Catching Up and Getting into Semantics

Announcements

- My office hours will be Wednesday 2–3 this week instead of Wednesday 4–5 (Oct. 7th)

Today

- The homework
- NLTK demo
- Finishing up probabilistic parsing
- Getting into semantics

Getting the Probabilities

- From an annotated database (a treebank)
  - So for example, to get the probability for a particular VP rule just count all the times the rule is used and divide by the number of VPs overall.

TreeBanks

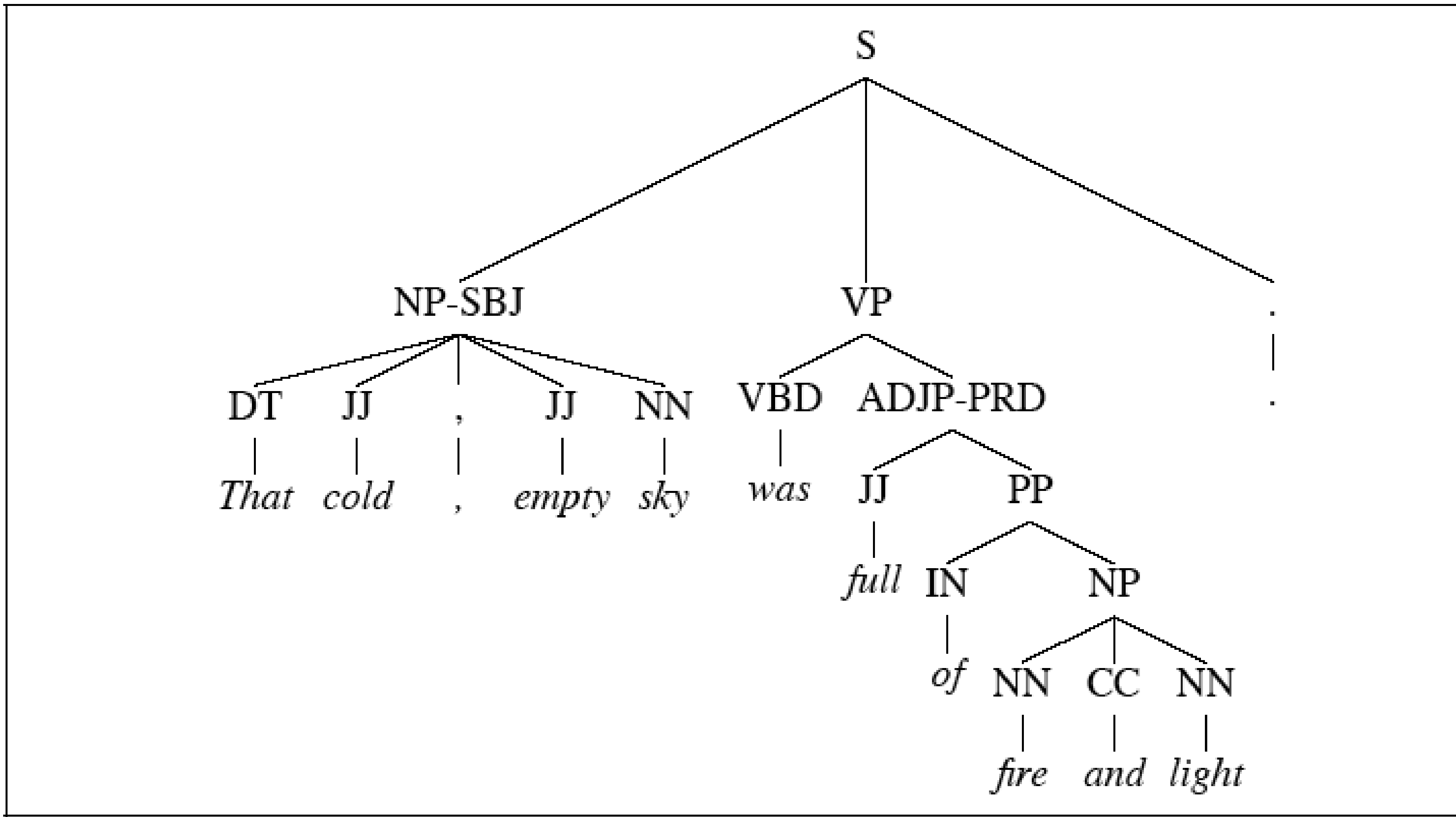
```
((S
   (NP-SBJ (DT That)
     (JJ cold) (, ,)
     (JJ empty) (NN sky) )
   (VP (VBD was)
     (ADJP-PRD (JJ full)
       (PP (IN of)
         (NP (NN fire)
           (CC and)
           (NN light) ))))
   (. .) ))
```

(a)

```
((S
   (NP-SBJ The/DT flight/NN )
   (VP should/MD
     (VP arrive/VB
       (PP-TMP at/IN
         (NP eleven/CD a.m/RB ))
       (NP-TMP tomorrow/NN )))))
```

(b)

Treebanks


S
NP-SBJ
VP
.
DT
JJ
,
JJ
NN
VBD
ADJP-PRD
.
That
cold
,
empty
sky
was
JJ
PP
full
IN
NP
of
NN
CC
NN
fire
and
light

Treebanks

```
( (S (`` ``)
    (S-TPC-2
      (NP-SBJ-1 (PRP We) )
      (VP (MD would)
        (VP (VB have)
          (S
            (NP-SBJ (-NONE- *-1) )
            (VP (TO to)
              (VP (VB wait)
                (SBAR-TMP (IN until)
                  (S
                    (NP-SBJ (PRP we) )
                    (VP (VBP have)
                      (VP (VBN collected)
                        (PP-CLR (IN on)
                          (NP (DT those) (NNS assets) ))))))))))))
    (, ,) ('' '')
    (NP-SBJ (PRP he) )
    (VP (VBD said)
      (S (-NONE- *T*-2) ))
    (. .) ))
```

Treebank Grammars

S	→	*NP VP .*	*PRP*	→	*we* | *he*
		NP VP	*DT*	→	*the* | *that* | *those*
		" S " , NP VP .	*JJ*	→	*cold* | *empty* | *full*
		-NONE-	*NN*	→	*sky* | *fire* | *light* | *flight*
		DT NN	*NNS*	→	*assets*
		DT NN NNS	*CC*	→	*and*
		NN CC NN	*IN*	→	*of* | *at* | *until* | *on*
		CD RB	*CD*	→	*eleven*
NP	→	*DT JJ , JJ NN*	*RB*	→	*a.m*
		PRP	*VB*	→	*arrive* | *have* | *wait*
		-NONE-	*VBD*	→	*said*
VP	→	*MD VP*	*VBP*	→	*have*
		VBD ADJP	*VBN*	→	*collected*
		VBD S	*MD*	→	*should* | *would*
		VB PP	*TO*	→	*to*
		VB S			
		VB SBAR			
		VBP VP			
		VBN VP			
		TO VP			
SBAR	→	*IN S*			
ADJP	→	*JJ PP*			
PP	→	*IN NP*			

Lots of flat rules

$NP \rightarrow$ DT JJ NN
$NP \rightarrow$ DT JJ NNS
$NP \rightarrow$ DT JJ NN NN
$NP \rightarrow$ DT JJ JJ NN
$NP \rightarrow$ DT JJ CD NNS
$NP \rightarrow$ RB DT JJ NN NN
$NP \rightarrow$ RB DT JJ JJ NNS
$NP \rightarrow$ DT JJ JJ NNP NNS
$NP \rightarrow$ DT NNP NNP NNP NNP JJ NN
$NP \rightarrow$ DT JJ NNP CC JJ JJ NN NNS
$NP \rightarrow$ RB DT JJS NN NN SBAR
$NP \rightarrow$ DT VBG JJ NNP NNP CC NNP
$NP \rightarrow$ DT JJ NNS , NNS CC NN NNS NN
$NP \rightarrow$ DT JJ JJ VBG NN NNP NNP FW NNP
$NP \rightarrow$ NP JJ , JJ ‘‘ SBAR ’’ NNS

Example sentences from those rules

- Total: over 17,000 different grammar rules in the 1-million word Treebank corpus

(9.19) [$_{DT}$ The] [$_{JJ}$ state-owned] [$_{JJ}$ industrial] [$_{VBG}$ holding] [$_{NN}$ company] [$_{NNP}$ Instituto] [$_{NNP}$ Nacional] [$_{FW}$ de] [$_{NNP}$ Industria]

(9.20) [$_{NP}$ Shearson's] [$_{JJ}$ easy-to-film], [$_{JJ}$ black-and-white] "[$_{SBAR}$ Where We Stand]" [$_{NNS}$ commercials]

Probabilistic Grammar Assumptions

- We're assuming that there is a grammar to be used to parse with.
- We're assuming the existence of a large robust dictionary with parts of speech
- We're assuming the ability to parse (i.e. a parser)
- Given all that… we can parse probabilistically

Typical Approach

- Bottom–up (CKY) dynamic programming approach
- Assign probabilities to constituents as they are completed and placed in the table
- Use the max probability for each constituent going up

What's that last bullet mean?

- Say we're talking about a final part of a parse
  - $S \to {}_0NP_iVP_j$

The probability of the S is…
P(S->NP VP)*P(NP)*P(VP)

The green stuff is already known. We're doing bottom-up parsing

Max

- I said the P(NP) is known.
- What if there are multiple NPs for the span of text in question (0 to i)?
- Take the max (where?)

Problems with PCFGs

- The probability model we're using is just based on the rules in the derivation…
  - Doesn't use the words in any real way
  - Doesn't take into account where in the derivation a rule is used

Solution

- Add lexical dependencies to the scheme…
  - Infiltrate the predilections of particular words into the probabilities in the derivation
  - I.e. Condition the rule probabilities on the actual words

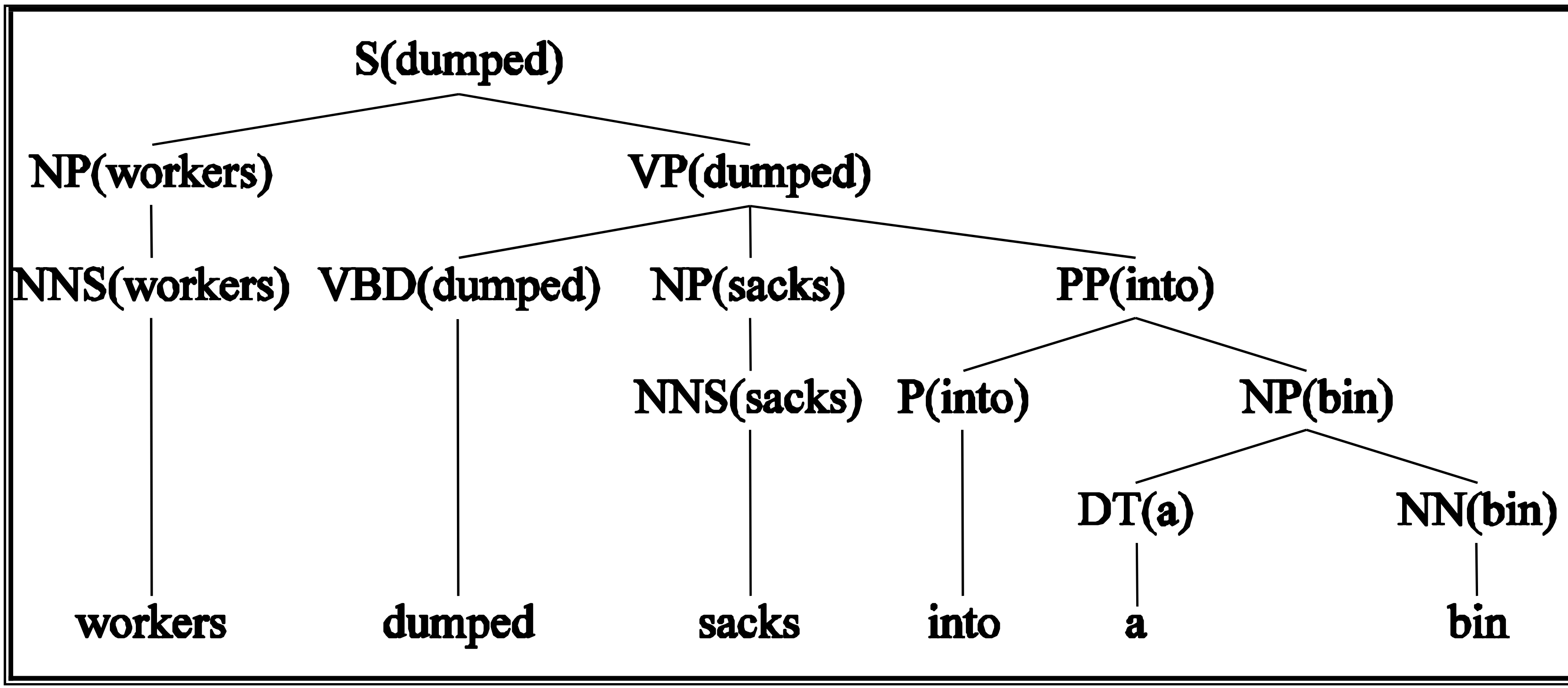
Heads

- To do that we're going to make use of the notion of the head of a phrase
  - The head of an NP is its noun
  - The head of a VP is its verb
  - The head of a PP is its preposition

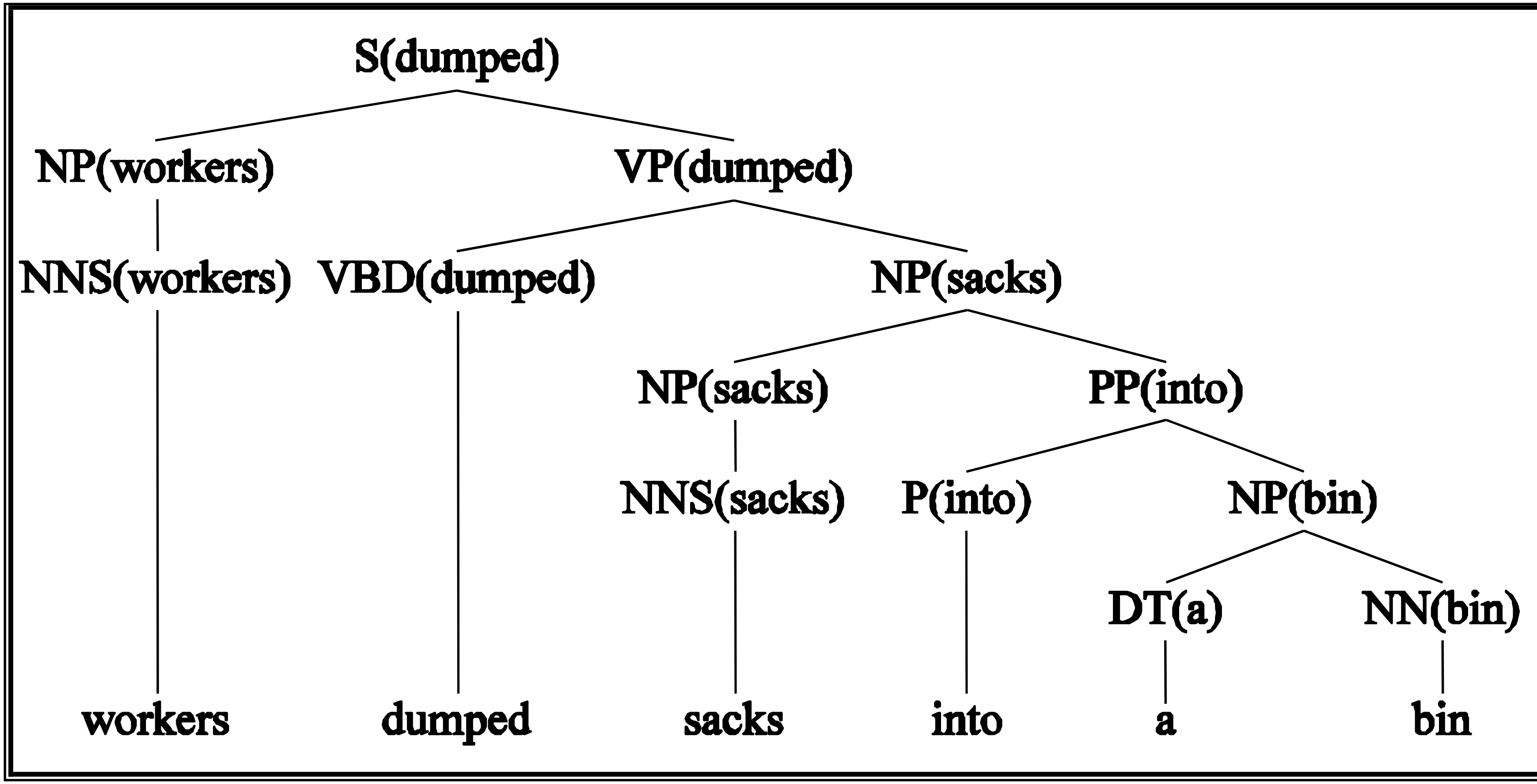
(It's really more complicated than that but this will do.)

Example (right)

Attribute grammar



Example (wrong)



How?

- We used to have
  - VP -> V NP PP   P(rule|VP)
    - That's the count of this rule divided by the number of VPs in a treebank
- Now we have
  - VP(dumped)-> V(dumped) NP(sacks)PP(in)
  - P(r|VP ^ dumped is the verb ^ sacks is the head of the NP ^ in is the head of the PP)
  - Not likely to have significant counts in any treebank

Declare Independence

- When stuck, exploit independence and collect the statistics you can…
- We'll focus on capturing two things
  - Verb subcategorization
    - Particular verbs have affinities for particular VPs
  - Objects affinities for their predicates (mostly their mothers and grandmothers)
    - Some objects fit better with some predicates than others

Subcategorization

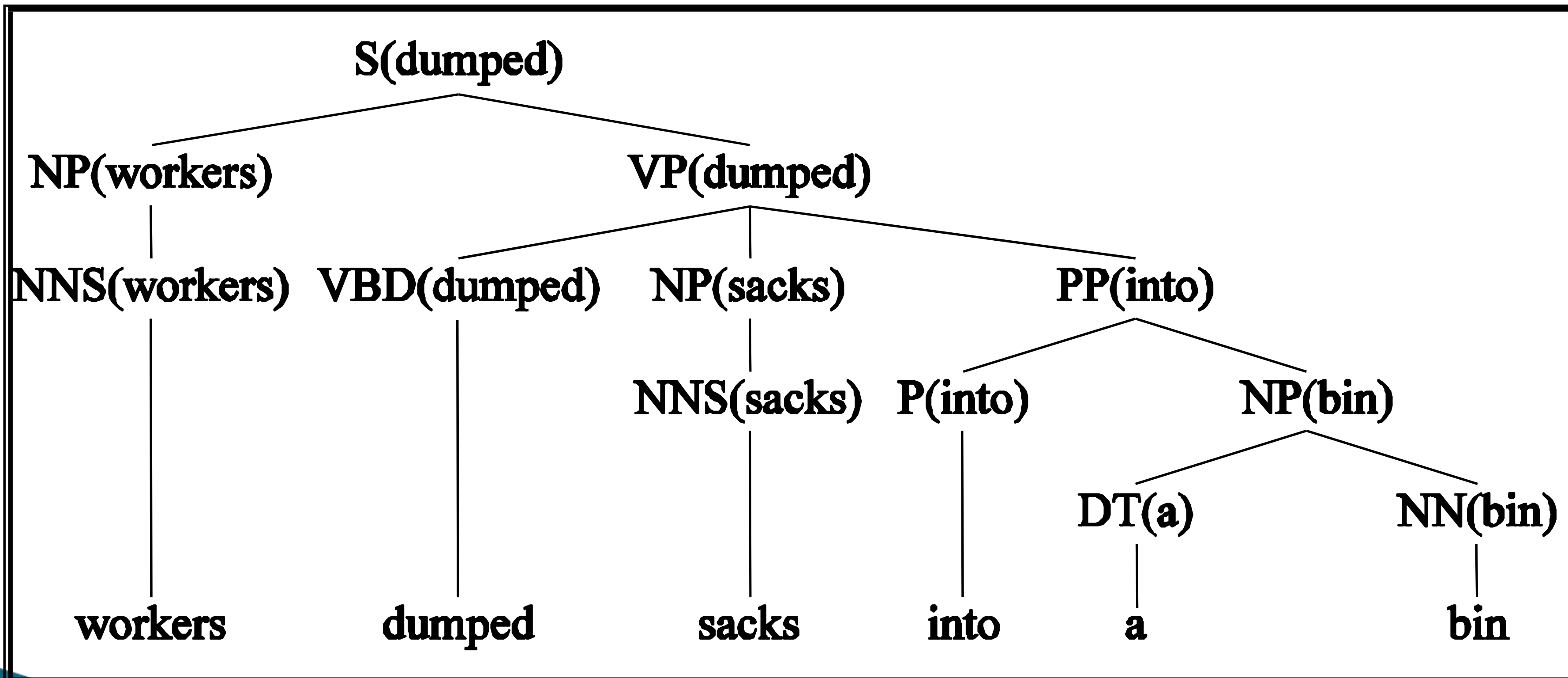
- Condition particular VP rules on their head… so
  r: VP -> V NP PP P(r|VP)
  Becomes
  P(r | VP ^ dumped)

  What's the count?
  How many times was this rule used with (head) dump, divided by the number of VPs that dump appears (as head) in total

  Think of left and right modifiers to the head

Example (right)

Attribute grammar



Probability model

$$P(T,S) = \prod_{n \in T} p(r_n)$$

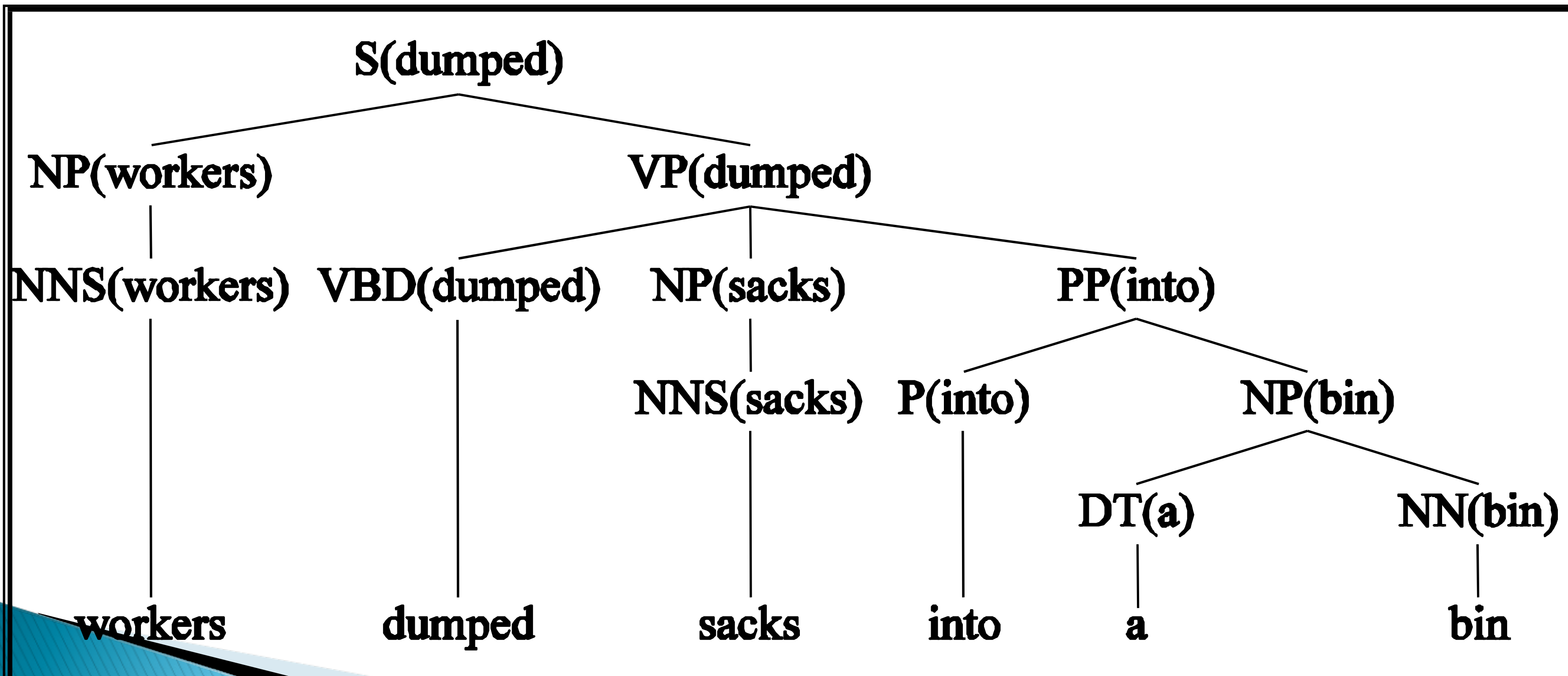
- P(T,S) = S-> NP VP (.5)*
- VP(dumped) -> V NP PP (.5) (T1)
- VP(ate) -> V NP PP (.03)
- VP(dumped) -> V NP (.2) (T2)

- What about VP -> VP PP?

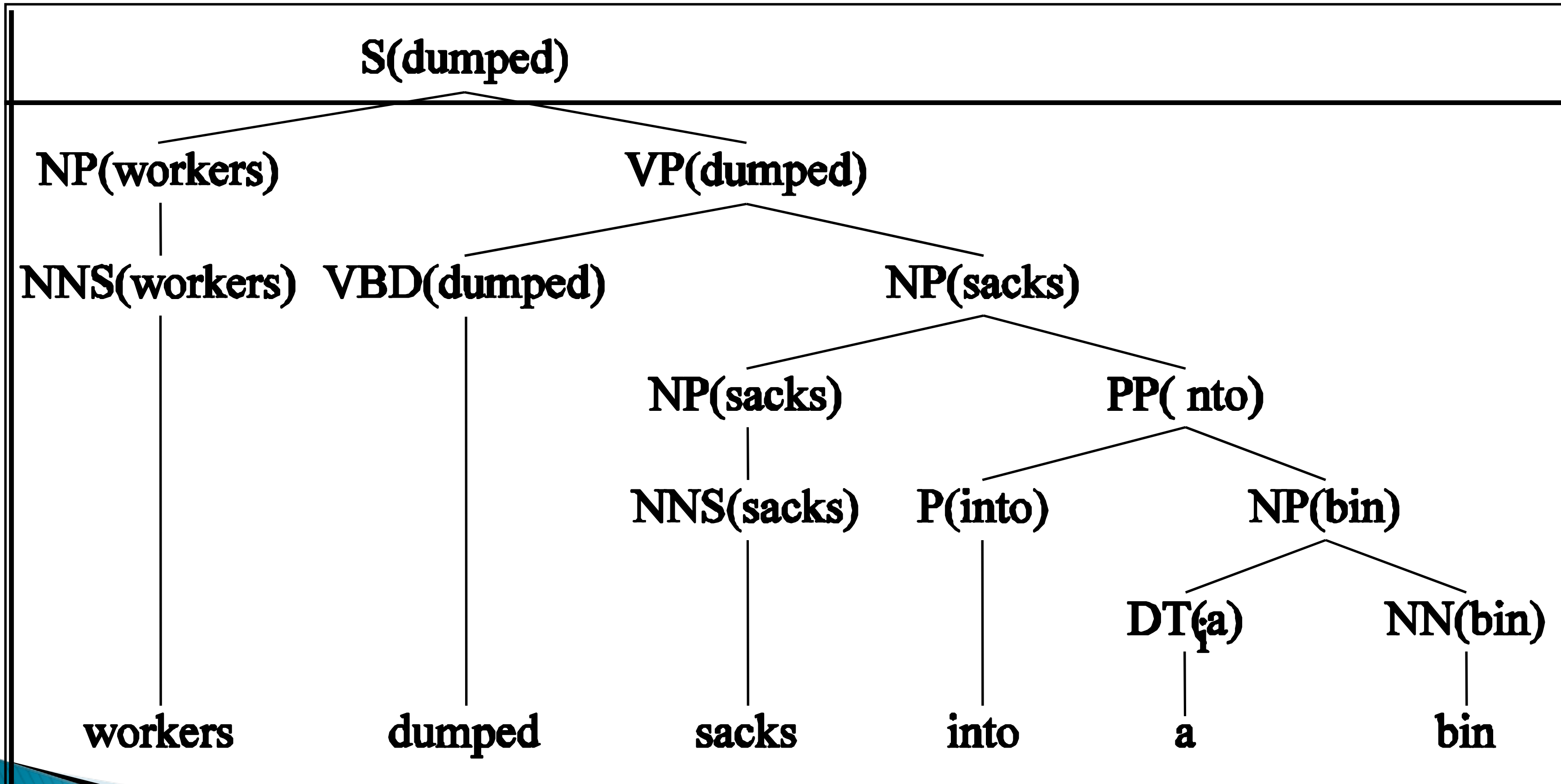
Preferences

- Subcategorization captures the affinity between VP heads (verbs) and the VP rules they go with.
- What about the affinity between VP heads and the heads of the other daughters of the VP
- Back to our examples…

Example (right)



Example (wrong)



Preferences

- The issue here is the attachment of the PP. So the affinities we care about are the ones between dumped and into vs. sacks and into.
- So count the places where dumped is the head of a constituent that has a PP daughter with into as its head and normalize
- Vs. the situation where sacks is a constituent with into as the head of a PP daughter.

Probability model

$$P(T,S) = \prod_{n \in T} p(r_n)$$

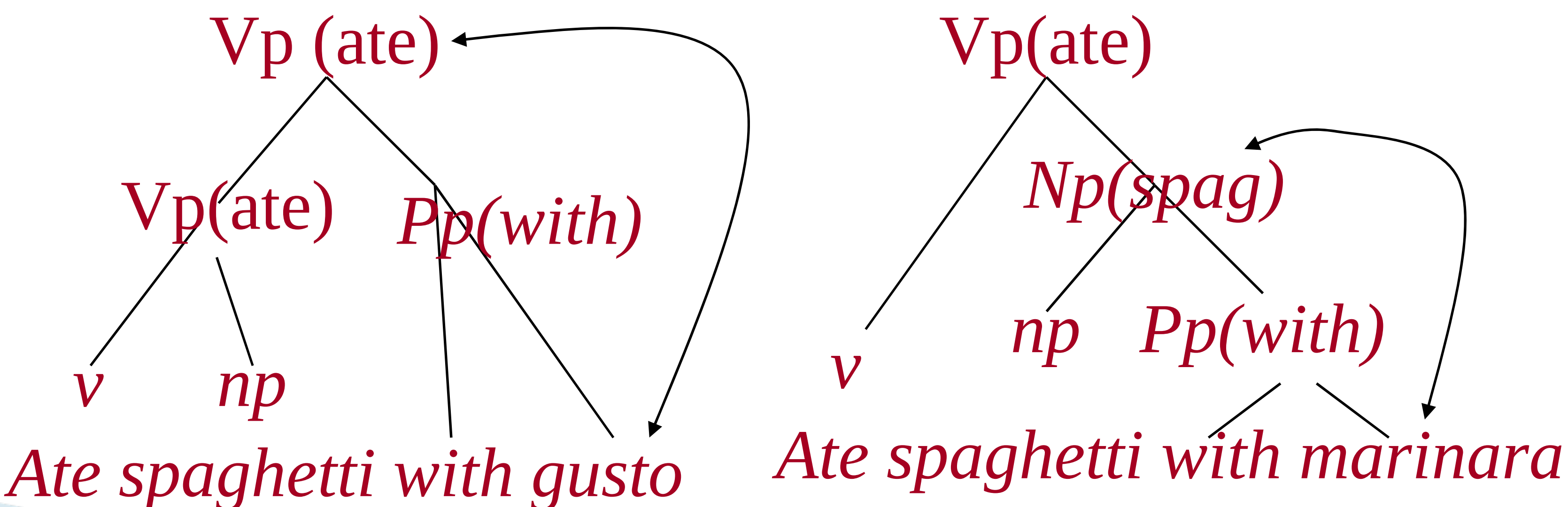
- P(T,S) = S-> NP VP (.5)*
- VP(dumped) -> V NP PP(into) (.7) (T1)
- NOM(sacks) -> NOM PP(into) (.01) (T2)

Preferences (2)

- Consider the VPs
  - Ate spaghetti with gusto
  - Ate spaghetti with marinara
- The affinity of gusto for eat is much larger than its affinity for spaghetti
- On the other hand, the affinity of marinara for spaghetti is much higher than its affinity for ate

Preferences (2)

- Note the relationship here is more distant and doesn't involve a headword since gusto and marinara aren't the heads of the PPs.



Summary

- Context-Free Grammars
- Parsing
  - Top Down, Bottom Up Metaphors
  - Dynamic Programming Parsers: CKY. Earley
- Disambiguation:
  - PCFG
  - Probabilistic Augmentations to Parsers
  - Tradeoffs: accuracy vs. data sparcity
  - Treebanks

Semantics: Representations and Analyses

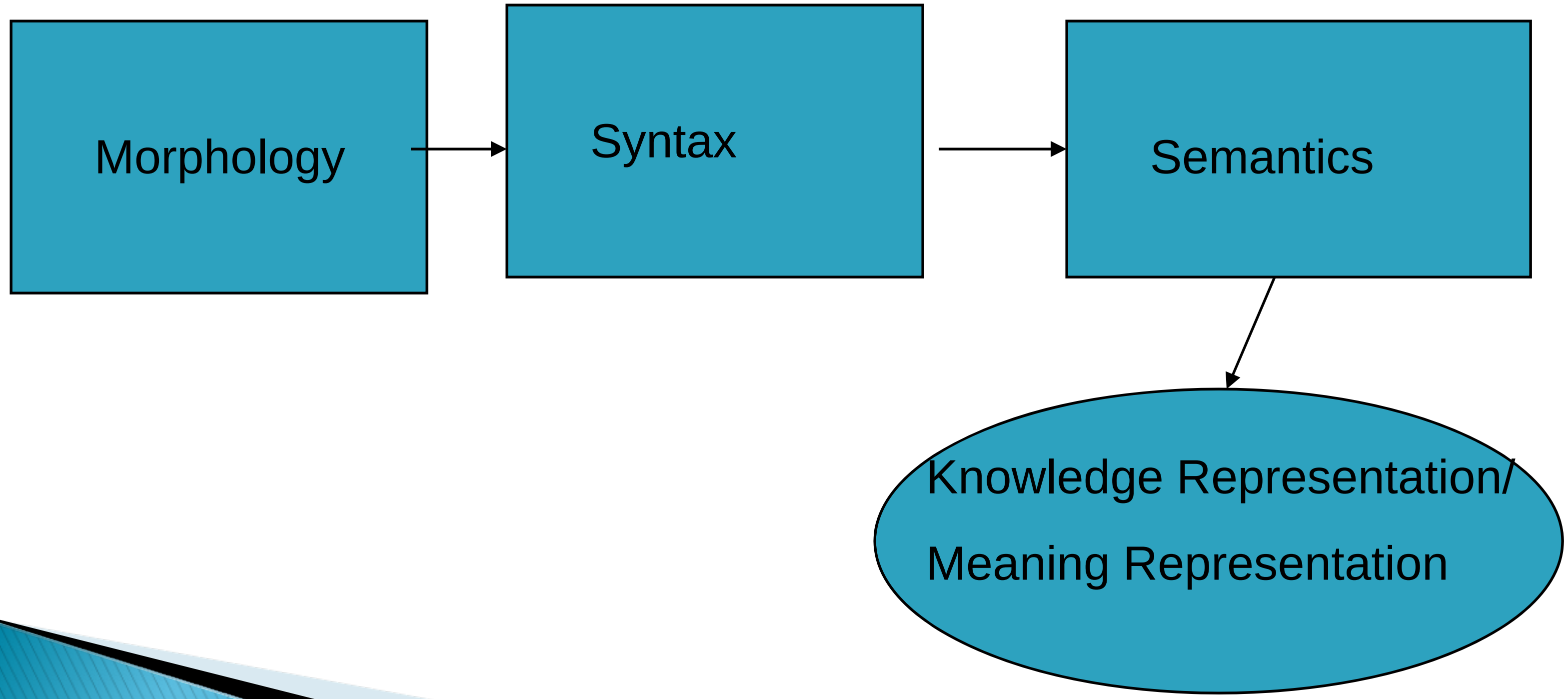
Slides adapted from Julia Hirschberg, Dan Jurafsky, Chris Manning

Question-Answering/Dialog

Where does the information come from?

- One possibility:
  - http://newyork.citysearch.com

NL Architecture



Semantic Considerations

- Meaning Representation
- Translation from syntax into the meaning representation
- Word meaning disambiguation
- Relations between words

Meaning Representation

- To represent questions
- To represent knowledge drawn from text

What Can Serve as a Meaning Representation?

- Anything that allows us to
  - Answer questions (What is the best French restaurant in the East Village?)
  - Determine truth (Is The Terrace in the Sky on 118th?)
  - Draw inferences (If The Terrace is in Butler Hall and Butler Hall is the tallest building on the West Side, then The Terrace is in the tallest building on the West Side.)

What kinds of meaning do we want to capture?

- Categories/entities
  - Tau, Jane, Asian cuisine, vegetarian
- Events
  - taking a taxi, nomination of Obama as Democratic candidate
- Time
  - Oct 30, next week, in 2 months
- Aspect
  - Kathy knows how to run. Kathy is running. Kathy ran to the restaurant in 5 min.
- Beliefs, Desires and Intentions (BDI)

Meaning Representations

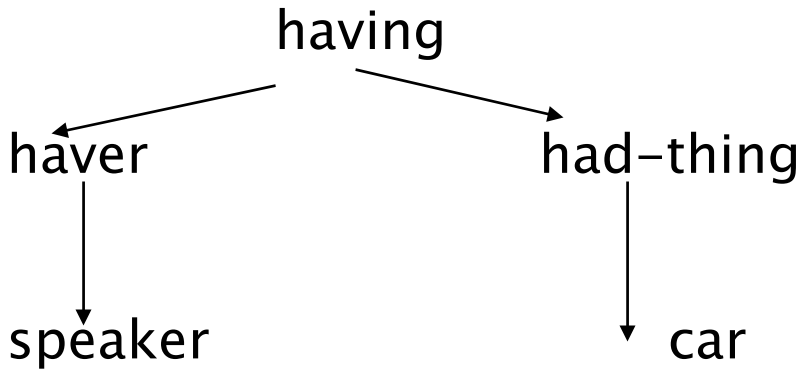
- All represent 'linguistic meaning' of

  I have a car

  and state of affairs in some world
- All consist of structures, composed of symbols representing objects and relations among them
  - FOPC:

$$\exists x,y\{Having(x)\wedge Haver(S,x)\wedge HadThing(y,x)\wedge Car(y)\}$$

- Semantic Net:



- Conceptual Dependency Diagram:
  Physical–object
  $\Uparrow$
  Car
  $\Uparrow$ Poss–By
  Speaker

- Frame
  Having
    Haver: S
    HadThing: Car

A Standard Representation: Predicate-Argument Structure

- Represents concepts and relationships among them
  - Nouns as concepts or arguments (red(ball))
  - Adjectives, adverbs, verbs as predicates (red(ball))
- Subcategorization (or, argument) frames specify number, position, and syntactic category of arguments
  - NP likes NP
  - NP likes Inf-VP
  - NP likes NP Inf-VP

Semantic (Thematic) Roles

- Subcat frames link arguments in surface structure with their semantic roles
  - Agent: George hit Bill. Bill was hit by George.
  - Patient: George hit Bill. Bill was hit by George.
- The claim of a theory of semantic roles is that these arguments of predicates can be usefully classified into a small set of semantically contentful classes

  And that these classes are useful for explaining lots of things

Common semantic roles

- Agent: initiator or doer in the event
- Patient: affected entity in the event; undergoes the action
  - Sue killed the rat.
- Theme: object in the event undergoing a change of state or location, or of which location is predicated
  - The ice melted
- Experiencer: feels or perceive the event
  - Bill likes pizza.
- Stimulus: the thing that is felt or perceived

Common semantic roles

- **Goal**:
  - Bill ran to Copley Square.
- **Recipient** (may or may not be distinguished from Goal):
  - Bill gave the book to Mary.
- **Benefactive** (may be grouped with Recipient):
  - Bill cooked dinner for Mary.
- **Source**:
  - Bill took a pencil from the pile.
- **Instrument**:
  - Bill ate the burrito with a plastic spork.
- **Location**:
  - Bill sits under the tree on Wednesdays

Common semantic roles

Try for yourself!

1. The submarine sank a troop ship.
2. Doris hid the money in the flowerpot.
3. Emma noticed the stain.
4. We crossed the street.
5. The boys climbed the wall.
6. The chef cooked a great meal.
7. The computer pinpointed the error.
8. A mad bull damaged the fence on Jack’s farm.
9. The company wrote me a letter.
10. Jack opened the lock with a paper clip.

Linking of thematic roles to syntactic positions

- John opened the door
- AGENT THEME
- The door was opened by John
- THEME AGENT
- The door opened
- THEME
- John opened the door with the key
- AGENT THEME INSTRUMENT

Deeper Semantics

- From the WSJ…
  - He melted her reserve with a husky-voiced paean to her eyes.
  - If we label the constituents He and her reserve as the Melter and Melted, then those labels lose any meaning they might have had.
  - If we make them Agent and Theme then we can do more inference.

Selectional Restrictions

- Selectional Restrictions: constraints on the *types* of arguments verbs take
  George assassinated the senator.
  **The spider assassinated the fly.*
  assassinate: intentional (political?) killing

- *The astronaut married the star.*

Problems

- What exactly is a role?
- What's the right set of roles?
- Are such roles universals?
- Are these roles atomic?
  - I.e. Agents
    - Animate, Volitional, Direct causers, etc
- Can we automatically label syntactic constituents with thematic roles?

First Order Predicate Calculus

- Not ideal as a meaning representation and doesn't do everything we want -- but better than many…
  - Supports the determination of truth
  - Supports compositionality of meaning
  - Supports question-answering (via variables)
  - Supports inference

NL Mapping to FOPC

- Terms: constants, functions, variables
  - Constants: objects in the world, e.g. Huey
  - Functions: concepts, e.g. sisterof(Huey)
  - Variables: x, e.g. sisterof(x)
- Predicates: symbols that refer to relations that hold among objects in some domain or properties that hold of some object in a domain

  likes(Kathy, pasta)

  female(Kathy) person(Kathy)

- Logical connectives permit compositionality of meaning

  pasta(x) → likes(Kathy,x) "Kathy likes pasta"

  cat(Vera) ^ odd(Vera) "Vera is an odd cat"

  sleeping(Huey) v eating(Huey) "Huey either is sleeping or eating"

- Sentences in FOPC can be assigned truth values
  - Atomic formulae are T or F based on their presence or absence in a DB (Closed World Assumption?)
  - Composed meanings are inferred from DB and meaning of logical connectives

- $cat(Huey)$
- $sibling(Huey,Vera)$
- $cat(Huey) \wedge sibling(Huey,Vera) \rightarrow cat(Vera)$

- Limitations:
  - Do 'and' and 'or' in natural language really mean '$\wedge$' and 'v'?

    Mary got married and had a baby. And then…

    Your money or your life!
  - Does '$\rightarrow$' mean 'if'?

    If you go, I'll meet you there.
  - How do we represent other connectives?

    She was happy but ignorant.

- Quantifiers: $\exists, \forall$
  - Existential quantification: There is a unicorn in my garden. Some unicorn is in my garden.
  - Universal quantification: The unicorn is a mythical beast. Unicorns are mythical beasts.
  - Many? A few? Several? A couple?

Temporal Representations

- How do we represent time and temporal relationships between events?

  It seems only yesterday that Martha Stewart was in prison but now she has a popular TV show. There is no justice.
- Where do we get temporal information?
  - Verb tense
  - Temporal expressions
  - Sequence of presentation
- Linear representations: Reichenbach '47

- Utterance time (U): when the utterance occurs
- Reference time (R): the temporal point-of-view of the utterance
- Event time (E): when events described in the utterance occur

George is eating a sandwich.

-- E,R,U →

George had eaten a sandwich (when he realized…)

E – R – U →

George will eat a sandwich.

--U,R – E →

While George was eating a sandwich, his mother arrived.

Verbs and Event Types: Aspect

- Statives: states or properties of objects at a particular point in time

  I am hungry.
- Activities: events with no clear endpoint

  I am eating.
- Accomplishments: events with durations and endpoints that result in some change of state

  I ate dinner.
- Achievements: events that change state but have no particular duration – they occur in an instant

  I got the bill

Beliefs, Desires and Intentions

- Very hard to represent internal speaker states like believing, knowing, wanting, assuming, imagining
  - Not well modeled by a simple DB lookup approach so..
  - Truth in the world vs. truth in some possible world

  George imagined that he could dance.

  George believed that he could dance.
- Augment FOPC with special modal operators that take logical formulae as arguments, e.g. believe, know

Believes(George, dance(George))

Knows(Bill,Believes(George,dance(George)))

- Mutual belief: I believe you believe I believe….
  - Practical importance: modeling belief in dialogue
  - Clark's grounding

Sum

- Many hard problems in full semantic representation:
  - Temporal relations: tense, aspect
  - BDI
- Current representations impoverished in many respects
- Read Ch 17.2-17.4, 18.1-18.7 (cover material through today)