1:06:00::) (a covering designed to be worn on a person's body)
      - {03127399} <noun.artifact>[06] S: (n) covering#2 (covering%1:06:00::) (an artifact that covers something else (usually to protect or shelter or conceal it))
        - {00022119} <noun.Tops>[03] S: (n) artifact#1 (artifact%1:03:00::), artefact#1 (artefact%1:03:00::) (a man–made object taken as a whole)
          - {00003553} <noun.Tops>[03] S: (n) whole#2 (whole%1:03:00::), unit#6 (unit%1:03:00::) (an assemblage of parts that is regarded as a single entity) *"how big is that part compared to the whole?"; "the team is a unit"*
            - {00002684} <noun.Tops>[03] S: (n) object#1 (object%1:03:00::), physical object#1 (physical_object%1:03:00::) (a tangible and visible entity; an entity that can cast a shadow) *"it was full of rackets, balls and other objects"*
              - {00001930} <noun.Tops>[03] S: (n) physical entity#1 (physical_entity%1:03:00::) (an entity that has physical existence)
                - {00001740} <noun.Tops>[03] S: (n) entity#1 (entity%1:03:00::) (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

- (1){03730361} <noun.artifact>[06] S: (n) **mask#1 (mask%1:06:00::)** (a covering to disguise or conceal the face)
  - *direct hyponym* / *full hyponym*
  - *direct hypernym* / ***inherited hypernym*** / *sister term*
    - {03127399} <noun.artifact>[06] S: (n) covering#2 (covering%1:06:00::) (an artifact that covers something else (usually to protect or shelter or conceal it))
      - {00022119} <noun.Tops>[03] S: (n) artifact#1 (artifact%1:03:00::), artefact#1 (artefact%1:03:00::) (a man–made object taken as a whole)
        - {00003553} <noun.Tops>[03] S: (n) whole#2 (whole%1:03:00::), unit#6 (unit%1:03:00::) (an assemblage of parts that is regarded as a single entity) *"how big is that part compared to the whole?"; "the team is a unit"*
          - {00002684} <noun.Tops>[03] S: (n) object#1 (object%1:03:00::), physical object#1 (physical_object%1:03:00::) (a tangible and visible entity; an entity that can cast a shadow) *"it was full of rackets, balls and other objects"*
            - {00001930} <noun.Tops>[03] S: (n) physical entity#1 (physical_entity%1:03:00::) (an entity that has physical existence)
              - {00001740} <noun.Tops>[03] S: (n) entity#1 (entity%1:03:00::) (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

- {03098030} <noun.artifact>[06] S: (n) consumer goods#1 (consumer_goods%1:06:00::) (goods (as food or clothing) intended for direct use or consumption)
  - {03080712} <noun.artifact>[06] S: (n) commodity#1 (commodity%1:06:00::), trade good#1 (trade_good%1:06:00::), good#4 (good%1:06:00::) (articles of commerce)
    - {00022119} <noun.Tops>[03] S: (n) artifact#1 (artifact%1:03:00::), artefact#1 (artefact%1:03:00::) (a man–made object taken as a whole)
      - {00003553} <noun.Tops>[03] S: (n) whole#2 (whole%1:03:00::), unit#6 (unit%1:03:00::) (an assemblage of parts that is regarded as a single entity) *"how big is that part compared to the whole?"; "the team is a unit"*
        - {00002684} <noun.Tops>[03] S: (n) object#1 (object%1:03:00::), physical object#1 (physical_object%1:03:00::) (a tangible and visible entity; an entity that can cast a shadow) *"it was full of rackets, balls and other objects"*
          - {00001930} <noun.Tops>[03] S: (n) physical entity#1 (physical_entity%1:03:00::) (an entity that has physical existence)
            - {00001740} <noun.Tops>[03] S: (n) entity#1 (entity%1:03:00::) (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Hierarchical Structure

Check out WordNet for yourself!

- You can browse WordNet using the link here: http://wordnetweb.princeton.edu/perl/webwn
- You can also programmatically access WordNet using NLTK: https://www.nltk.org/howto/wordnet.html

This Week's Topics

Tuesday

- Word Senses
- WordNet
- Word Sense Disambiguation

Thursday

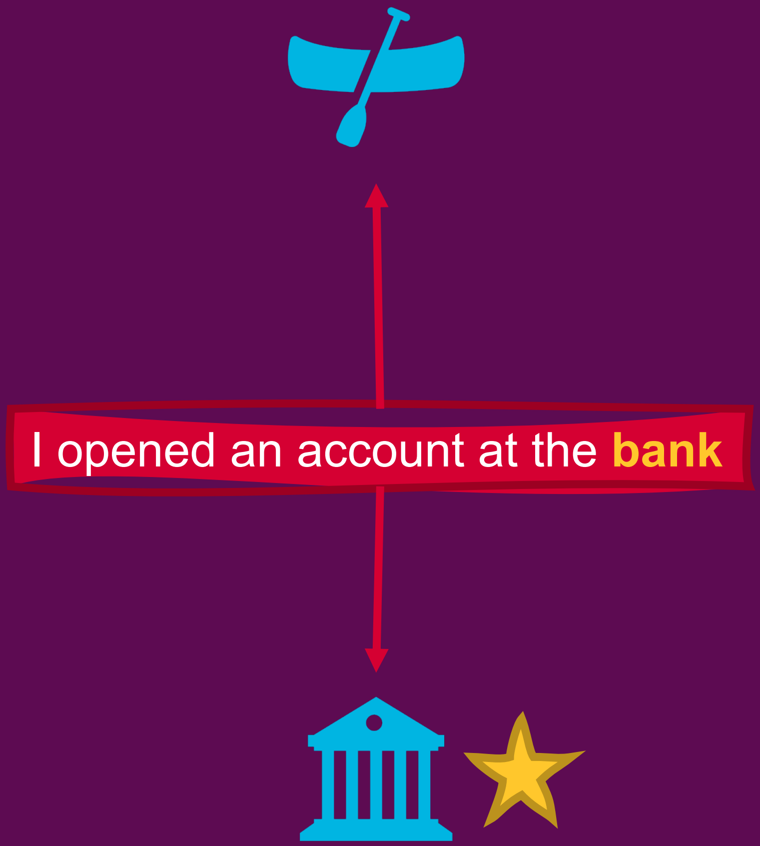
- Semantic Roles
- Semantic Role Labeling
- Selectional Preferences

Somehow, NLP systems need to be able to determine which sense is used in a given context.

- How can we do this?
  - Word sense disambiguation

What is word sense disambiguation?

- **Word sense disambiguation:** The task of automatically selecting the correct sense for a given word
- Input: A word in context
- Output: The correct word sense from a fixed inventory of potential word senses
- The best approach for solving this will depend on your domain and the size of your word and sense sets



Popular Sense-Tagged Corpora

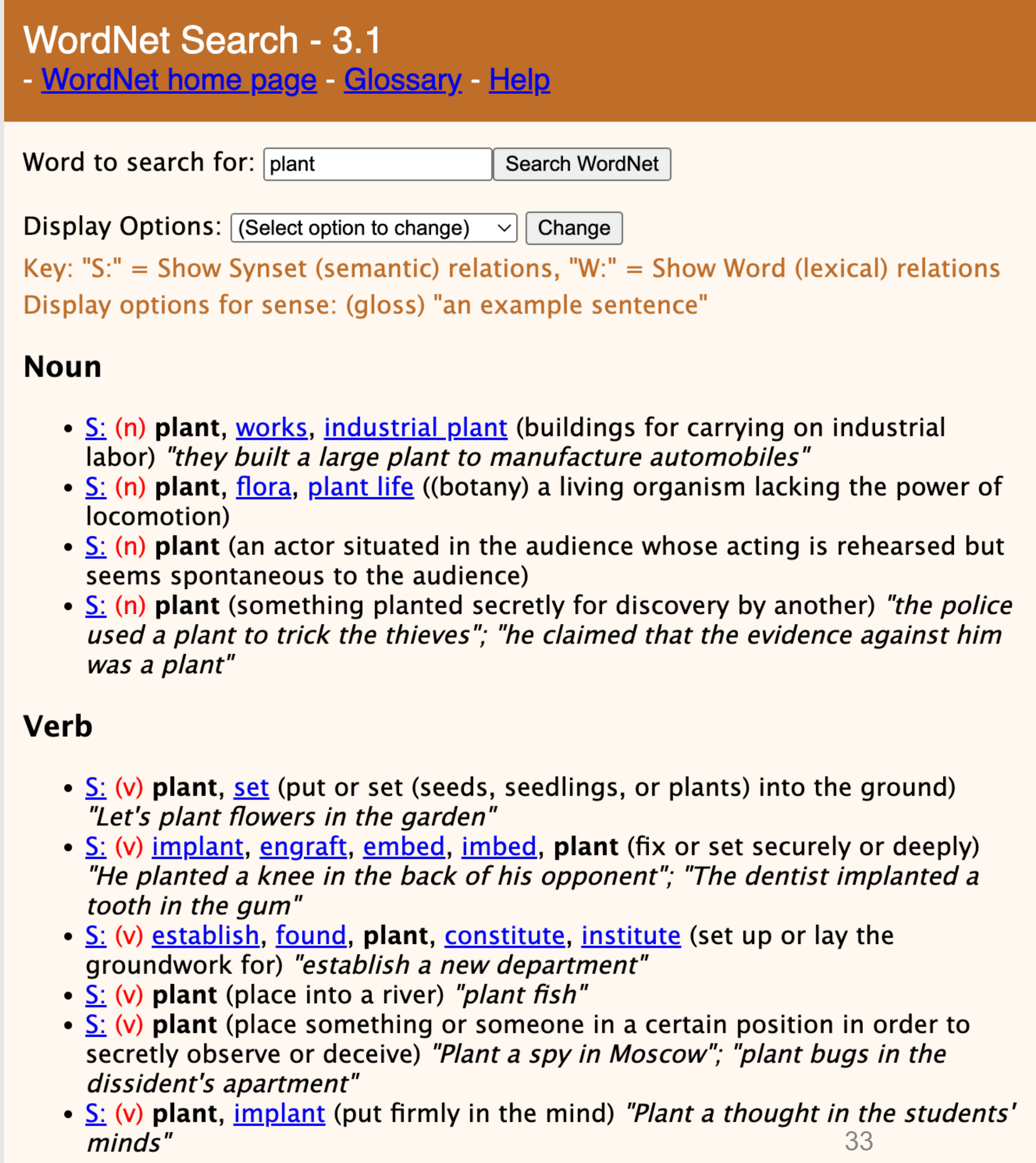
- SemCor: https://www.sketchengine.eu/semcor-annotated-corpus/
- Senseval Corpora: https://web.eecs.umich.edu/~mihalcea/senseval/senseval3/tasks.html
- Certain SemEval corpora: http://alt.qcri.org/semeval2015/task13/
- Sense tag inventories may be domain-specific
  - A word may have many senses in a specialized domain, but fewer senses in the general domain

Word Sense Disambiguation

Given a word, what is its correct sense?

I love my new purple plant!

?



WordNet Search - 3.1
- WordNet home page - Glossary - Help

Word to search for: plant [Search WordNet]

Display Options: (Select option to change) [Change]

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
Display options for sense: (gloss) "an example sentence"

Noun

- S: (n) **plant**, works, industrial plant (buildings for carrying on industrial labor) *"they built a large plant to manufacture automobiles"*
- S: (n) **plant**, flora, plant life ((botany) a living organism lacking the power of locomotion)
- S: (n) **plant** (an actor situated in the audience whose acting is rehearsed but seems spontaneous to the audience)
- S: (n) **plant** (something planted secretly for discovery by another) *"the police used a plant to trick the thieves"; "he claimed that the evidence against him was a plant"*

Verb

- S: (v) **plant**, set (put or set (seeds, seedlings, or plants) into the ground) *"Let's plant flowers in the garden"*
- S: (v) implant, engraft, embed, imbed, **plant** (fix or set securely or deeply) *"He planted a knee in the back of his opponent"; "The dentist implanted a tooth in the gum"*
- S: (v) establish, found, **plant**, constitute, institute (set up or lay the groundwork for) *"establish a new department"*
- S: (v) **plant** (place into a river) *"plant fish"*
- S: (v) **plant** (place something or someone in a certain position in order to secretly observe or deceive) *"Plant a spy in Moscow"; "plant bugs in the dissident's apartment"*
- S: (v) **plant**, implant (put firmly in the mind) *"Plant a thought in the students' minds"*

Task Complexity

- WSD grows more challenging as the number of words being disambiguated grows
- Lexical sample tasks
  - Small pre-selected set of target words
  - Inventory of senses for each word from a lexicon
- All-words tasks
  - Entire large texts
  - Inventory of senses for each word from a lexicon
  - Conceptually similar to POS tagging with a much larger tagset

Semantic Concordances

- All-words tasks are often trained using **semantic concordances**
  - Corpora for which each open-class word in a sentence is labeled with its word sense
- Word senses are then predicted similarly to other sequence tagging tasks

Effective word sense disambiguation is required for many tasks.

- Question answering
  - To which form of “mouse” is the user referring?
- Machine translation
  - Word senses associated with a source language word may not all directly transfer to its target language translation!
- Evaluating NLP models
  - Do word representations accurately reflect relevant word sense similarities?
- Word sense disambiguation tends to be especially challenging in low-resource or highly specialized domains

WSD Baselines

- **Most frequent sense**
  - Given a new word, assign the most frequent sense to it based on counts from a training corpus
  - Often used as a default method when a supervised model has insufficient data to learn the task effectively

WSD Baselines

- **One sense per discourse**
  - Given a new word, if an instance of the same word has already been assigned a sense earlier in the current discourse (by selecting the most frequent sense or applying some other method), assign that same sense
  - Words appearing multiple times in a text or discourse often appear with the same sense (Gale et al., 1992)
    - Gale, W.A., Church, K.W. & Yarowsky, D. A method for disambiguating word senses in a large corpus. Comput Hum 26, 415–439 (1992). https://doi.org/10.1007/BF00136984
  - Works especially well with coarse-grained senses that are unrelated
  - Less popular than most frequent sense baseline

What are some more sophisticated WSD techniques?

- Lesk algorithm
- Feature-based models
- Contextual embedding models

Lesk Algorithm

- Classic, powerful, **knowledge-based approach**
- Intuition: Given the glosses for all possible senses of a word, the gloss that shares the most words with the immediate context of the target word corresponds to the correct sense

Simplified Lesk Algorithm

```
best_sense ← most frequent sense for word
max_overlap ← 0
context ← set of words in sentence
for each sense in senses of word do:
      signature ← set of words in the gloss and examples of sense
      overlap ← compute_overlap(signature, context)
      if overlap > max_overlap then:
            max_overlap ← overlap
            best_sense ←  sense
return best_sense
```

Case Example: Simplified Lesk Algorithm

The **bank** can guarantee deposits will eventually cover future tuition costs because it invests in adjustable-rate mortgage securities.

bank^1	Gloss	A financial institution that accepts deposits and channels the money into lending activities
	Examples	“he cashed a check at the bank,” “that bank holds the mortgage on my home”
bank^2	Gloss	Sloping land (especially the slope beside a body of water)
	Examples	“they pulled the canoe up on the bank,” “he sat on the bank of the river and watched the currents”

Case Example: Simplified Lesk Algorithm

The **bank** can guarantee **deposits** will eventually cover future tuition costs because it invests in adjustable-rate **mortgage** securities.

☺

$bank^1$	Gloss	A financial institution that accepts **deposits** and channels the money into lending activities
	Examples	“he cashed a check at the bank,” “that bank holds the **mortgage** on my home”
$bank^2$	Gloss	Sloping land (especially the slope beside a body of water)
	Examples	“they pulled the canoe up on the bank,” “he sat on the bank of the river and watched the currents

Feature-Based WSD

- Choose the best sense based on feature representations and feature-based classification algorithms
- Common features:
  - **Part-of-speech tags** for words before and after the target word
  - **N-grams** before and after the target word
  - **Weighted average of embeddings** for words before and after the target word

Contextual Embedding Models

- Current best-performing models for word sense disambiguation
- Task is framed similar to other neural sequence labeling tasks
- Contextual word embeddings:
  - Word embeddings that differ depending on a word's specific use
  - Word2Vec does *not* produce contextual word embeddings!
  - Contextual embeddings are generally produced using encoder-based approaches
    - **ELMo** was a pioneering implementation of this: https://aclanthology.org/N18-1202.pdf

Contextual Embedding Models

- To train:
  - Extract a contextual embedding for each word in a sense-labeled training set
  - For a given word sense *c*, average the contextual embeddings of all instances of that sense $\mathbf{c_i}$:
    - $\mathbf{v_s} = \frac{1}{n}\sum_i \mathbf{c_i}$
- To test:
  - Compute a contextual embedding $\mathbf{t_i}$ for the target word
  - Select the sense embedding $\mathbf{v_s}$ associated with that target word that has the highest cosine similarity with $\mathbf{t_i}$

What about words that didn't exist in the training data?

One option: Develop simple heuristics for these cases

More sophisticated option: Impute the missing sense embeddings using the WordNet taxonomy and supersenses

Imputing Missing Sense Embeddings

- Find sense embeddings for higher-level nodes in the WordNet taxonomy by averaging the embeddings of their children
  - For each missing sense in WordNet, $\hat{s} \in W$:
    - Let the sense embeddings for other members of its synset be $S_{\hat{s}}$
    - Let the hypernym-specific synset embeddings be $H_{\hat{s}}$
    - Let the lexicographic synset embeddings be $L_{\hat{s}}$
- This produces:
  - An embedding for each synset as the average of its sense embeddings
    - If $|S_{\hat{s}}| > 0$, $\mathbf{v}_{\hat{s}} = \frac{1}{|S_{\hat{s}}|} \sum \mathbf{v}_s, \forall \mathbf{v}_s \in S_{\hat{s}}$
  - An embedding for each hypernym as the average of its synset embeddings
    - Else if $|H_{\hat{s}}| > 0$, $\mathbf{v}_{\hat{s}} = \frac{1}{|H_{\hat{s}}|} \sum \mathbf{v}_{syn}, \forall \mathbf{v}_{syn} \in H_{\hat{s}}$
  - An embedding for each supersense as the average of the synset embeddings belonging to that lexicographic category
    - Else if $|L_{\hat{s}}| > 0$, $\mathbf{v}_{\hat{s}} = \frac{1}{|L_{\hat{s}}|} \sum \mathbf{v}_{syn}, \forall \mathbf{v}_{syn} \in L_{\hat{s}}$

This is guaranteed to produce a representation for every missing sense.

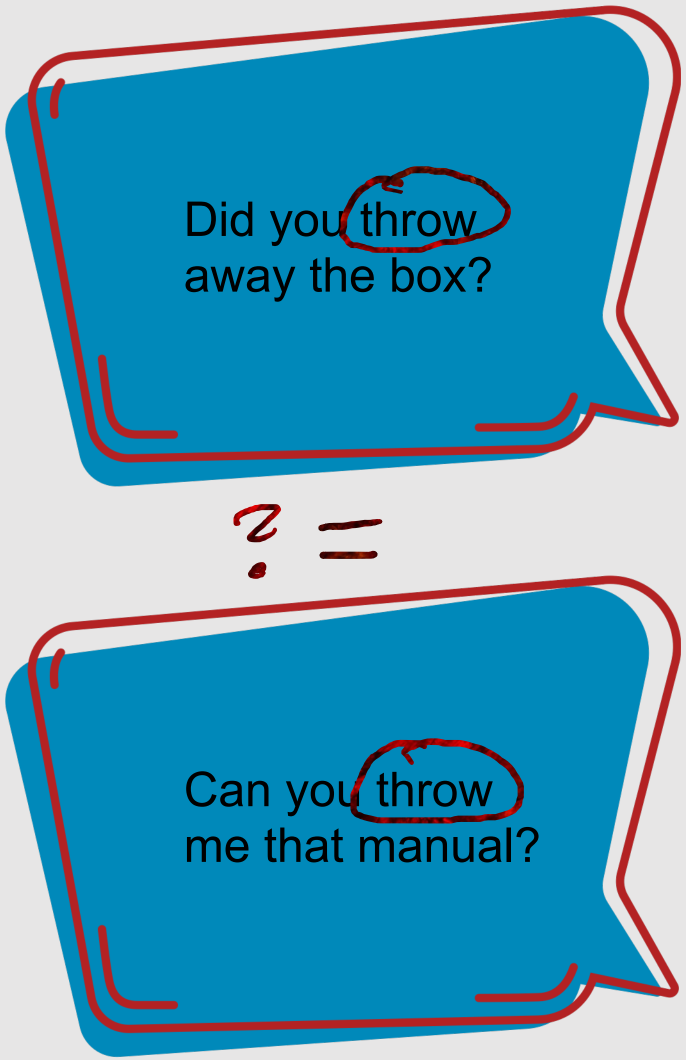
- All supersenses have labeled data in SemCor
- Thus, the algorithm will have some representation for all possible senses by the time it backs off to the lexicographic (supersense) information
- Using information from higher taxonomic levels will produce more coarse-grained sense embeddings

Word Similarity at Different Granularities

- WSD is more fine-grained than earlier word similarity tasks
- Context-free word similarity (how similar is “Chicago” to “Dallas”?)
- This is because word sense disambiguation is a contextualized similarity task
  - Goal is to distinguish the meaning of a word in one context from its meaning in another
- The **word-in-context** task lies between these two extremes

Word-in-Context Evaluation

- Given two sentences with the same target word but different context, decide whether the target words are used:
  - In the same sense, or
  - In different senses
- Word-in-context models generally first cluster word senses in to coarser-grained groups
  - First-degree connections are clustered together
  - Senses belonging to the same supersense are clustered together
- Words are considered as belonging to the same "sense" if they belong to the same cluster



How can we solve word-in-context tasks?

- Simple approach:
  - Compute the contextual embedding for the target word in each of the two sentences
  - Compute the cosine similarity between those embeddings
  - If the cosine similarity is above a threshold, predict that the words are used in the same sense
  - Otherwise, predict that they are used in different senses

Additional Data Acquisition for WSD

- SemCor is often used for WSD, but other data sources can also be leveraged
- One useful resource: Wikipedia
  - Hyperlinks to concepts can be used as sense annotations
  - However, Wikipedia concepts must be mapped to relevant senses for WSD

How can we map Wikipedia concepts to WordNet senses?

- For a given WordNet synset, find the words in the:
  - Synset
  - Gloss
  - Related senses
- For a given Wikipedia concept, find the words in the:
  - Page title
  - Outgoing links
  - Page category
- Select the WordNet sense with the greatest lexical overlap with the Wikipedia concept

Using Lexical Resources to Improve Word Embeddings

- Beyond assisting with WSD, resources like WordNet can be used to improve the quality of learned word embeddings
- This can resolve well-known systemic embedding issues, such as poor estimation of antonymy in static word embeddings
- How can these resources be used?
  - **Retraining**
  - **Retrofitting**

Retraining Word Embeddings

- Modify the embedding's training process to incorporate word sense relations
  - Synonymy
  - Antonymy
  - Hypernymy
- In Word2Vec, this can be done by modifying the static embedding loss function to make use of this information

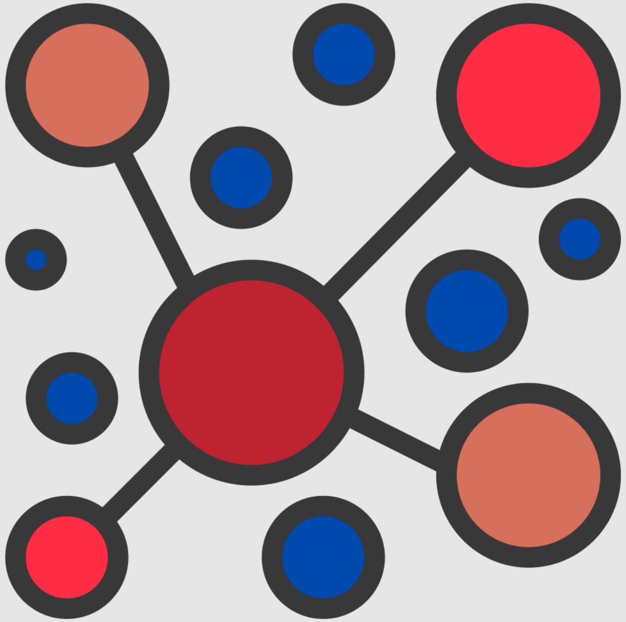
Retrofitting Word Embeddings

- Learn a second mapping based on the lexical resource that shifts the embeddings in such a way that synonyms are pushed closer together and antonyms are pulled further apart
- Also referred to as **counterfitting**

When working with large or unconstrained vocabularies, supervised WSD can be difficult.

- Expensive (and sometimes impractical) to build large corpora labeled with word senses!
- Alternative: Unsupervised word sense disambiguation, or **word sense induction**

Word Sense Induction



- Creates sets of words automatically from a large, unlabeled training set
- Often done using **clustering techniques**
  - Centroid of a cluster represents the **sense vector** corresponding to a sense
  - To induce word senses for new words, algorithms can assign them to the sense vector that is closest to the contextual vector for a given word

More formally, to train….

- For each token w_i of word w in a corpus, compute a context vector $\mathbf{c}$
- Use a clustering algorithm to cluster the context vectors $\mathbf{c}$ into a predefined number of clusters, each of which define a sense of w
- Compute the vector centroid, $\mathbf{s_j}$, of each cluster to produce the sense vectors for w

To test….

- Compute a context vector $\mathbf{c}$ for a test token t of word w
- Retrieve all sense vectors $\mathbf{s_j}$ for w
- Assign t to the sense represented by the vector $\mathbf{s_j}$ that is closest to $\mathbf{c}$

Clustering

- Unsupervised machine learning approach that groups data points into “clusters” with similar representations
- Many clustering algorithms exist
  - K-means clustering
  - Density-based clustering
  - Gaussian mixture models
  - And many more!

What clustering method should we use?

- In theory we can use any clustering algorithm for word sense induction
- Common in NLP tasks: **Agglomerative clustering**
  - Each training instance is initially assigned to its own cluster
  - New clusters are formed using a bottom-up process in which the two most similar clusters are successively merged
  - This process continues until the specified number of clusters is reached, or a global cluster quality measure is achieved

Evaluating Unsupervised Word Sense Induction Approaches

- Best approach: Extrinsic evaluation
- If intrinsic evaluation is needed:
  - Measure cluster overlap
  - Map sense clusters to predefined senses
  - Devise other approaches that map automatically-derived sense classes to an established gold standard for performance comparison
- There is no standardized evaluation metric (yet!) for this task

Summary: Word Senses and WordNet

- Word **senses** define a word's meaning in context
- Many words are **polysemous**
- Word senses can be related to one another in many ways, such as through **synonymy**, **antonymy**, **meronymy**, and **hypernymy**
- **WordNet** is a large lexical database with word sense information for nouns, verbs, adjectives, and adverbs
- **Word sense disambiguation** is the task of determining the correct sense for a word, given its context
- WSD can be performed in a variety of ways, including with contextual embedding approaches, feature-based algorithms, the **Lesk algorithm**, or a most frequent sense baseline
- Word senses can also be **induced** using unsupervised clustering methods

This Week's Topics

Word Senses
WordNet
Word Sense Disambiguation

Thursday

Tuesday

Semantic Roles
Semantic Role Labeling
Selectional Preferences

Semantic Roles

- When extracting information from text, it is useful to understand how participants relate to events
  - Who did what?
  - When?
  - Where?
- We can learn this information using **semantic roles**

Semantic Roles

- The underlying purpose of a word with respect to a predicate
- Many possible semantic roles!
- Set of roles may vary depending on the application

Recall the meaning representations we've already seen....

Natalie baked the cake.

$\exists e, x, y \text{ Baking}(e) \wedge \text{Baker}(e, \text{Natalie}) \wedge \text{BakedThing}(e, y) \wedge \text{Cake}(y)$

Recall the meaning representations we've already seen….

Natalie baked the cake.

- Subject of “bake”
- Deep role specific to the “baking” event

$\exists e, x, y \; \text{Baking}(e) \land \text{Baker}(e, \text{Natalie}) \land \text{BakedThing}(e, y) \land \text{Cake}(y)$

What if we consider another sentence?

Natalie ate the cake.

- Subject of “ate”
- Deep role specific to the “eating” event

$\exists e, x, y \text{ Eating}(e) \wedge \text{Eater}(e, \text{Natalie}) \wedge \text{EatenThing}(e, y) \wedge \text{Cake}(y)$

There are commonalities between these roles!

- “Bakers” and “Eaters” are both:
  - Volitional actors
  - Generally animate
  - Have causal responsibility for their events
- How can we capture this commonality more formally?

Semantic Roles

- **Semantic roles:** Underlying semantic commonalities among arguments to different types of events
- Also sometimes referred to as **thematic roles**

$$\exists e, x, y \ \mathrm{Baking}(e) \wedge \mathrm{Baker}(e, \mathrm{Natalie}) \wedge \mathrm{BakedThing}(e, y) \wedge \mathrm{Cake}(y)$$

Theme

Agent

$$\exists e, x, y \ \mathrm{Eating}(e) \wedge \mathrm{Eater}(e, \mathrm{Natalie}) \wedge \mathrm{EatenThing}(e, y) \wedge \mathrm{Cake}(y)$$

Semantic roles are ancient!

- First formalized by Pāṇini sometime between 700-400 BCE
- More recently formalized in the 1960s
  - Fillmore (1968): https://files.eric.ed.gov/fulltext/ED019631.pdf
  - Gruber (1965): http://www.ai.mit.edu/projects/dm/theses/gruber65.pdf
- No universally agreed-upon roles, but some are common across numerous papers

Common Semantic Roles

THEMATIC ROLE	DEFINITION	EXAMPLE
Agent	The volitional causer of an event	The **waiter** spilled the soup.
Experiencer	The experiencer of an event	**John** has a headache.
Force	The non-volitional causer of the event	The **wind** blows debris from the mall into our yards.
Theme	The participant most directly affected by an event	Only after Benjamin Franklin broke the **ice**….
Result	The end product of an event	The city built a **regulation-size baseball diamond**….
Content	The proposition or content of a propositional event	Mona asked, **"You met Mary Ann at the supermarket?"**
Instrument	An instrument used in an event	He poached catfish, stunning them with a **shocking device**….
Beneficiary	The beneficiary of an event	Whenever Ann Callahan makes hotel reservations for her **boss**….
Source	The origin of the object of a transfer event	I flew in from **Boston**.
Goal	The destination of an object of a transfer event	I drove to **Portland**.

How many semantic roles are typically considered?

Some sets use smaller numbers of roles, each of which are more abstract

Some sets use larger numbers of roles, each of which are more specific

We can refer to all sets of roles as **semantic roles**

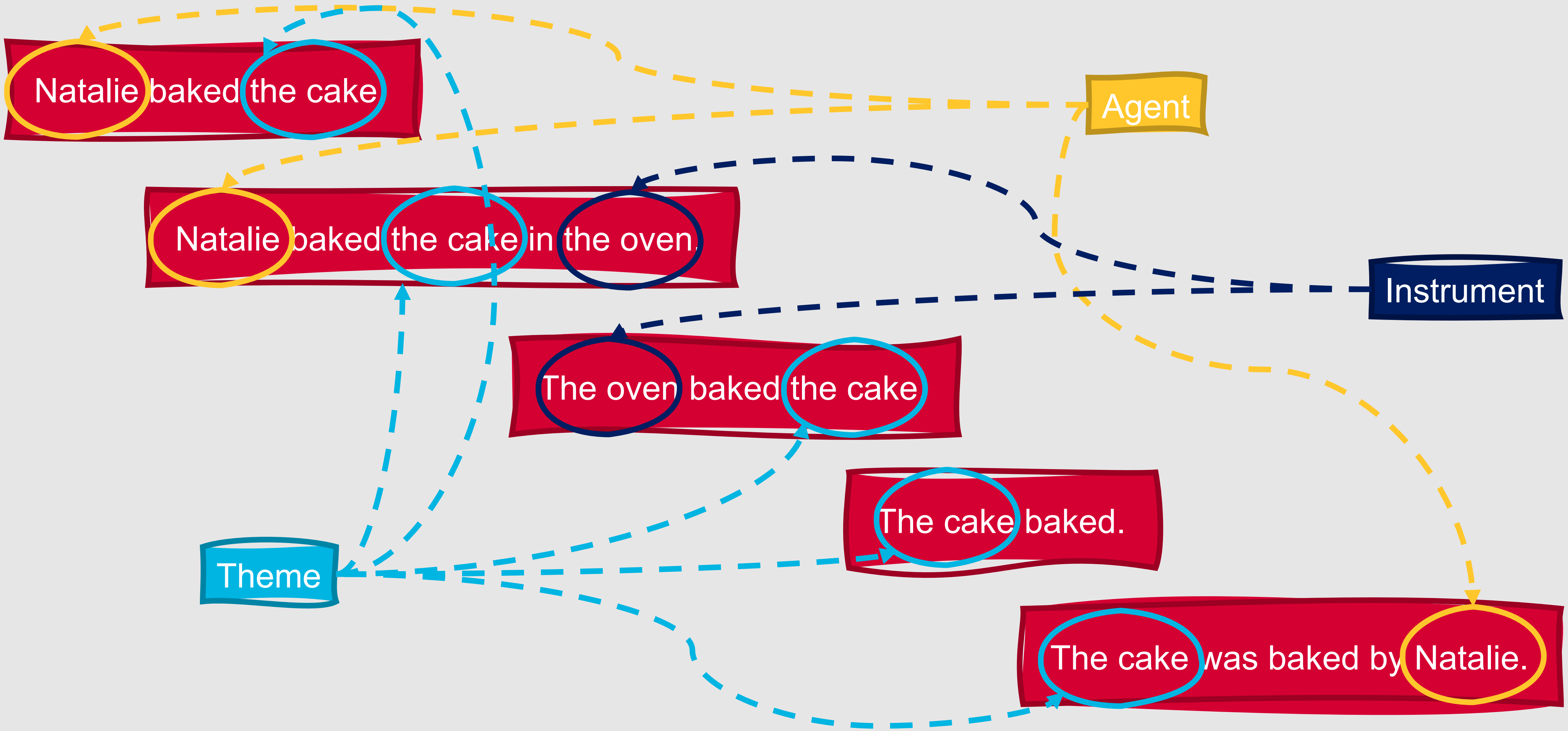
Semantic roles thus offer another way for us to construct shallow meaning representations.

They allow us to:

- Make inferences that aren't possible from surface representations or parse trees
- Create intermediate languages for downstream tasks

In general, semantic roles help us generalize over different surface realizations of the same predicate arguments

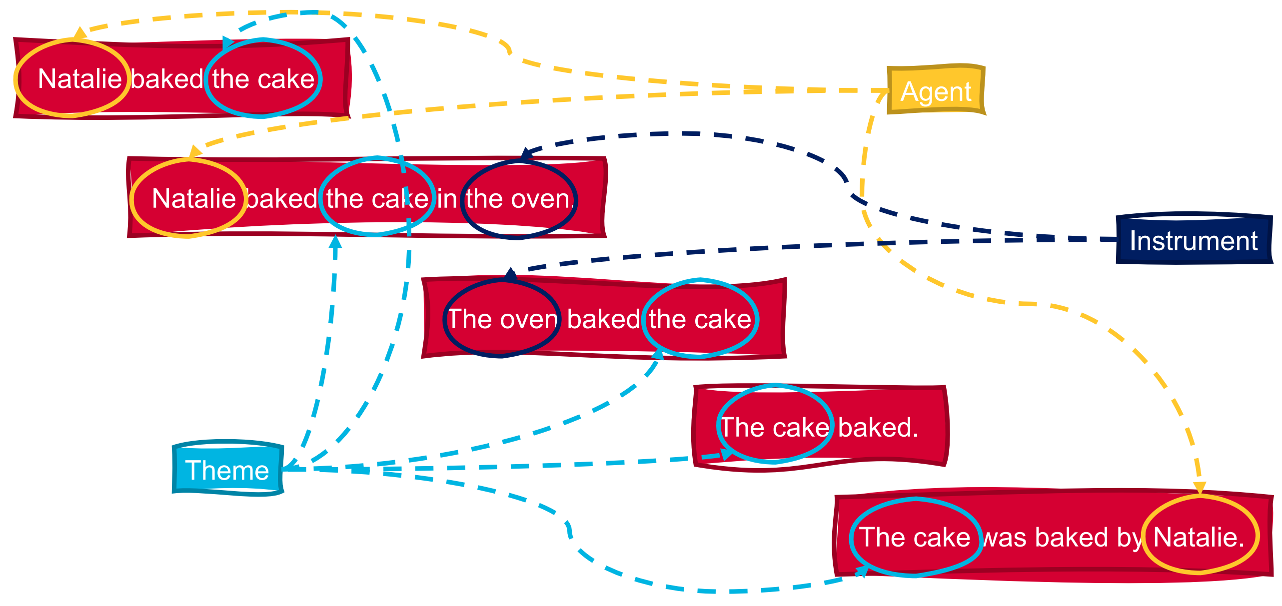
For example….



Natalie baked the cake
Natalie baked the cake in the oven.
The oven baked the cake
The cake baked.
The cake was baked by Natalie.
Agent
Instrument
Theme

Thematic Grid

- The set of semantic role arguments taken by a verb
  - Also sometimes referred to as a **case frame**
- Semantic roles can often be realized in different syntactic positions
  - For example:
    - Agent=Subject; Theme=Object
    - Instrument=Subject; Theme=Object
    - Theme=Subject



Diathesis Alternations

- **Diathesis Alternations:** Alternate acceptable structural realizations for arguments
- This facilitates generalization over different surface realizations
- Different verbs can participate in different alternations

Challenges Associated with Semantic Roles

- Difficult to create a standard set of roles
- Difficult to define roles in general
  - Formally, what is an *Agent*?
- Difficult to reason about and generalize across semantic roles

Defining Role Sets

- Researchers often find it necessary to fragment more general roles (e.g., Agent) into more specific roles

Instrument

Intermediary: Can appear as subjects

Enabling: Cannot appear as subjects

Conformity to Predefined Properties

- Individual noun phrases may not conform to all properties of an *Agent*, but they might conform to most …can they still be labeled with this role?
  - Might require even more fragmentation!

How can these challenges be addressed?

- **Generalized semantic roles**
  - Proto-Agents
  - Proto-Patients
  - Fewer, more abstract roles
- **Semantic roles tailored to specific semantic classes**
  - Additional, more specific roles

VerbNet

- An online resource indicating the semantic classes to which many English verbs belong
- Linked to WordNet and FrameNet entries
- Link: https://verbs.colorado.edu/verbnet/
  - Also an API: https://github.com/cu-clear/verbnet/
  - Also accessible via NLTK: https://www.nltk.org/_modules/nltk/corpus/reader/verbnet.html

VerbNet

CREATE-26.4

Back to search

Full Class View

- create-26.4
  - create-26.4-1
    - create-26.4-1-1

Member Verb Lemmas:

AUTHOR | COIN | COMPUTE | CONCOCT | CONSTRUCT | CONTRIVE | COWRITE | CREATE | DERIVE | FABRICATE | FORM | FORMULATE | LAY | MANUFACTURE | MASS-PRODUCE | MODEL | ORGANIZE | PRODUCE | PUBLISH | REARRANGE | REBUILD | RECONSTITUTE | REORGANIZE | STYLE | SYNTHESIZE | TURN-OUT

ROLES:

Agent [+animate | +machine]
Result
Material
Beneficiary [+animate]
Attribute

- NP V NP
- NP V NP PP.material
- NP V NP PP.beneficiary
- NP V NP PP.attribute

EXAMPLE:

David constructed a house.

SHOW DEPENDENCY PARSE TREE

SYNTAX:

Agent VERB Result

SEMANTICS:

¬ HAS_STATE(e1 , ?Material , V_Final_State)
¬ BE(e1 , Result)
DO(e2 , Agent)
BE(e3 , Result)
HAS_STATE(e3 , ?Material , V_Final_State)
CAUSE(e2 , e3)

FORCE DYNAMICS:

Volitional Create FD representation

Subclasses:

CREATE-26.4-1

Generalized Semantic Roles

- Abstract over specific thematic roles
- Roles are defined by heuristic features that accompany properties likely to correspond with the generalized class
  - Proto-Agent: Agent-like properties
- More overlapping properties → argument likelier to be labeled with that role

Specialized Semantic Roles

- Define roles that are specific to a particular verb or a group of semantically related verbs or nouns
  - A **Cook** creates a **Produced_food** from (raw) **Ingredients**.
  - The **Heating_instrument** and/or the **Container** may also be specified.

What are some popular resources for semantic role labeling?

PropBank

- https://propbank.github.io/
- Both generalized and verb-specific roles

FrameNet

- https://framenet.icsi.berkeley.edu/fndrupal/
- Semantic roles that are specific to general ideas or *frames*

PropBank

- Proposition Bank
- Available in numerous languages
  - English
  - Hindi
  - Chinese
  - Arabic
  - Finnish
  - Portuguese
  - Basque
  - Turkish

PropBank

- Provides semantic roles associated with different verb senses
- Senses are given numbered arguments as roles
  - Arg0
  - Arg1
  - …
  - ArgN
- Arg0: Generally represents a proto-agent
- Arg1: Generally represents a proto-patient
- Other arguments tend to be more verb-specific

PropBank Entries

- Referred to as **frame files**
- Definitions for each role are informal glosses

agree.01
- Arg0: Agreer
- Arg1: Proposition
- Arg2: Other entity agreeing

- Ex1: [$_{Arg0}$ The group] agreed [$_{Arg1}$ it wouldn't make an offer].
- Ex2: [$_{ArgM-TMP}$ Usually] [$_{Arg0}$ John] agrees [$_{Arg2}$ with Mary] [$_{Arg1}$ on everything].

fall.01
- Arg1: Logical subject, patient, thing falling
- Arg2: Extent, amount fallen
- Arg3: start point
- Arg4: end point, end state of arg1

- Ex1: [Arg1 Sales] fell [Arg4 to $25 million] [Arg3 from $27 million].
- [Arg1 The average junk bond] fell [Arg2 by 4.2%].

PropBank can be useful for....

- Recovering shallow semantic information
  - Inferring commonality in event structures for varying surface forms
- Representing modification or adjunct meanings
  - Denoted using non-numbered arguments called **ArgMs**
  - ArgMs aren't listed in individual frame files since they're generalizable across predicates

Common Modifier Arguments

ArgM	Description	Example
TMP	When?	Yesterday evening, now
LOC	Where?	At the museum, in Chicago
DIR	Where to/from?	Down, to Chicago
MNR	How?	Clearly, with much enthusiasm
PRP/CAU	Why?	Because, in response to the ruling

PropBank

forecast

forecast.01 - *tell the future*

FORECAST-V NOTES: In the latter example there really should be a trace in objectposition, but treebank didn't put it there. (from forecast.01-v)
FORECAST-N NOTES: Based on sentences in nouns-9998. Comparison to forecast.01-v. No VN class. Framed by Katie. (from forecast.01-n)
FORECASTING-N NOTES: Based on sentences in nouns-9998. Comparison to forecast.01-v. No VN class. Framed by Katie. (from forecasting.01-n)

Aliases:

forecast (v.)
forecasting (n.)
forecast (n.)

Roles:

ARG0-PAG: *fortune teller*
ARG1-PPT: *prediction*
ARG2-PRD: *secondary predication*

transitive

The company forecast that fourth - quarter income from continuing operations would be `` significantly '' lower than a year earlier .

missing object

Saab 's problems were underscored Friday when the company announced that its car division had a 1.2 billion kronor ($ 186.1 million) loss during the first eight months of this year , slightly worse than Saab - Scania had forecast in its first - half report last month .

args 0 and 1

its forecast for economic growth in the EC in 1989

Check out PropBank!

- Link:
  - https://propbank.github.io/
- Paper:
  - Paul Kingsbury and Martha Palmer. From Treebank to PropBank. 2002. In Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002), Las Palmas, Spain.
- PropBank is focused on verbs, but a related project also annotates nominal predicates with the same types of semantic roles:
  - NomBank: https://nlp.cs.nyu.edu/meyers/NomBank.html

Making inferences about semantic commonalities is useful….

- Even more useful: Making inferences across different verbs, or between verbs and nouns
- Potentially applicable to more situations

FrameNet

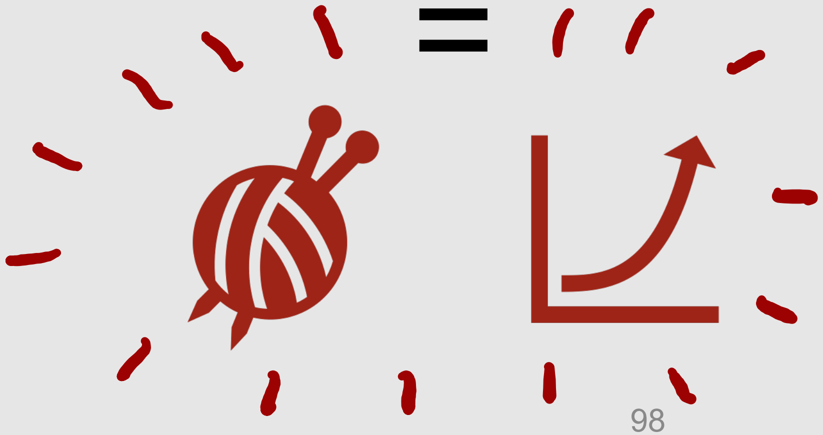
- Semantic role labeling project where roles are specific to frames rather than individual verbs
- **Frame:** A set of background information that unites a group of words

[$_{\text{Arg1}}$ The price of yarn] increased [$_{\text{Arg2}}$ 500%].

[$_{\text{Arg1}}$ The price of yarn] rose [$_{\text{Arg2}}$ 500%].

There has been a [$_{\text{Arg2}}$ 500%] rise in [$_{\text{Arg1}}$ the price of yarn].

=



Frames

- Background knowledge structures that define:
  - Specific **frame elements** associated with a given topic
  - Predicates that use these frame elements
- **Frame element:** A frame-specific semantic role

Attention

Lexical Unit Index

Definition:

This frame concerns a Perceiver's state of readiness to process and consider impressions of a Figure within a Ground. It is often unknown to the Perceiver whether or not the Figure exists within the Ground. Alternatively, the Expressor may be expressed as showing signs of the Perceiver's state of attentiveness.

Legislator tells consumers to be ALERT to dioxin levels.

They demand an ATTENTIVE gaze, a careful accounting of parts.

FEs:

Core:

Expressor [] **Excludes:** Perceiver	An entity (or event) associated with a Perceiver that gives evidence for a Perceiver's attentiveness.
Figure []	The entity that the Perceiver is specifically focussing on within the Ground.
Perceiver [] **Semantic Type:** Sentient	The individual that pays attention to the Ground.

Non-Core:

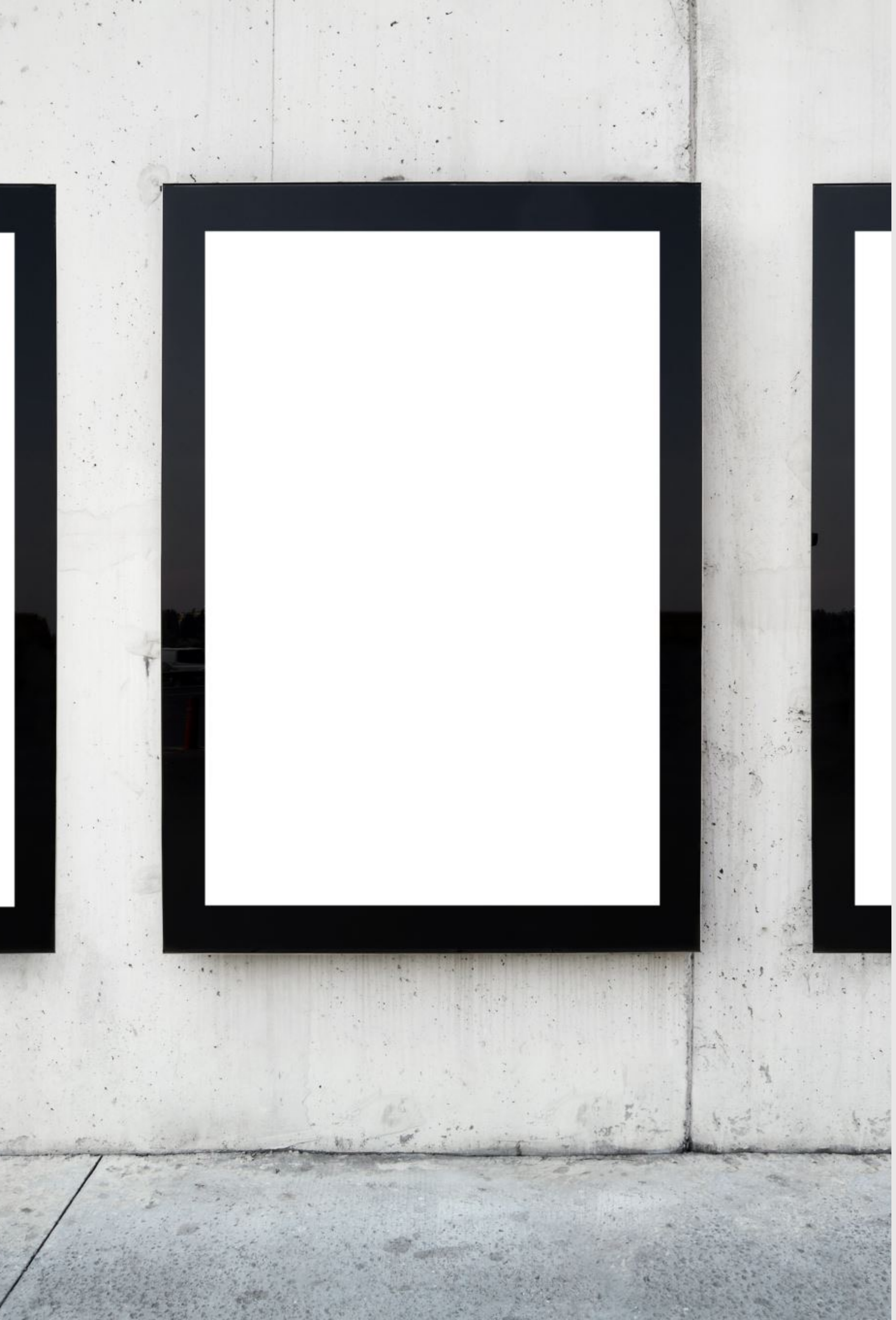
Circumstances []	The situation within which the Perceiver is alert.
Degree [] **Semantic Type:** Degree	The amount of attention that the Perceiver is paying to the Figure or Ground.
Ground []	The sensory field or subset of a sensory field that the Perceiver is attending to.
Manner [] **Semantic Type:** Manner	Any description of the event which is not covered by more specific FEs, including epistemic modification (probably, presumably, mysteriously), force (hard, softly), secondary effects (quietly, loudly), and general descriptions comparing events

Frame-frame Relations:

Inherits from: State
Is Inherited by:
Perspective on:
Is Perspectivized in:
Uses:
Is Used by: Emotions_of_mental_activity, Perception_active, Searching_scenario
Subframe of:
Has Subframe(s):
Precedes:
Is Preceded by:
Is Inchoative of:
Is Causative of:
See also:

Lexical Units:

alert.a, attend.v, attention.n, attentive.a, close.a, closely.adv, ignore.v, keep an eye.v



Frames

- Each word within a sentence or clause is understood to evoke a frame, and participate in that frame in some way
- FrameNet includes:
  - Manually specified frames and frame elements
  - Example sentences

Frame Elements

Core roles

- Frame-specific elements

Non-core roles

- More general elements
  - Time, location, etc.
- Similar to the ArgM arguments in PropBank

Example Sentences

Frame: **change_position_on_a_scale**

[ITEM Oil] rose [ATTRIBUTE in price] [DIFFERENCE by 2%].

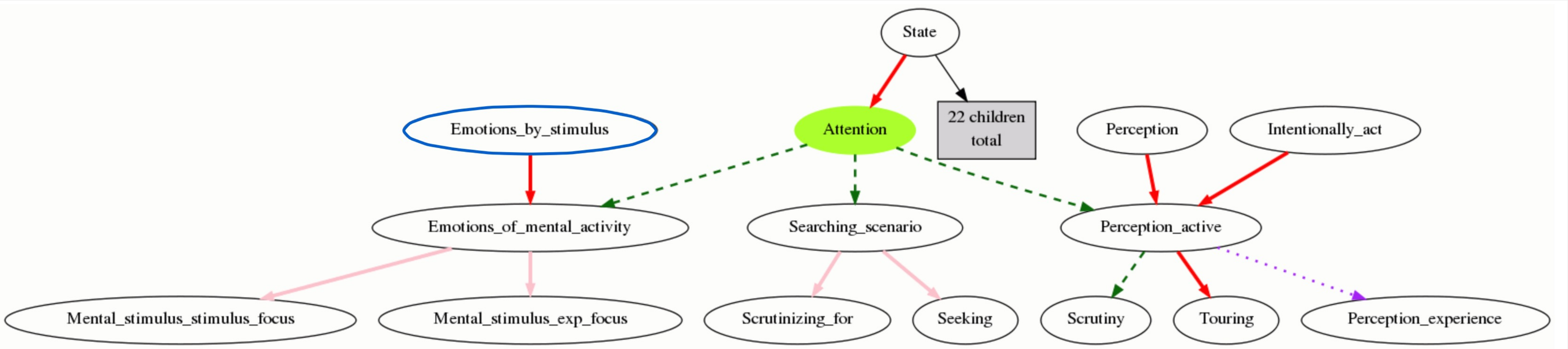
a steady increase [INITIAL VALUE from 9.5] [FINAL VALUE to 14.3] [ITEM in dividends]

[ITEM It] has increased [FINAL STATE to having them 1 day a month].

[ITEM Microsoft shares] fell [FINAL VALUE to 7 5/8].

[ITEM Colon cancer incidence] fell [DIFFERENCE by 50%] [GROUP among men].

a [DIFFERENCE 5%] [ITEM dividend] increase...



Frame Relationships

- Inheritance
- Causation

FrameNet

- Frame relationships allow us to understand common event semantics across verbal and nominal causative and non-causative uses
- FrameNet databases have been developed for a variety of languages including:
  - English
  - Spanish
  - German
  - Japanese
  - Portuguese
  - Italian Chinese
- Link:
  - https://framenet.icsi.berkeley.edu/fndrupal/
- Manual:
  - Josef Ruppenhofer, Michael Ellsworth, Miriam R. L Petruck, Christopher R. Johnson, Collin F. Baker, Jan Scheffczyk: FrameNet II: Extended Theory and Practice (Revised November 1, 2016.): https://framenet2.icsi.berkeley.edu/docs/r1.7/book.pdf

This Week's Topics

Tuesday

- Word Senses
- WordNet
- Word Sense Disambiguation

Thursday

- Semantic Roles
- Semantic Role Labeling
- Selectional Preferences

Semantic Role Labeling

- **Semantic role labeling:** Automatically assigning semantic roles to predicate arguments
- Often solved using supervised machine learning methods

The University of Illinois Chicago offered free flu shots.

? ?

How are roles defined?

- Depends on the resource!
- Often, FrameNet and/or PropBank are used to:
  - Specify predicates
  - Define roles
  - Provide training and test data

- Feature-based algorithms:
  - Parse the input string
  - Traverse the parse to find predicates
  - Decide the semantic role (if any) of each node in the parse tree with respect to each predicate
- Feature-based algorithms employ standard supervised machine learning algorithms and a wide variety of feature representations

Numerous approaches have been used to perform semantic role labeling.

Global Optimization

- Semantic roles are not independent of one another!
- Many approaches also perform a second pass to address **global consistency**
  - Constituents in FrameNet and PropBank cannot overlap
  - PropBank does not allow multiple arguments of the same type
- To choose the most globally consistent set of labels, SRL systems often include an additional step leveraging Viterbi decoding or a reranking algorithm

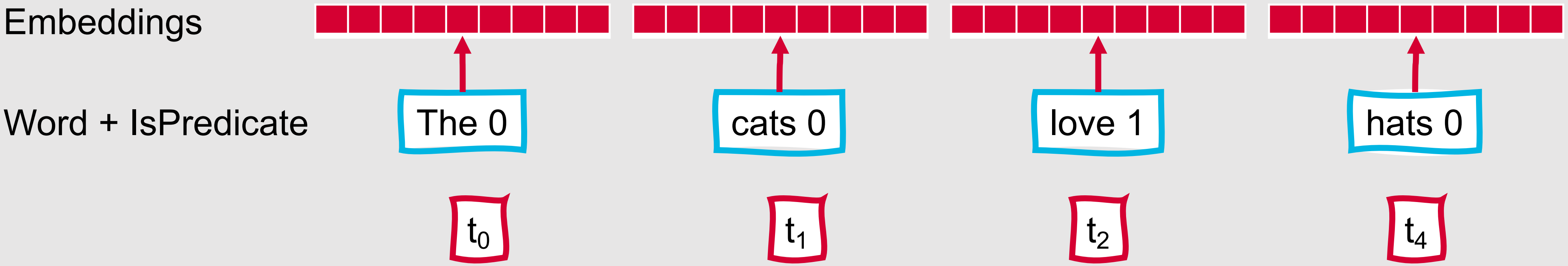
Features for Semantic Role Labeling

- Common features:
  - Governing predicate
  - Constituent type
  - Head word of the constituent
  - Part of speech of the head word
  - Path in the parse tree from the constituent to the predicate
  - Whether the voice of the surrounding clause is active or passive
  - Whether the constituent appears before or after the predicate
  - Set of expected arguments for the verb phrase
  - Named entity type of the constituent
  - First and last word(s) of the constituent

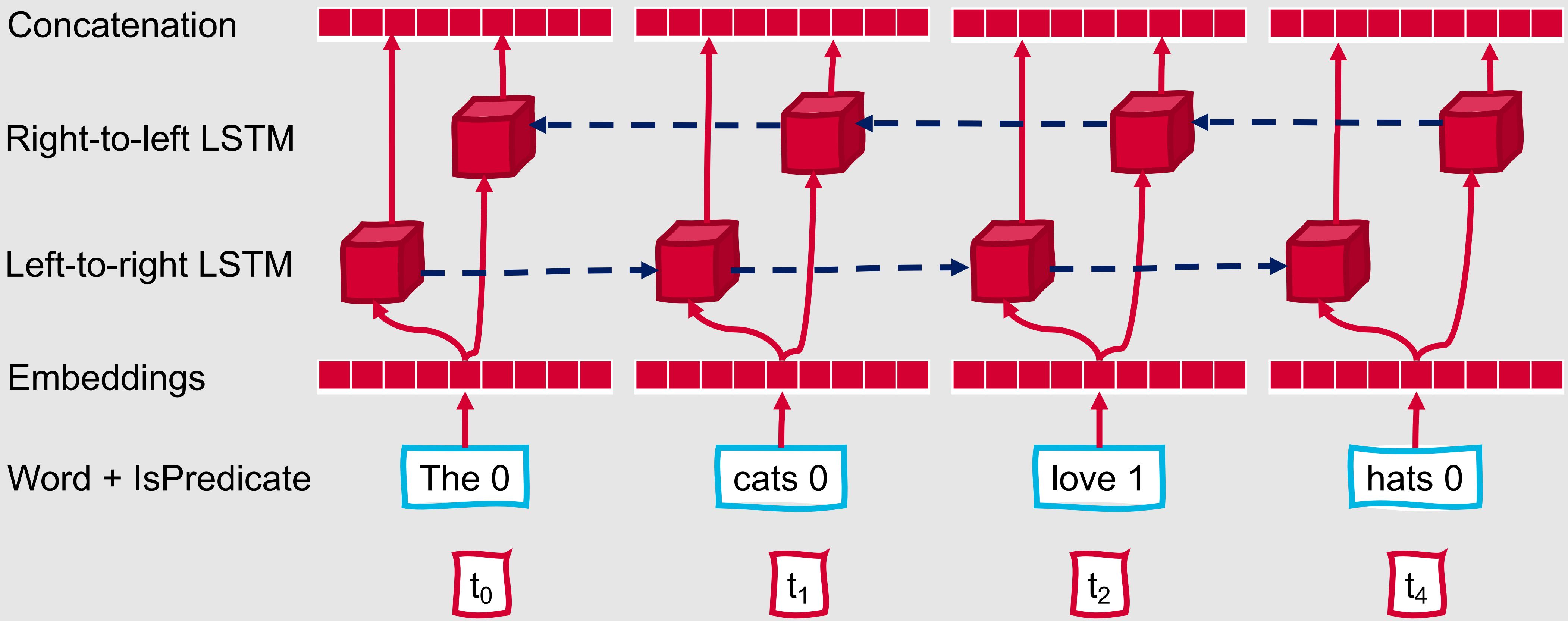
Modern SRL is also often performed using neural models.

- Frame SRL like other sequence labeling tasks
  - Given a predicate, detect and label spans with semantic roles
  - Use BIO tagging for this process
- Goal: Compute the highest probability tag sequence $\hat{y}$, given an input sequence of words w:
  - $\hat{y} = \underset{y \in T}{\operatorname{argmax}} P(y|w)$
- Global optimization can be addressed by applying Viterbi decoding directly to the softmax output

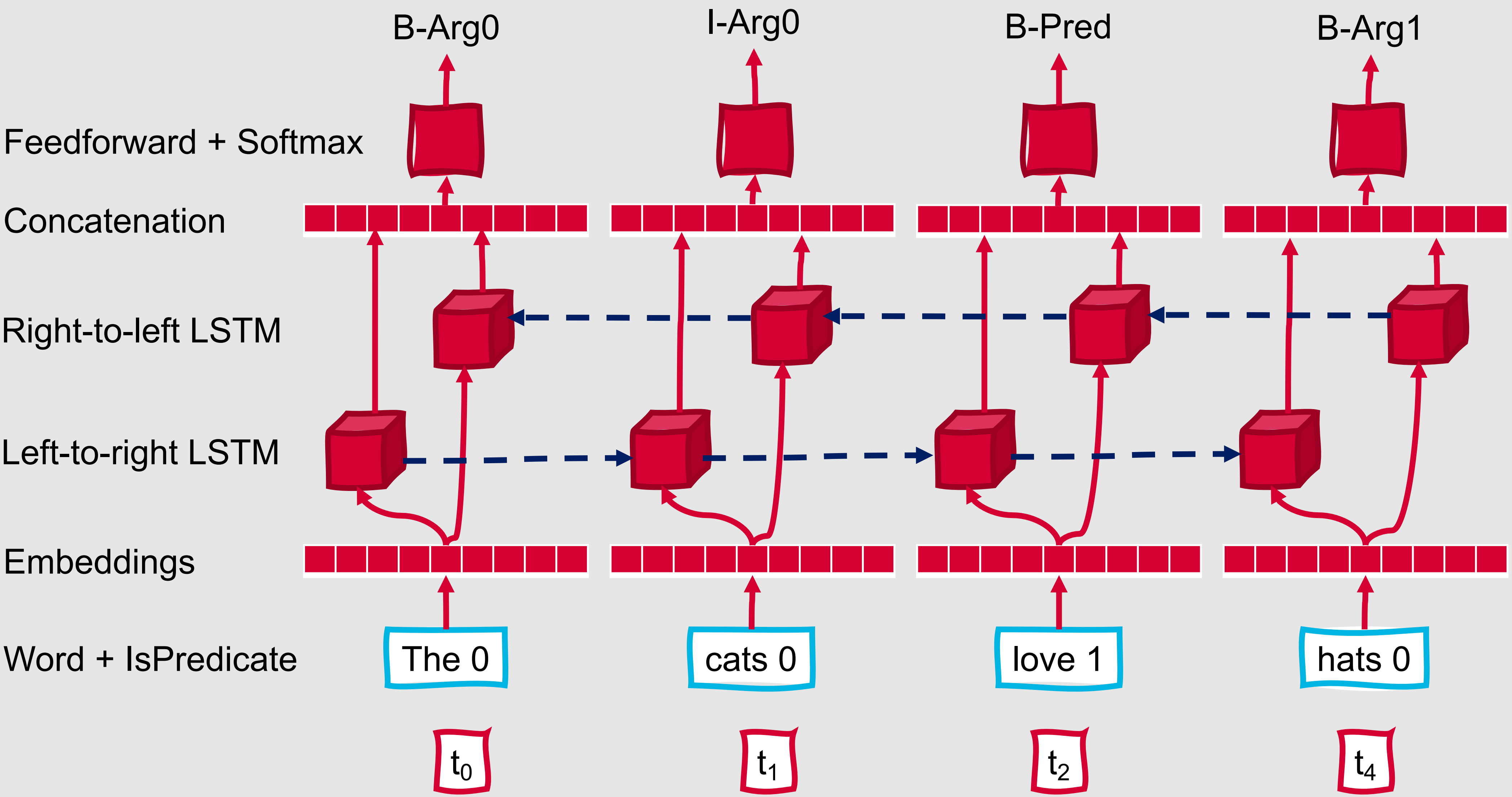
Neural Semantic Role Labeling



Neural Semantic Role Labeling



Neural Semantic Role Labeling



Evaluation of Semantic Role Labeling

- **True positives:** Argument labels assigned to the correct word sequence or parse constituents
- Then, we can compute our standard NLP metrics:
  - Precision
  - Recall
  - F-measure

This Week's Topics

Word Senses
WordNet
Word Sense Disambiguation

Tuesday

Thursday

Semantic Roles
Semantic Role Labeling
Selectional Preferences

Relationships between predicates and arguments can also be defined in other ways.

- Sometimes, there are conceptual limitations on which words can act as arguments to predicates
- We refer to these as **selectional restrictions**

What are selectional restrictions?

- **Selectional restrictions:** Semantic constraints placed upon predicates, governing the types of concepts that can fill those predicates' semantic roles

What are selectional restrictions?

- **Selectional restrictions:** Semantic constraints placed upon predicates, governing the types of concepts that can fill those predicates' semantic roles

Selectional Restrictions

- Associated with senses, not words themselves
- Vary in their specificity
  - To eat: THEME should be edible
  - To sip: THEME should be edible and liquid

Representing Selectional Restrictions

- Set of concepts needed for representing selectional restrictions is open-ended
  - Being a liquid
  - Being edible
  - …
- This makes selectional restrictions different from other ways to represent lexical knowledge
  - For example, parts of speech are finite and limited

One way to represent selectional restrictions....

- Extend the logical representations we've already seen
  - Use the same components we've used for representing events
    - Event variable
    - Predicate denoting event
    - Variables and relations for event roles

Representing Selectional Restrictions

$\exists e, x, y \; \mathrm{Eating}(e) \wedge \mathrm{Agent}(e, x) \wedge \mathrm{Theme}(e, y)$

$\exists e, x, y \; \mathrm{Eating}(e) \wedge \mathrm{Agent}(e, x) \wedge \mathrm{Theme}(e, y) \wedge \mathrm{EdibleThing}(y)$

$\exists e, x, y \; \mathrm{Eating}(e) \wedge \mathrm{Eater}(e, x) \wedge \mathrm{Theme}(e, y) \wedge \mathrm{EdibleThing}(y) \wedge \mathrm{Pizza}(y)$

Some issues with using logical representations….

Simpler formalisms can also enforce selectional restrictions with less computational overhead

Knowledge bases containing the facts needed to enforce logical rules associated with selectional restrictions aren't always available or comprehensive enough

What's another way we can represent selectional restrictions?

- WordNet synsets!
  - Selectional restriction for semantic role = one or more synsets
  - Input is considered reasonable if the word filling that semantic role is a member or hyponym of the specified synset

Selectional Preferences

- **Selectional restrictions** → hard constraints
- **Selectional preferences** → soft constraints
- Modern systems tend to use selectional preferences rather than selectional restrictions

She was way faster than everyone else …the other runners **ate her dust**.

Spit that out, you **can't eat plastic**!

Selectional Preference

- Selectional preferences, $S_P(v)$, are defined as the difference between two distributions:
  - Distribution of the expected semantic classes, $P(c)$
  - Distribution of the expected semantic classes for a specific verb, $P(c|v)$
- This difference can be quantified using **Kullback-Leibler (KL) divergence**, $D(P||Q)$:
  - $D(P||Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)}$
  - $S_P(v) = D(P(c|v)||P(c)) = \sum_c P(c|v) \log \frac{P(c|v)}{P(c)}$

Selectional Association

- **Selectional association** then indicates how much a given class contributes to a verb's overall selectional preference
  - $A_R(v, c) = \frac{1}{S_P(v)} P(c|v) \log \frac{P(c|v)}{P(c)}$

Selectional Preference via Conditional Probability

- We can also directly model the strength of association between a predicate and its nominal argument using conditional probability
  - Probability of an argument noun *n* given a predicate verb *v* for a particular relation *r*
- Can be computed using modified maximum likelihood estimates
  - $P(v|n,r) = \begin{cases} \frac{C(n,v,r)}{C(n,r)} \text{ if } C(n,v,r) > 0 \\ 0 \text{ otherwise} \end{cases}$
- Log co-occurrence frequency can also be used instead of the full conditional probability

How do we evaluate selectional preferences?

- **Pseudoword task**
  - Determine which of two words are more preferred by a given verb, and compute how often the selectional preference model makes the correct choice
- **Human selectional preference scores**
  - Check correlation between human selectional preference scores and those predicted by the model

Summary: Semantic Role Labeling

- **Semantic roles** define argument roles with respect to a predicate
- **PropBank** and **FrameNet** also define various general and specific semantic role types
- **Semantic role labeling** is the task of automatically assigning semantic roles to words or spans of words in a specific context
- **Selectional restrictions** are hard constraints placed upon the semantic properties of arguments
- **Selectional preferences** are soft constraints placed on those properties, and can have varying **selectional association** strength