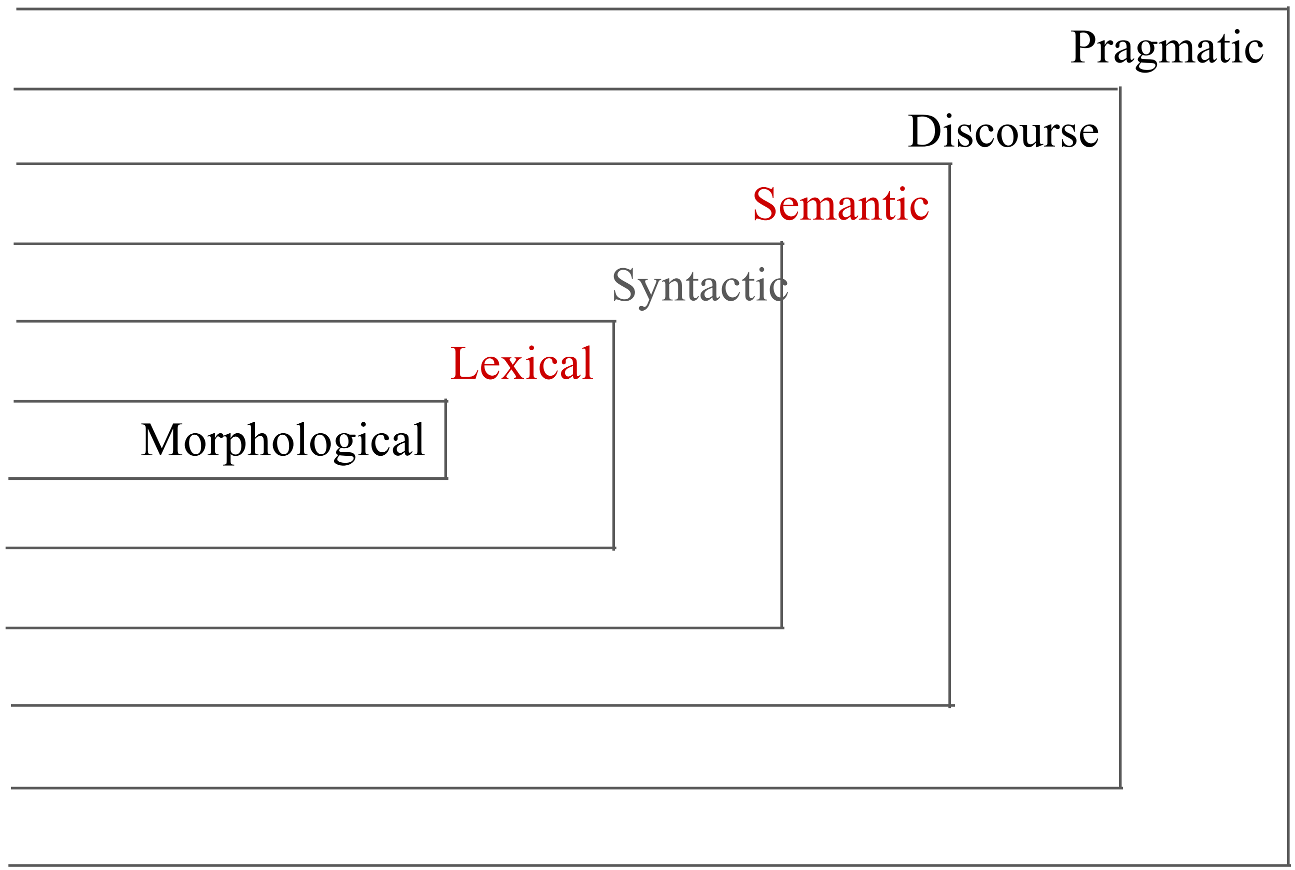
Lexical Semantics: Part 1

School of Information Studies
Syracuse University

Levels of Language



Lexical Semantics

Lexicons: words (or lexemes or stems) together with some information

Dictionaries: a lexicon with definitions for each word sense

- Most are now available online

Thesauruses: add synonyms for each word sense

- WordNet

Semantic networks: add more semantic relations, including semantic categories

Ontologies: add semantic relations and rules about entities, concepts and relations, semantic categories

Semantic lexicons: each word given value in one particular semantic category

- Sentiment lexicons

Word Senses

We say that a word has more than one word sense if there is more than one definition.

Online dictionary definitions for the noun *plant*

1. a living organism of the kind exemplified by trees, shrubs, herbs, grasses, ferns, and mosses, typically growing in a permanent site, absorbing water and inorganic substances through its roots, and synthesizing nutrients in its leaves by photosynthesis using the green pigment chlorophyll
2. a place where an industrial or manufacturing process takes place

Word senses may be:

- Coarse-grained, if not many distinctions are made
- Fine-grained, if there are many distinctions of meanings

WordNet

WordNet is a database of facts about words.

- Meanings and the relations among them
- As a semantic network, we use it to illustrate aspects of lexicons

Words are organized into clusters of synonyms called synsets.

- http://wordnet.princeton.edu/

Words are organized into nouns, verbs, adjectives, and adverbs.

- Currently 170,000 synsets
- More developed for nouns and verbs
- Available for download, arranged in separate files (DBs)

Dictionary

For each word in the language vocabulary, a dictionary provides:

- A list of meanings
- Definitions (for all word meanings)
- Typical usage examples (for most word meanings)

WordNet definitions (called glosses)/examples for synsets of the noun *plant*

1. buildings for carrying on industrial labor; “they built a large plant to manufacture automobiles”
2. a living organism lacking the power of locomotion
3. 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”
4. an actor situated in the audience whose acting is rehearsed but seems spontaneous to the audience

Thesauruses: Synonyms

A thesaurus adds:

- An explicit synonymy relation between word meanings

WordNet synsets for the noun "plant"

1. plant, works, industrial plant
2. plant, flora, plant life

Semantic Network: Relations

A semantic network adds relations for each word sense.

- Hypernymy/hyponymy (**is-a**)
  - Hypernyms are more general; hyponyms are more specific
- Meronymy/holonymy (**part-of**),
- Antonymy, entailment, etc.

WordNet related concepts for the meaning "plant life"

{plant, flora, plant life}

hypernym: {organism, being}

hypomym: {house plant}, {fungus},…

meronym: {plant tissue}, {plant part}

holonym: {Plantae, kingdom Plantae, plant kingdom}

More detailed WordNet list for nouns from Jurafsky and Martin

Relation	Also Called	Definition	Example
Hypernym	Superordinate	From concepts to superordinates	*breakfast* [1] → *meal* [1]
Hyponym	Subordinate	From concepts to subtypes	*meal* [1] → *lunch* [1]
Instance hypernym	Instance	From instance to their concepts	*Austen* [1] → *author* [1]
Instance hyponym	Has – instance	From concepts to concept instances	*composer* [1] → *Bach* [1]
Member meronym	Has – member	From groups to their members	*faculty* [2] → *professor* [1]
Member holonym	Member – of	From members to their groups	*copilot* [1] → *crew* [1]
Part meronym	Has – part	From wholes to parts	*table* [2] → *leg* [3]
Part holonym	Part – of	From parts to wholes	*course* [7] → *meal* [1]
Substance meronym		From substance to their subparts	*water* [1] → *oxygen* [1]
Substance holonym		From parts of substance to wholes	*gin* [1] → *martini* [1]
Antonym		Semantic opposition between lemmas	*leader* [1] ⇔ *follower* [1]
Derivationally Related form		Lemmas with same morphological root	*destruction* [1] ⇔ *destroy* [1]

School of Information Studies
Syracuse University

WordNet Verb Relations

A more detailed list for verbs from Jurafsky and Martin

Relation	Definition	Example
Hypernym	From events to superordinate events	*fly*^{9} → *travel*^{5}
Troponym	From events to subordinate event (often via specific manner)	*walk*^{1} → *stroll*^{1}
Entails	From verbs (events) to the verbs (events) they entail	*snore*^{1} → *sleep*^{1}
Antonym	Semantic opposition between lemmas	*increase*^{1} ⟺ *decrease*^{1}
Derivationally Related form	Lemmas with same morphological root	*destroy*^{1} ⟺ *destruction*^{1}

WordNet Hierarchies

Notes: “person” has two different hypernyms based on person viewed as a biological organism or by its functionality as causal agent

Topmost object is “entity”

```
Sense 3
bass, basso --
(an adult male singer with the lowest voice)
=> singer, vocalist, vocalizer, vocaliser
   => musician, instrumentalist, player
      => performer, performing artist
         => entertainer
            => person, individual, someone...
               => organism, being
                  => living thing, animate thing,
                     => whole, unit
                        => object, physical object
                           => physical entity
                              => entity
               => causal agent, cause, causal agency
                  => physical entity
                     => entity
```

Lexical Semantics: Part 2

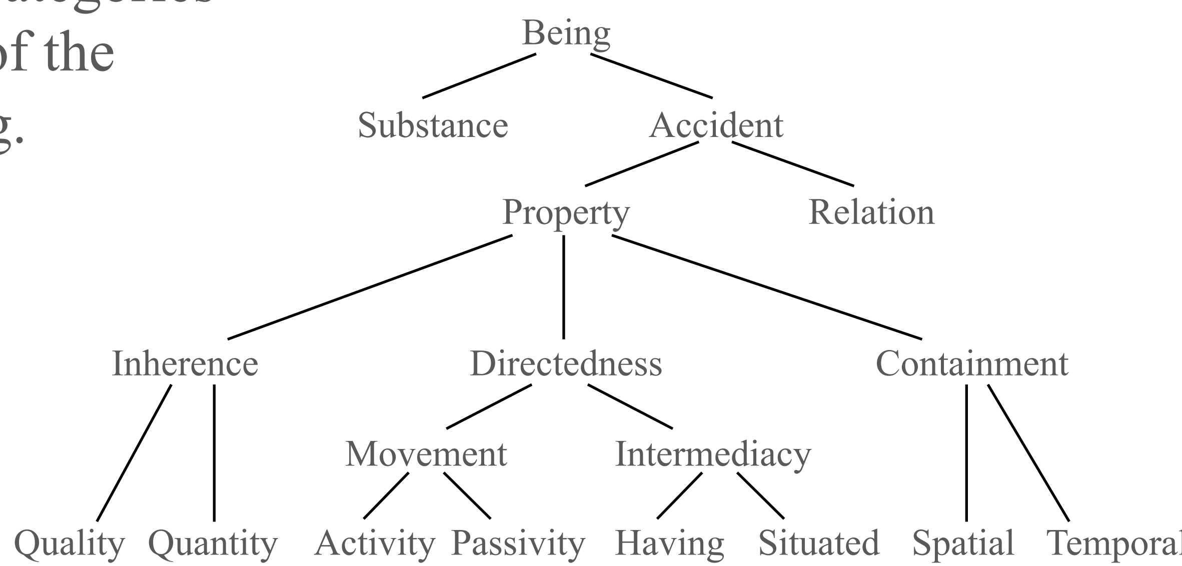
School of Information Studies
Syracuse University

Origins of Ontology

In philosophy, ontology studies the existence/being of the world.

- We can think of ontology as categorizing everything in the world.

In his work "categories," Aristotle listed ten categories to which all things of the world should belong.



http://www.jfsowa.com/talks/ontology.htm

Ontology in Information Science

Ontology is an approach of knowledge organization.

In general, ontologies are about the **representations of semantics.**

- **Concepts** (e.g., person, animal, food, table, movie, etc.)
- **Instances** (or entities) (e.g., Barack Obama is an instance of the concept "person")
- **Properties** (e.g., a person has properties of gender, height, weight, father, mother, etc.)
- **Relations** (e.g., Syracuse University is located in Syracuse)
- **Rules between concepts, properties, and relations** (e.g., if someone is married, then he/she should have a spouse)

Among the relations, concepts can be used as **categories** of instances and other concepts.

Ontology Example: UMLS

The Unified Medical Language System (UMLS) aggregates various controlled vocabularies and mapped them to a comprehensive biomedical ontology. It has three knowledge sources:

1. Metathesaurus: mapping concepts and terms in different thesauruses and organizing them in the UMLS structure
2. Semantic network: connecting semantic types of concepts in metathesaurus by semantic relations
3. Specialist lexicon: containing lexical information of biomedical terms

This is an example of a word- and phrase-level resource.

It is online but not publicly available.

Semantic Lexicons

Lexicon where each word is assigned to a semantic class or category

Lexical resources have been developed to assign words to semantic classes in support of applications that need to detect opinion, sentiment, or other more subjective meanings

Three examples given here for **sentiment lexicons**; additional examples will be given when we cover sentiment analysis

Sentiment: Subjectivity Lexicon

Subjectivity lexicon from the MPQA project with Jan Wiebe

- Gives a list of 8,000+ words that have been judged to be weakly or strongly positive, negative, or neutral in subjectivity
- Examples

type=weaksubj len=1 word1=**abandoned** pos1=adj stemmed1=n priorpolarity=negative

type=weaksubj len=1 word1=abandonment pos1=noun stemmed1=n priorpolarity=negative

type=weaksubj len=1 word1=abandon pos1=verb stemmed1=y priorpolarity=negative

type=strongsubj len=1 word1=**abase** pos1=verb stemmed1=y priorpolarity=negative

type=strongsubj len=1 word1=abasement pos1=anypos stemmed1=y priorpolarity=negative

type=strongsubj len=1 word1=abash pos1=verb stemmed1=y priorpolarity=negative

type=weaksubj len=1 word1=abate pos1=verb stemmed1=y priorpolarity=negative

type=strongsubj len=1 word1=**absolve** pos1=verb stemmed1=y priorpolarity=positive

type=strongsubj len=1 word1=absolute pos1=adj stemmed1=n priorpolarity=neutral

Semantic Classes: LIWC

Linguistic Inquiry and Word Count (LIWC): http://www.liwc.net/

- Text analysis software based on dictionaries of word dimensions
- Dimensions can be syntactic
  - Pronouns, past-tense verbs
- Dimensions can be semantic
  - Social words, affect, cognitive mechanisms
- Other categories
  - See http://liwc.wpengine.com/compare-dictionaries/
  - James Pennebaker, University of Texas at Austin

Often used for positive and negative emotion words in opinion mining

Semantic Classes: ANEW

Affective Norms for English Words (ANEW)

- Provides a set of emotional ratings for a large number of words in the English language

Participants gave graded reactions from 1–9 on three dimensions

- Good/bad, psychological valence
- Active/passive, arousal valence
- Strong/weak, dominance valence

From the NIMH Center for the Study of Emotion and Attention at the University of Florida

- http://csea.phhp.ufl.edu/Media.html
- See also the paper by Dodds and Danforth on Happiness of Large-Scale Written Expressions

Lexical Semantics: Word Sense Disambiguation

School of Information Studies
Syracuse University

Word Sense Disambiguation (WSD)

We need to look up words in lexicons to find semantics, but which meaning (sense) of the word should we use?

Definition of word sense disambiguation

- Correct selection of the appropriate sense/meaning of a polysemous word in context

In English, the most frequently occurring nouns have seven senses, and the most frequently occurring verbs have 11 senses

Defining a Word Sense

How can we define different word senses?

- Give a list of synonyms.
- Give a definition, which will necessarily use words that will have different senses, and these will (perhaps circularly) use words for definitions.

Coarse-grained senses distinguish core aspects of meaning.

Fine-grained senses also distinguish peripheral aspects of meaning.

- Example: In WordNet, which has fairly fine-grained senses, the word "bass" has five senses related to music and three senses related to fish (and the sense determines how you pronounce the word).

Difficulties With Synonym Lists

True synonyms are non-existent or very rare

Near-synonyms (Edmonds and Hirst)

- Examples
  - Error, blunder, mistake
  - Order, command, bid, enjoin, direct
- Dimensions of synonym differentiation
  - Stylistic variation
    - Pissed, drunk, inebriated
  - Expressive variation
    - Attitude: skinny, thin, slim
    - Emotion: father, dad, daddy
  - . . .

Human Sense Disambiguation

Sources of influence known from psycholinguistics research

- Local context (most important)
  - The sentence or other surrounding text containing the ambiguous word restricts the interpretation of the ambiguous word
    - Example: *book* in a sentence that has flight, travel, etc.
- Domain knowledge
  - The fact that a text is concerned with a particular domain activates only the sense appropriate to that domain
    - Example: *plant* in a biology article
- Frequency data
  - The frequency of each sense in general usage affects its accessibility to the mind

Lesk Algorithm

Original Lesk definition: measure overlap between sense definitions for all words in context (Michael Lesk, 1986)

- Identify simultaneously the correct senses for all words in context

Simplified Lesk (Kilgarriff & Rosensweig, 2000): measure overlap between sense definitions of a word and current context

- Identify the correct sense for one word at a time
- Current context is the set of words in the surrounding sentence/paragraph/document

Lesk Algorithm: A Simplified Version

Algorithm for simplified Lesk:

1. Retrieve from lexicon all sense definitions of the target word.
2. Determine the overlap between each sense definition and the current context.
3. Choose the sense that leads to highest overlap.

Example: disambiguate PINE in

"Pine cones hanging in a tree"

- PINE
  1. kinds of evergreen tree with needle-shaped leaves
  2. waste away through sorrow or illness

Pine#1 $\cap$ Sentence = 1

Pine#2 $\cap$ Sentence = 0

WSD Algorithm Development in Senseval

All-word task

- Given an entire text, disambiguate every content word in the text
- Use general-purpose lexicon with senses
- Can use a labeled corpus
  - SemCor is a sub-set of the Brown Corpus with 234,000 words labeled with
    WordNet senses
  - Additional corpora developed through Senseval

Sense-Tagged Corpus

Examples of text where words are annotated with their sense from WordNet

Bonnie and Clyde are two really famous criminals. I think they were **bank/1** robbers.
My **bank/1** charges too much for an overdraft.
I went to the **bank/1** to deposit my check and get a new ATM card.
The University of Minnesota has an East and a West **Bank/2** campus right on the Mississippi River.
My grandfather planted his pole in the **bank/2** and got a great big catfish!
The **bank/2** is pretty muddy. I can’t walk there.

Classification Approach to WSD

Train a classification algorithm that can label each (open-class) word with the correct sense, given the context of the word.

Training set is the hand-labeled corpus of senses.

The context is represented as a set of "features" of the word and includes information about the surrounding words.

- Typical features shown on next slides

The result of training is a model that is used by the classification algorithm to label words in the test set and, ultimately, in new text examples.

- In the Senseval conferences, a number of systems in the range of 70–80% accuracy for English lexical sample task

Word Similarity Features

Important in these classifications is that, instead of just using words that overlap in context, look for "similar" words

For each word, compute a similarity measure between that word and the words in the definitions to be disambiguated

Similarity measures

- Can be defined from a semantic relation lexicon, such as WordNet
- One example is path similarity
  - For any two words, gives a number between 0 and 1 based on the shortest path between the two words in the WordNet hypernym/hyponym hierarchy
    - For example, the words "plant" and "tree" should have a shorter path through a common ancestor than words like "plant" and "piano"

WSD Classification Features

Collocational features from the target word

- Information about words in specific positions (i.e., previous word)
- Typical features include the word itself, its stem, and its POS tag

Similar word features

- Whether words surrounding the target word are "similar" to those in the word definition

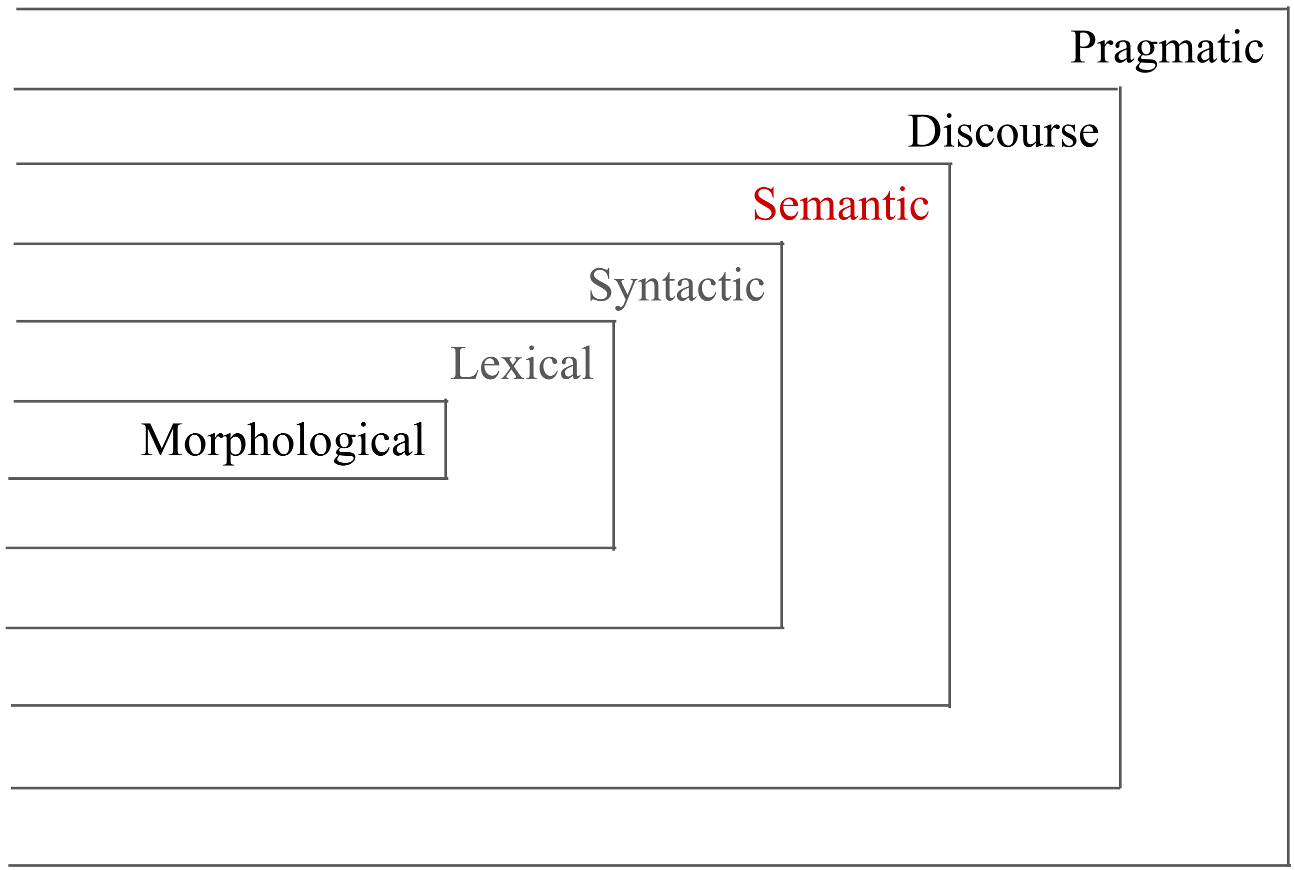
Associated word features

- For each word to be disambiguated, collect a small number of frequently used context words from a large corpus
- Example: for each word, collect the 12 most frequent context words
- For *bass*, the 12 context words from the *Wall Street Journal* are:
  - [fishing, big, sound, player, fly, rod, pound, double, runs, playing, guitar, band]

Semantic Processing

School of Information Studies
Syracuse University

Levels of Language



School of Information Studies
Syracuse University

Semantic Processing

Implement the human ability to understand the meaning of sentences in their language.

- Given text (or speech), give a semantic representation of the meaning

Semantic processing attempts to build representations of the meaning of sentences in a decompositional way based on the syntax of the language.

- The interpretation of syntactic words
- Semantic relations between words, with selectional restrictions
  - The agent of the verb “kick” must be something active
- Example: syntactic phrase structure should map to semantic structure

Related Tasks for Semantic Processing

Detect **semantics ambiguities**. If a sentence is two ways ambiguous, characterize the meaning of each reading.

The bill is large.

- It may include word sense disambiguation and semantic similarity of words.
  - Senses of bill: the bill of a bird, paper money—and both can be large

Decide if one sentence is a **paraphrase** of another (two way).

- "Your marks on the tests were excellent."
- "You scored very high on the exams."

Entailment: Decide if the truth of one sentence implies the truth of another (one way).

- "John lives in Toronto" implies that John's residence is in Canada.

Relation Between Syntax and Semantics in NLP

Syntactic analysis:

- Determines the syntactic category of the words
- Decides phrase structure—how words are grouped
- Assigns structural analysis to a sentence

Semantic analysis:

- Creates a representation of the meaning of a sentence

Relation Between Syntax and Semantics in NLP

Clearly syntactic structure affects meaning (e.g., word order, phrase attachment).

- “The man with the telescope watched Mary.”
- “Mary watched the man with the telescope.”

But meaning can determine syntactic structure.

- Recall that lexicalized statistical parsing used head word affinities (probabilities) to help determine parsing.

Semantic Systems and Representation

A semantic system consists of different types of building blocks: entities, concepts, relations, and predicates.

A representation shows how to put together entities, concepts, relations, and predicates to describe a situation or "semantic world."

- Enables reasoning about that semantic world

Why Do We Need Semantic Representations?

To link the surface, linguistic elements to the non-linguistic knowledge of the world

- Many words, few concepts

To represent the variety at the lexical level at a unified conceptual level

- Unambiguous representations; canonical forms

Structures composed from a set of symbols

- All languages have a predicate–argument structure
- Corresponds to relationships that hold among concepts underlying constituent words and phrases of a sentence, then across sentences

Can be used to reason, both to verify what is true in the world and to infer knowledge from the semantic representation

Building Blocks of Semantic Systems

Semantics that words (or base noun phrases) represent—the objects

- **Entities:** individuals such as a particular person, location, or product
  - John F. Kennedy, Washington, D.C., Cocoa Puffs
- **Concepts:** the general category of individuals
  - Person, city, breakfast cereal

Building Blocks of Semantic Systems

Semantics indicated by verbs, prepositional phrases, and other structures

- **Relations** between entities and concepts
  - John F. Kennedy "is-a" person
- **Relations** between entities or between concepts
  - Hierarchy of specific to more general concepts
  - Wide variety of other relations
- **Predicates** representing verb structures, sometimes called events
  - Semantic roles, case grammar
  - Can also be used for relations between objects

Semantic Representations

School of Information Studies
Syracuse University

Semantic Representation Approaches

Some possible knowledge representation approaches include:

- First-order logic
- Semantic nets
- Conceptual dependency
- Frames
- Rule-based
- Conceptual graphs
- Case grammar

First-Order Logic for Semantic Representation

Also known as predicate calculus

A symbolic language whose symbols have precisely stated meanings and uses

- The symbols can be used as meanings in the real world
- Typically express properties of entities in the world

Example: if Socrates is a man, then Socrates is a mortal

man (Socrates) -> mortal (Socrates)

First-order logic (FOL) often used in AI systems found in such applications as robotics and computational control systems

- Allows a natural language interface to such systems
- Systems have automatic reasoning to make decisions or supply information

FOL Language

FOL uses terms to represent objects in the real world.

- Constants are specific objects in the world—entities
  - Socrates, Pastabilities
- Functions represent concepts about objects
  - LocationOf(Pastabilities)
    - Note the value of a function is a concept or entity
- Variables are used to stand for any object
  - X

FOL Language

FOL uses predicates to state relations between objects.

- Note that the value of a predicate is true or false representing facts in the world.
- “IsRestaurant” could be a predicate that, when applied to an object, returns true if it is a restaurant.
  - IsRestaurant(Pastabilities)
- If “Serves” is a predicate taking a restaurant and a type of food as arguments, we can state that a restaurant serves a type of food.
  - Serves(Pastabilities, VegetarianFood)

FOL Language, Operations, and Quantifiers

FOL uses connectives “and” and “or” to combine statements.

- Serves(Pastabilities, VegetarianFood) ^ IsExpensive(Pastabilities)

FOL uses the implication connection to mean that, if the first statement is true, then the second one is also true.

- Serves(Pastabilities, VegetarianFood) => IsRestaurant(Pastabilities)
  - Is this true?

FOL Language, Operations, and Quantifiers

FOL uses the existential quantifier to assert that an object with particular properties exists

- $\exists$ X IsRestaurant(X) ^ Serves(X, VegetarianFood)

FOL uses the universal quantifier to assert that particular properties are true for all objects (using $\forall$ for the "forall" symbol).

- $\forall$ X IsRestaurant(X) => Serves(X, VegetarianFood)
- This is definitely false because not all restaurants serve vegetarian food

Reasoning With FOL

FOL allows inference to make conclusions of new information

- Inference rule is called "modus ponens" and informally is if-then reasoning;
  if we know that A is true and we know that A => B is true, we can conclude that B is true

This type of inference has efficient implementations to allow systems to reason from facts in the semantic world or in text

- For example, reasoning could find answers for a question-answering system
  " Find me a restaurant serving Mexican food near the warehouse"
- Find the X such that

IsRestaurant(X) ^ Serves(X, MexicanFood) ^ Near(LocationOf(X)), TheWarehouse)

Difficulties With First-Order Logic

Problem for NLP is that "semantics" of logic does not necessarily equate to "meaning" in the real world

- Not everything is as clear-cut as required by a formal logic

May not be enough "real-world" predicates in the FOL system to capture semantics of text

- This is a problem for all the semantic representations

Semantic systems better developed for objects and actions

- Not as well developed to represent ideas and beliefs

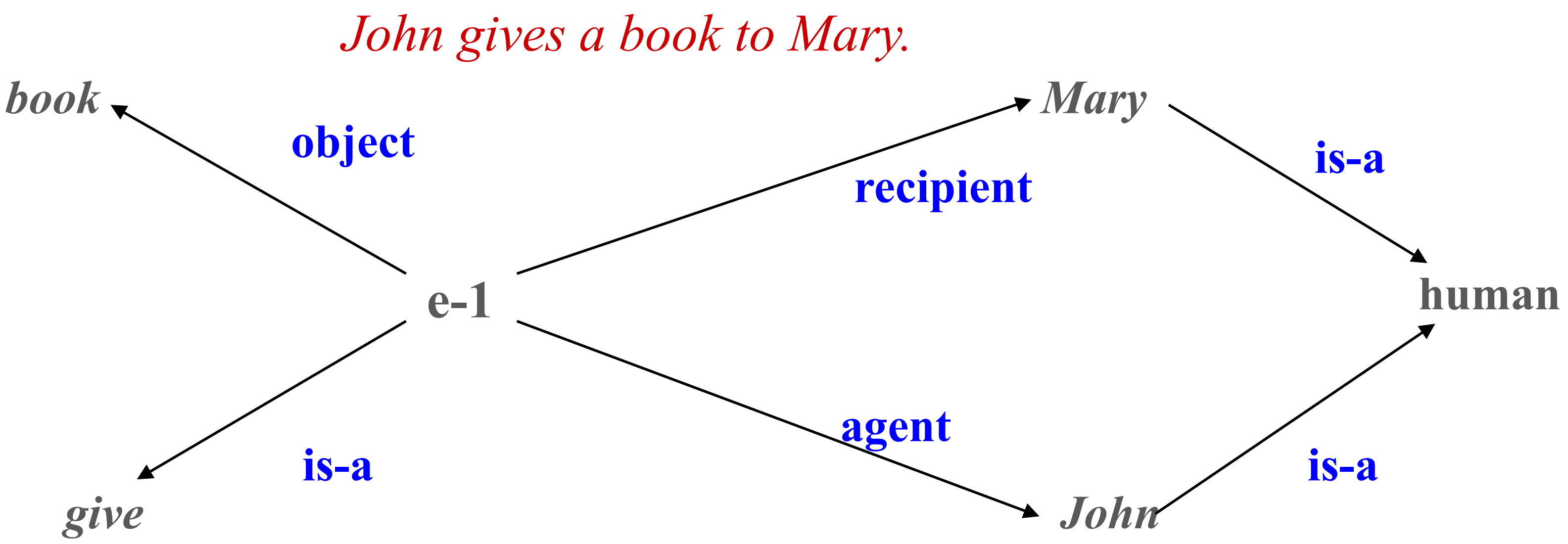
See Cycorp efforts to embody all world knowledge in (essentially) first-order logic in their "knowledge base"

- http://www.cyc.com/kb/
- Captures properties of the real world (e.g., coffee cups have handles), that you assume everyone knows

Semantic Networks

A network or graph of nodes joined by links where:

- Nodes represent **concepts** (book, human) and entities (John, Mary)
- Links (labeled, directed arcs) represent **relations** (e.g., **is-a**)



Frames

Frames are a type of structured representation or schema and are widely used for knowledge organization

- Introduced by Marvin Minsky in 1975, "A Framework for Representing Knowledge"
  - Most widely referenced paper on knowledge representation; explicitly attempts to represent human processing

Frames group information about an entity or an event in terms of a record of "slots" and "fillers"

- Each object has a frame with slots
- One slot filled by the name of the object
- Other slots filled with a property or relation and the value of the property or the entity that is related

Graph structure of concepts (frames) and links (slot relations)

Example of Frames

Wikipedia infobox is an example of a frame structure

- Slot names are properties or relations
- A property value is information such as a date or height
- A relation value is another entity, which may have its own frame
  - e.g. Country of origin has value Norway

Norwegian Lundehund

A Norwegian Lundehund

Other names	Norsk Lundehund Norwegian Puffin Dog
Nicknames	Lundehund
Country of origin	Norway

Traits [hide]

Weight	6–7 kilograms (13–15 lb)
Height	30–40 centimetres (12–16 in)

Classification / standards [hide]

FCI	Group 5 Section 2 #265	standard
AKC	Non-Sporting	standard
UKC	Northern Breed	standard

Dog (*Canis lupus familiaris*)

Example of Frames

More formal frame systems (such as those for information extraction) require uniformity of slot names and value syntax

Name	Barack Obama
Birthdate	August 4, 1961
Birthplace	Honolulu, Hawaii
Height	6’ 1” (1.85 m)
Parents	Barack Obama Sr., Ann Dunham
Children	Natasha Obama, Malia Ann Obama

Reasoning with frames can use FOL:

($\exists$ X) (Name(X) = Barack Obama) ^ Birthplace(X) = Honolulu) etc.

Applications of Semantic Representations

Paraphrase task: two sentences map to the same semantic representation (Microsoft Research Paraphrase Corpus).

Entailment task: the semantics of the first sentence implies the semantics of the second under reasoning (EDITS: http://edits.fbk.eu/).

Semantic representations are used to represent entities with their properties and relations in information extraction and question-answering systems.

Semantic representations are used in reasoning in AI systems such as robot manipulations.

- Could also be used in dialog systems
- Works best in a small environment where the amount of world knowledge needed is small

Getting Semantic Representation From Text

Use a syntactic parse tree to identify predicates and possible relations structures.

Algorithms map syntactic structure to relations, given the words in the text.

- Semantic role labeling is one important algorithm (next section).
- Some systems employ a first-order logic mapper.
- Watson (IBM question answering system) mapped dependency parses to frames.

Semantic Roles

School of Information Studies
Syracuse University

Semantics of Events in Sentences

In a sentence, a verb and its semantic roles form a proposition; the verb can be called the predicate and the roles are known as arguments.

When Disney offered to pay Mr. Steinberg a premium for his shares, the New York investor didn't demand the company also pay a premium to other shareholders.

Example semantic roles for the verb "pay" (using verb-specific roles) show which noun and prepositional phrases are related to the verb

When [$_{\text{payer}}$ Disney] offered to [$_{\text{V}}$ pay] [$_{\text{recipient}}$ Mr. Steinberg] [$_{\text{money}}$ a premium] for [$_{\text{commodity}}$ his shares], the New York investor…

Case Grammar

Fillmore, Charles, "The Case for Case," 1968.

- A response to Chomsky's disregard for any semantics
  - "A semantically justified syntactic theory"

Given a sentence, it is possible to say much more than this NP is the subject and this NP is the object.

Chomsky's transformational grammar would reduce active and passive versions of the same deep structure but doesn't go far enough to reveal why this is possible semantically.

- *A crowbar could open that door easily.*
- *That door could be opened easily with a crowbar.*

Case Grammar

Focuses on conceptual events

- For each event or situation, there is a limited number of roles/cases that people or objects play in the situation
- Roles reflect ordinary human judgments about:
  - Who did the action?
  - Who/what was it done to?
  - What was it done with?
  - Where was it done?
  - What was the result?
  - When was it done?

Case roles show semantic structures, not syntactic structures

Semantic Structure Is Not the Same as Syntactic Structure

Syntactic similarities hide semantic dissimilarities.

- ***We*** *baked every Saturday morning.*
- ***The pie*** *baked to a golden brown.*
- ***This oven*** *bakes evenly.*

- Three subject NPs perform very different roles in regard to "bake."
  - "We" is the agent, "This oven" is the instrument, "The pie" can be the theme (or similar).

Syntactic dissimilarities hide semantic similarities.

- *John*$_{agent}$ *broke the window*$_{theme}$.
- *John*$_{agent}$ *broke the window*$_{theme}$ *with a rock*$_{instrument}$.
- *The rock*$_{instrument}$ *broke the window*$_{theme}$.
- *The window*$_{theme}$ *broke.*
- *The window*$_{theme}$ *was broken by John*$_{agent}$.
  - Note that the last two sentences are "passive voice."

Cases
(a.k.a. Thematic Roles or Theta Roles)

Some of Fillmore's original set of roles still in use as general descriptors of roles:

- **Agentive (A)**
  - The instigator of the action, an animate being
    - John opened the door.
    - The door was opened by John.
- **Instrumental (I)**
  - The thing used to perform the action, an inanimate object
    - The key opened the door.
    - John opened the door with the key.

Cases
(a.k.a. Thematic Roles or Theta Roles)

Some of Fillmore's original set of roles still in use as general descriptors of roles:

- **Locative (L)**
  - The location or spatial orientation of the state or action of the verb
    - It's windy in Chicago.

Other original roles not typically used

Verb-Specific Roles

General thematic roles don't work for many verbs and roles.

- Many general sets are proposed; not uniform agreement
- Generalized semantic roles now often called:
  - Proto roles: proto-agent, proto-patient, etc.
  - Or, theta roles

Verb-specific roles are proposed in treebanks.

- PropBank annotates the verbs of Penn Treebank
  - Extended with NomBank for nominalizations
- FrameNet annotates the British National Corpus
  - Uses domains of semantically similar verbs called frames

Semantic Role Labeling (SRL) Treebanks

School of Information Studies
Syracuse University

PropBank

PropBank is a corpus with annotation of semantic roles, capturing the semantic role structure of each verb sense.

- Funded by ACE to Martha Palmer and Mitch Marcus at U Penn

Each verb sense has a frameset, listing possible semantic roles.

- Argument notation uses numbers for the annotation, for example, first sense of accept (accept.01)
  - Arg0: acceptor
  - Arg1: thing accepted
  - Arg2: accepted-from
  - Arg3: attribute

The frameset roles are standard across all syntactic realizations in the corpus of that verb sense.

- Each verb has a frameset file describing the args as above with example texts.

Roles Consistent With VerbNet

PropBank builds on VerbNet to assign more specific roles

VerbNet is one extension of Levin's verb classes, giving semantic roles from about 20 possible roles

- Agent, Patient, Theme, Experiencer, etc., similar to theta roles

Whenever possible, the PropBank argument numbering is made consistent for all verbs in a VerbNet class

- There is only 50% overlap between PropBank and VerbNet verbs

Example from frameset file for "explore," which has a VN class:

```
<roleset id="explore.01" name="explore, discover new places or things" vncls="35.4">
<roles> <role descr="explorer" n="0">
                <vnrole vncls="35.4" vntheta="Agent"/></role>
        <role descr="thing (place, stuff) explored" n="1">
                <vnrole vncls="35.4" vntheta="Location"/></role>
</roles>
```

Semantic Role Notation for PropBank

The first two numbered arguments correspond, approximately, to the core case roles.

- Arg0—Prototypical Agent
- Arg1—Prototypical Patient or Theme
- Remaining numbered args are verb-specific case roles, Arg2 through Arg5

Another large group of roles is the adjunctive roles (which can be applied to any verb). They are annotated as ArgM with a suffix.

- ArgM-LOC—location
- ArgM-EXT—extent
- ArgM-DIR—direction
- ArgM-ADV—general adverbial
- ArgM-DIS—discourse connective
- ArgM-MOD—modal verb

ArgM-CAU—cause
ArgM-TMP—time
ArgM-PNC—purpose
ArgM-MNR—manner
ArgM- NEG—negation

Adjunctive and Additional Arguments

Example of adjunctive arguments

- Not all core arguments are required to be present, for example, Arg2 below.
- Arguments can be phrases, clauses, even partial words.

When Disney offered to pay Mr. Steinberg a premium for his shares, the New York investor didn't demand the company also pay a premium to other shareholders.

Example of PropBank annotation (on demand):

[$_{\text{ArgM-TMP}}$ When Disney offered to pay Mr. Steinberg a premium for his shares], [$_{\text{Arg0}}$the New York investor] did [$_{\text{ArgM-NEG}}$ n' t] [v demand] [$_{\text{Arg1}}$ the company also pay a premium to other shareholders].

Where for demand, Arg0 is "asker," Arg1 is "favor," Arg2 is "hearer"

PropBank Annotations

Framesets were created by looking at sample sentences containing each verb sense

- ~4,500 frames (in 3,314 framesets for each verb)

Corpus is primarily newswire text from Penn Treebank

- Annotated the *Wall Street Journal* section and, more recently, the "Brown" Corpus
- Verbs and semantic role annotations added to the parse trees by human annotators

PropBank Annotations

Annotators are presented with **role-set descriptions** of a verb and the (gold) **syntactic parses** of a sentence in Treebank, and they annotate the roles of the verb

- Lexical sampling: annotated on a verb-by-verb basis
- ~40,000 sentences were annotated

Interannotater agreement

- Identifying argument and classifying role: 99% accuracy
  - Another agreement measure is the Kappa statistic, where they achieved .91 overall and .93 if ArgMs excluded, which is high

FrameNet

Project at International Computer Science Institute with Charles Fillmore

- http://framenet.icsi.berkeley.edu/

Similar goal to document the syntactic realization of arguments of predicates in the English language

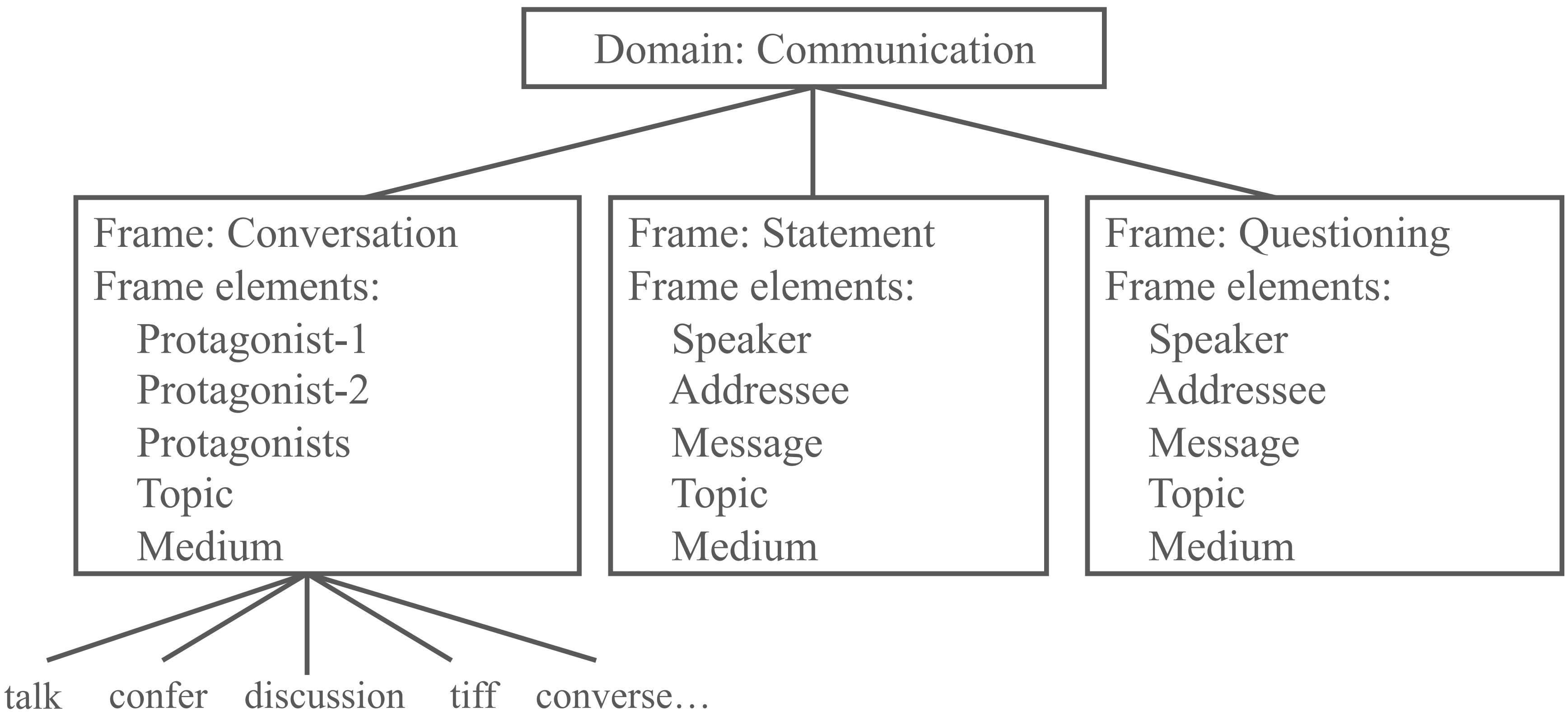
Starts from semantic frames (e.g., Commerce) and defines frame elements (e.g., Buyer, Goods, Seller, Money)

Annotates example sentences chosen to illustrate all possibilities

- But one release includes 132,968 sentences
- British National Corpus

Example of FrameNet Frames

Semantic frames are related by topic domain



Comparison of FrameNet and PropBank

FrameNet semantic roles are consistent for semantically related verbs (not just synonyms as in PropBank).

- Commerce examples

FrameNet annotation:

[$_{\text{Buyer}}$ Chuck] *bought* [$_{\text{Goods}}$ a car] [$_{\text{Seller}}$ from Jerry][$_{\text{Payment}}$ for $1000].
[$_{\text{Seller}}$ Jerry] *sold* [$_{\text{Goods}}$ a car] [$_{\text{Buyer}}$ to Chuck] [$_{\text{Payment}}$ for $1000].

Propbank annotation:

[$_{\text{Arg0}}$ Chuck] *bought* [$_{\text{Arg1}}$ a car] [$_{\text{Arg2}}$ from Jerry][$_{\text{Arg3}}$ for $1000].
[$_{\text{Arg0}}$ Jerry] *sold* [$_{\text{Arg1}}$ a car] [$_{\text{Arg2}}$ to Chuck] [$_{\text{Arg3}}$ for $1000].

Frame for buy:
Arg0: buyer
Arg1: thing bought
Arg2: seller
Arg3: price paid
Arg4: benefactive

Frame for sell:
Arg0: seller
Arg1: thing sold
Arg2: buyer
Arg3: price paid
Arg4: benefactive

SRL Task

School of Information Studies
Syracuse University

Automatic SRL

Define an algorithm that will process text and recognize roles for each verb.

Assume previous levels of Natural Language Processing (NLP) on text.

- Part-of-speech (POS) tagging
- Parse trees or dependency trees

Machine-learning classification approaches are typical.

Machine-Learning Approach

Given a verb in a sentence, the problem is to find and label all arguments

Reformulate as a classification task: For each constituent in the parse tree of the sentence, label it as to what argument, if any, it is for the verb

For each constituent, define features of semantic roles

- Each feature describes some aspect of a text phrase that can help determine its semantic role of a verb
  - Examples include what the verb is, POS tags, position in parse tree, etc.

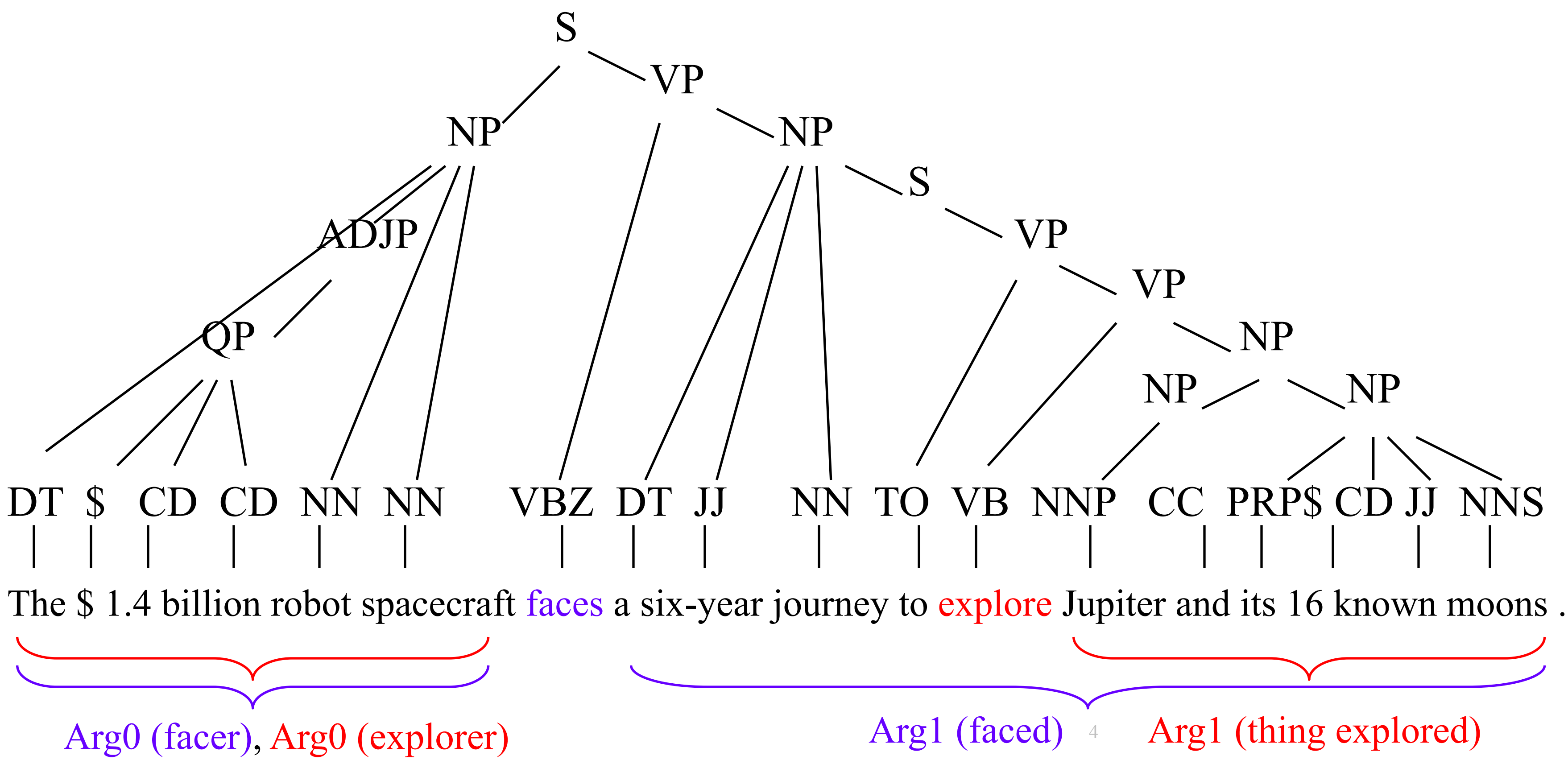
Machine-learning process

- Training a classifier on treebank annotated with semantic roles
  - PropBank or FrameNet
- Can then classify syntactic phrases as to their roles

Parse Tree Constituents

Each syntactic constituent is a candidate for labeling.

Define features from parse tree with part-of-speech tags on words.



Difficulties for Classification

For each verb in a sentence, the number of constituents in the parse tree are large compared with the number of semantic roles.

- Can be hundreds of constituents eligible to be labeled a role
- Leads to the problem of too many “negative” examples

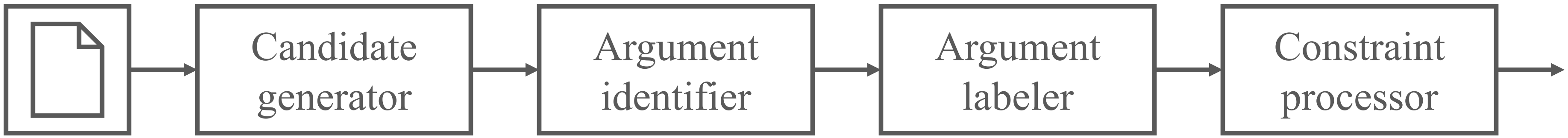
What should the features be?

- Words are typically the features for an NLP problem
  - Sometimes called bag of words (BOW)
- Need more about the syntactic structure as well as other potential clues
- Typical number of features can be up to 20,000, requiring a classification algorithm that is robust for large numbers of features

Typical Architecture With Two-Step Classifier

Steps of the architecture

- Candidate generator: filter out implausible constituents from the parse trees
- Argument identifier: use a machine-learning classifier to decide if each of the remaining constituents is an argument to the verb
- Argument labeler: N binary classifiers, each producing a probability estimate of whether an argument should have that label (Arg0–Arg5, ArgM's, etc.)
- Do some final constraint processing



Typical Argument Features

These features are defined for each constituent:

Predicate: the predicate word from the training data

- "face" and "explore"
- Usually stemmed or lemmatized

Phrase type: the phrase label of the argument candidate

- Examples are NP, S for phrases, or may be POS tag if a single word

Position: whether the argument candidate is before or after the predicate

Voice: whether the predicate is in active or passive voice

- Passive voice is recognized if a past participle verb is preceded by a form of the verb "be" within three words

Argument Features

Subcategory: the phrase labels of the children of the predicate's parent in the syntax tree

- Subcategory of "faces" is "VP -> VBZ NP"

Path: the syntactic path through the parse tree from the argument constituent to the predicate

- Arg0 for "faces": NP -> S -> VP -> VBZ

Head word: the head word of the argument constituent

- Main noun of NP (noun phrase)
- Main preposition of PP (prepositional phrase)

Many additional features

- Head word POS: the part-of-speech tag of the head word of the argument constituent
- Temporal cue words: special words occurring in ArgM-TMP phrases
- Etc.

SRL Problem Constraints

Results of the labeling classifier are probabilities for each label of whether it labels that constituent.

- Constraints combine the results of the classification for each label.

Use these with constraints to assign a label.

- Two constituents cannot have the same argument label.
- A constituent cannot have more than one label.
- If two constituents have (different) labels, they cannot have any overlap.
- No argument can overlap the predicate.

CoNLL-2005 Shared Task

Each year, CoNLL (Conference on Natural Language Learning) defines a task to develop some aspect of natural language processing with systems that use machine learning.

- Provides data for training and developing systems for three months
- Then provides test data; everyone runs their system and returns the results for scoring
- Competitive in that scores are published in a comparative way
- Collaborative in that a session of the annual conference is devoted to discussion of the progress in this task

The 2005 shared task evaluated machine learning SRL systems based on full parse information. Best results:

	Precision	Recall	$F_{\beta=1}$
Koomen et al.	80.05%	74.83%	77.35

Current Direction of SRL

Best English SRL results combining parse trees or combining the parsing task with the SRL task (joint inference) are at F-measure of 80–82

- Similar to other semantic tasks with performance in low 80s

CoNLL 2009 shared task is SRL again, but systems combined **dependency parsing** with semantic role labeling

- English, Catalan, Chinese, Czech, German, Japanese, Spanish

Question: **Can applications make good use of SRL?**

- SRL tools are not as generally available as good parsing tools.
- Results are not as accurate as POS tagging (~97) or parsing (~92).
- But there are systems requiring the semantics in general domain text that have used SRL to give semantic representations.
  - IBM Watson Question Answering system