

# The Role of Linguistics in Natural Language Processing

Martha Palmer  
CS598  
September 9, 2004

# Outline

---



- Introduction
- Syntactic Structure
- Syntactic ambiguities
- Semantic Structure
- Semantic ambiguities
- Conclusion

What do you call a successful movie? **Blockbuster**

- Tips on Being a **Successful Movie** Vampire ... I shall **call** the police.
- **Successful** Casting **Call** & Shoot for ``Clash of Empires" ... thank everyone for their participation in the making of yesterday's **movie**.
- Demme's casting is also highly entertaining, although I wouldn't go so far as to **call** it **successful**. This **movie's** resemblance to its predecessor is pretty vague...
- **VHS Movies: Successful** Cold **Call** Selling: Over 100 New Ideas, Scripts, and Examples from the Nation's Foremost Sales Trainer.

What do you call a successful movie?

- Tips on Being a **Successful Movie** Vampire ... I shall **call** the police.

- 

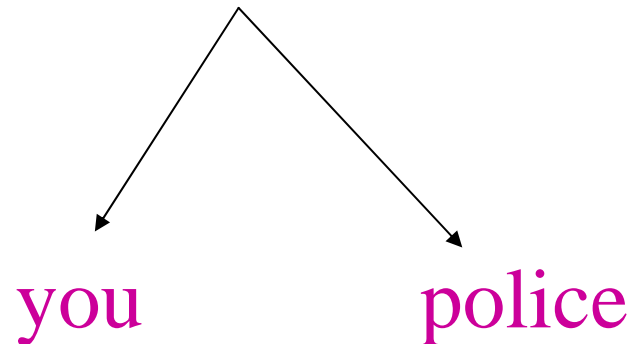
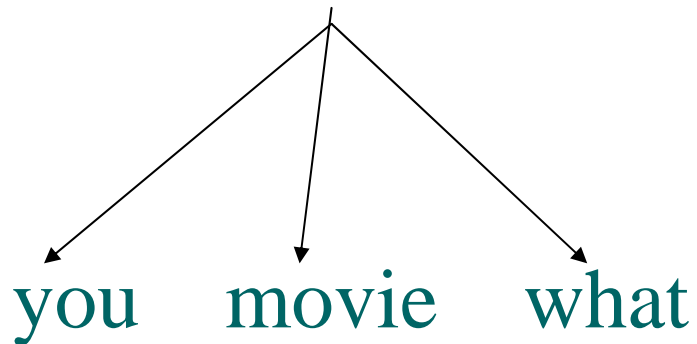
- Demme's casting is also highly entertaining, although I wouldn't go so far as to **call** it **successful**. This **movie's** resemblance to its predecessor is pretty vague...

- 

# Filtering out “call the police”

- Different senses,
- different syntax,
  - different participants

call(you,movie,what)  $\neq$  call(you,police)



# Syntactic Structure

---



- Syntactic categories and parsers
- Structural ambiguities in sentence interpretation
- Features
- Machine Translation

- Syntax
  - Grammars, parsers, parse trees, dependency structures
- Semantics
  - Subcategorization frames, semantic classes, ontologies, formal semantics
- Pragmatics
  - Pronouns, reference resolution, discourse models

# Syntactic Categories

---



- Nouns, pronouns, Proper nouns
- Verbs, intransitive verbs, transitive verbs, ditransitive verbs (subcategorization frames)
- Modifiers, Adjectives, Adverbs
- Prepositions
- Conjunctions



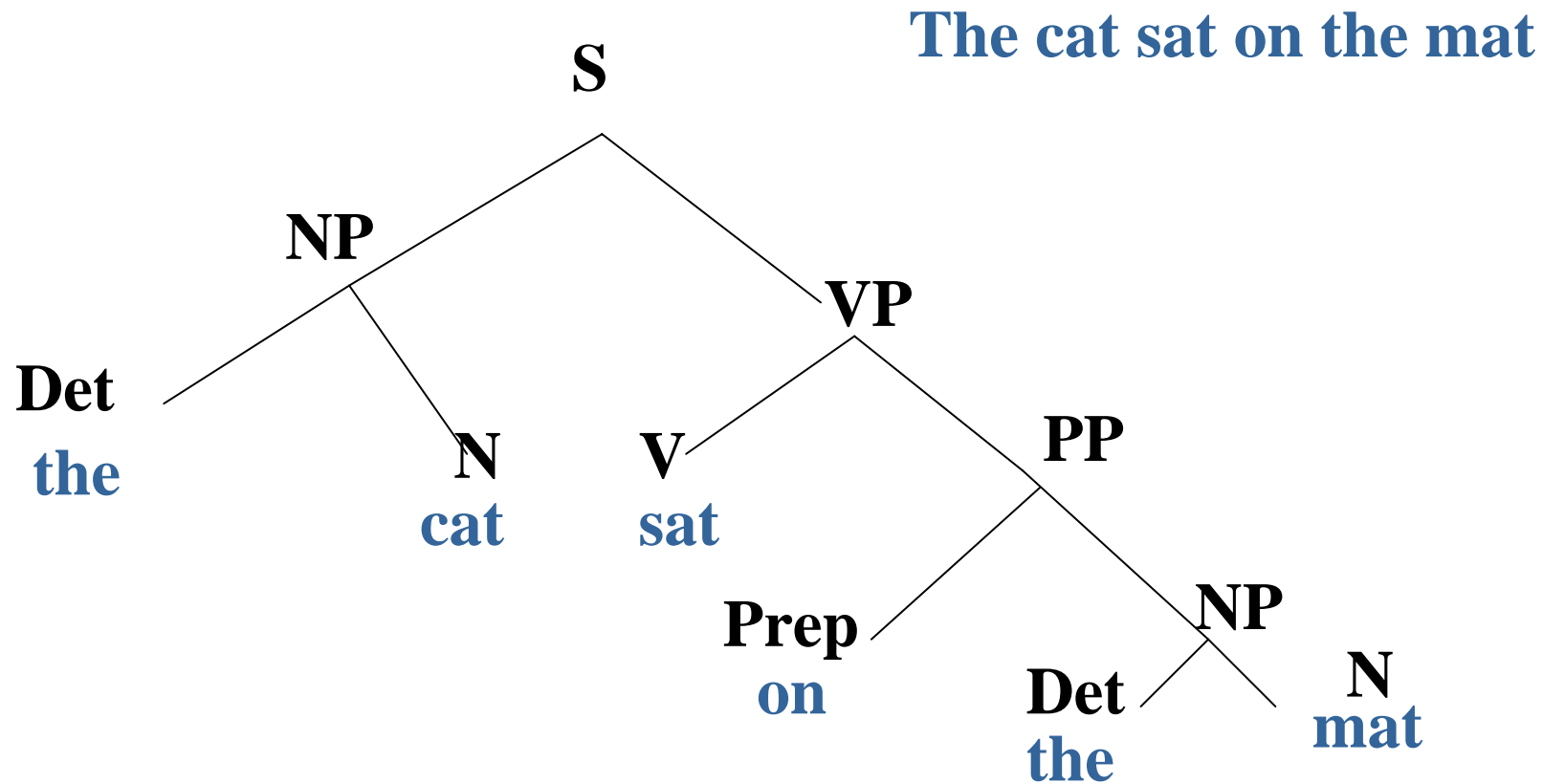
# Syntactic Parsing

---

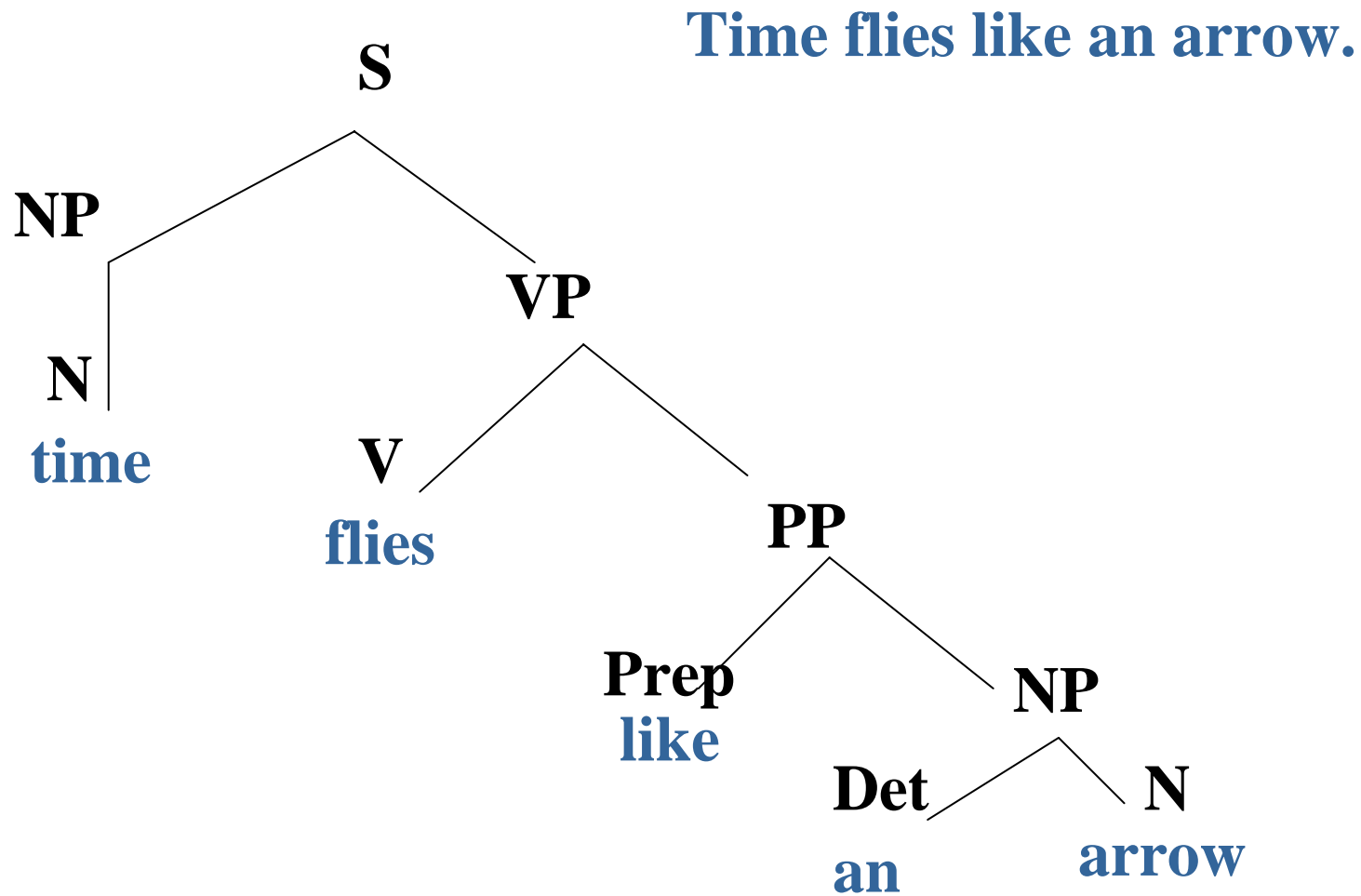


- The cat sat on the mat.  
Det Noun Verb Prep Det Noun
- Time flies like an arrow.  
Noun Verb Prep Det Noun
- Fruit flies like a banana.  
Noun Noun Verb Det Noun

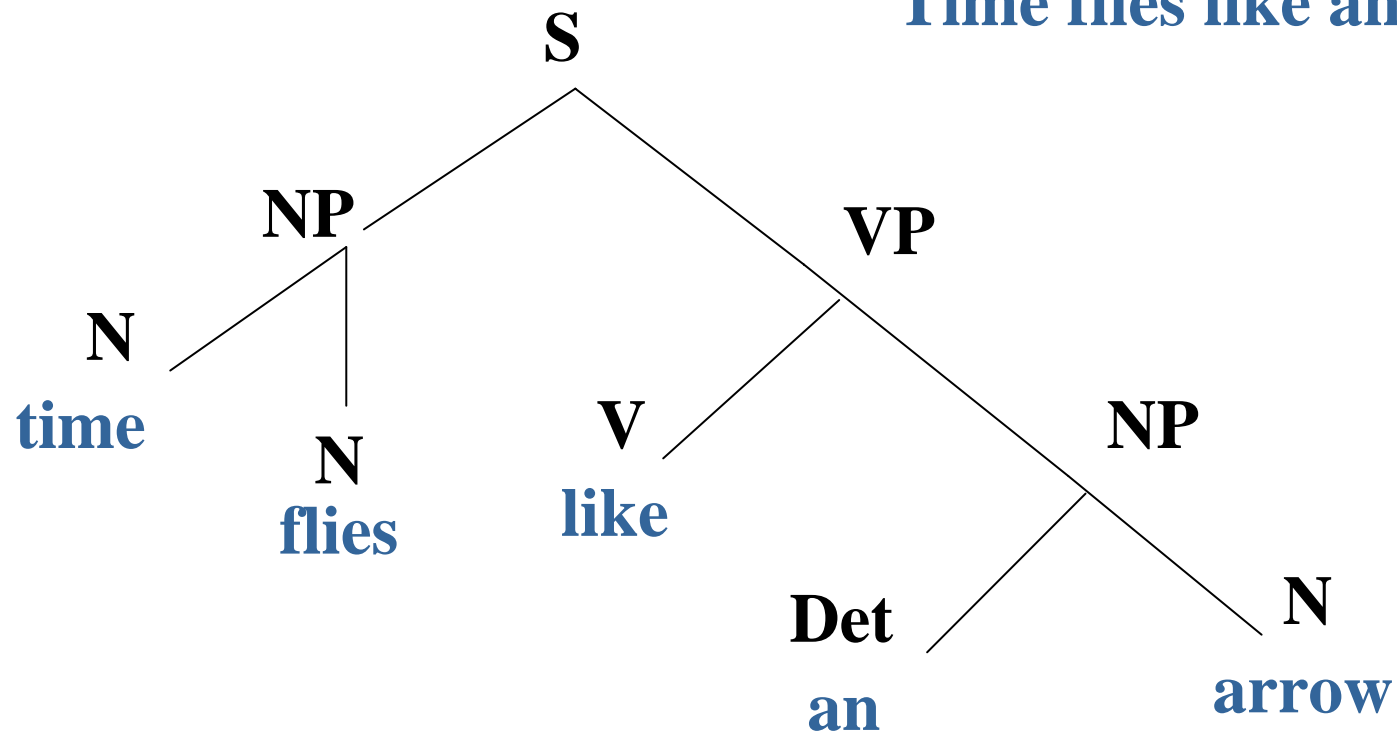
# Parses



# Parses



**Time flies like an arrow.**



# Lexicon

---



noun(cat).

noun(mat).

det(the).

det(a).

verb(sat).

prep(on).

noun(flies).

noun(time).

noun(arrow).

det(an).

verb(flies).

verb(time).

prep(like).

# Lexicon with Roots

---



noun(cat,cat).

noun(mat,mat).

det(the,the)

det(a,a).

verb(sat,sit).

prep(on,on).

noun(flies,fly).

noun(time,time).

noun(arrow,arrow).

det(an,an).

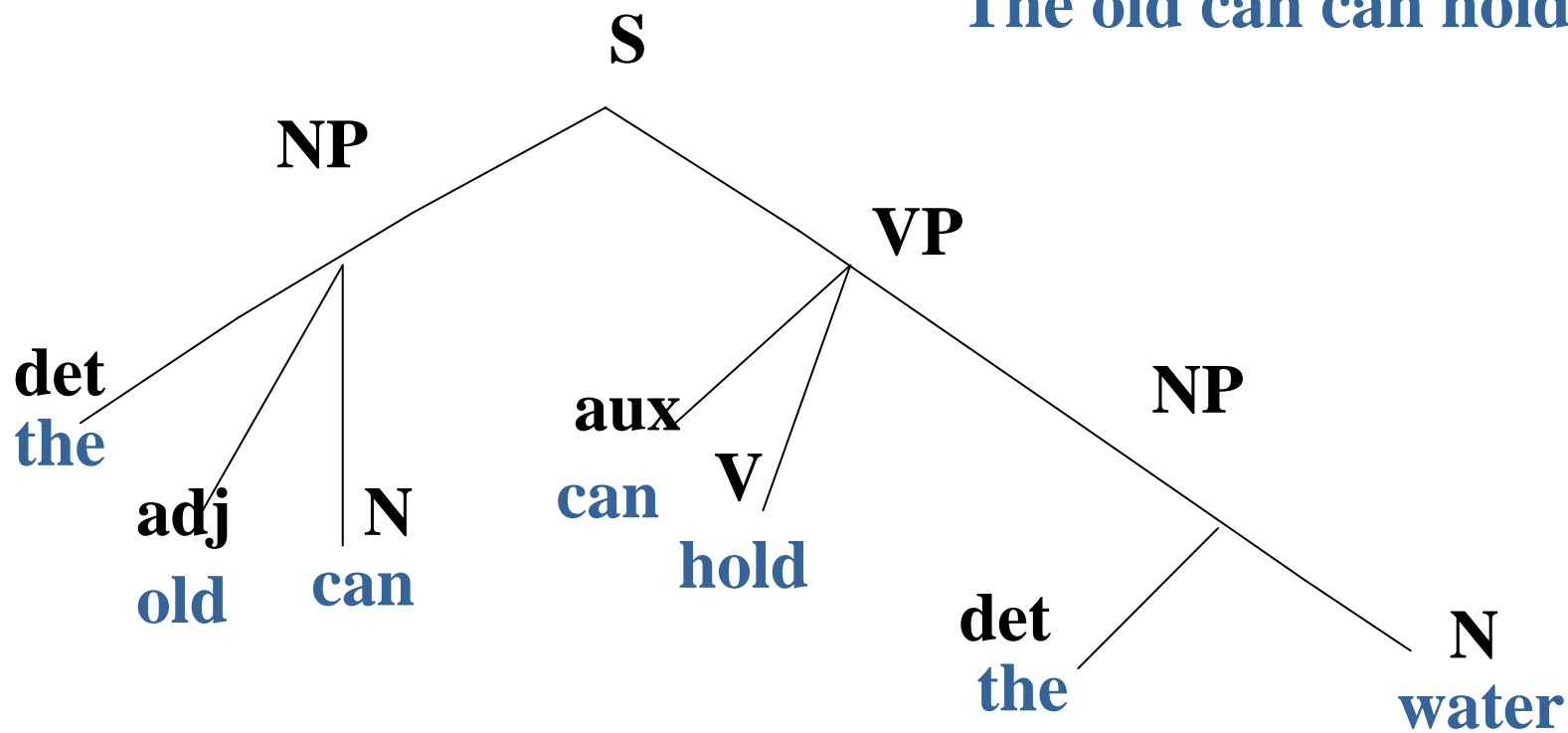
verb(flies,fly).

verb(time,time).

prep(like,like).

# Parses

The old can can hold the water.



# Lexicon

*The old can can hold the water.*

---



Noun(can,can)

Verb(hold,hold)

Noun(cans,can)

Verb(holds,hold)

Noun(water,water)

Aux(can,can)

Noun(hold,hold)

Adj(old,old)

Noun(holds,hold)

Det(the,the)



# Simple Context Free Grammar in BNF notation

---



S → NP VP  
NP → Pronoun | Noun | Det Adj Noun | NP PP  
PP → Prep NP  
V → Verb | Aux Verb  
VP → V | V NP | V NP NP | V NP PP | VP PP

# Top-down parse in progress

[The, old, can, can, hold, the, water]



S → NP VP

NP → NP?

NP → Pronoun?

Pronoun? fail

NP → Noun?

Noun? fail

NP → Det Adj Noun?

Det? **the**

ADJ? **old**

Noun? **Can**

**Succeed.**

**Succeed.**

VP?

# Top-down parse in progress

[can, hold, the, water]



VP → VP?

V → Verb?

Verb? **fail**

V → Aux Verb?

Aux? **can**

Verb? **hold**

**succeed**

**succeed**

**fail** [the, water]

# Top-down parse in progress

[can, hold, the, water]



VP → VP NP

V → Verb?

Verb? **fail**

V → Aux Verb?

Aux? **can**

Verb? **hold**

NP → Pronoun?

Pronoun? **fail**

NP → Noun?

Noun? **fail**

NP → Det Adj Noun?

Det? **the**

ADJ? **fail**

# Lexicon

---



Noun(can,can)

Verb(hold,hold)

Noun(cans,can)

Verb(holds,hold)

Noun(water,water)

Aux(can,can)

Noun(hold,hold)

Adj(old,old)

Noun(holds,hold)

Adj( , )

Det(the,the)

# Top-down parse in progress

[can, hold, the, water]



VP → V NP?

V → Verb?

Verb? **fail**

V → Aux Verb?

Aux? **can**

Verb? **hold**

NP → Pronoun?

Pronoun? **fail**

NP → Noun?

Noun? **fail**

NP → Det Adj Noun?

Det? **the**

ADJ?

Noun? **water**

**SUCCEED**  
**SUCCEED**

# Lexicon

---

Noun(can,can)

Verb(hold,hold)

Noun(cans,can)

Verb(holds,hold)

Noun(water,water)

Aux(can,can)

Noun(hold,hold)

Adj(old,old)

Noun(holds,hold)

Adj( , )

Det(the,the)

Noun(old,old)

# Syntactic Structure

---



- Syntactic categories and parsers
- Structural ambiguities in sentence interpretation
- Features
- Machine Translation



# Structural ambiguities

---



- *That factory can can tuna.*
- *That factory cans cans of tuna and salmon.*
- *Have the students in cse91 finish the exam in 212.*
- *Have the students in cse91 finished the exam in 212?*

# Top-down approach

---



- Start with goal of sentence
  - $S \rightarrow NP VP$
  - $S \rightarrow Wh\text{-word Aux NP VP}$
- Will try to find an NP 4 different ways before trying a parse where the verb comes first.
- What does this remind you of?
  - search
- What would be better?

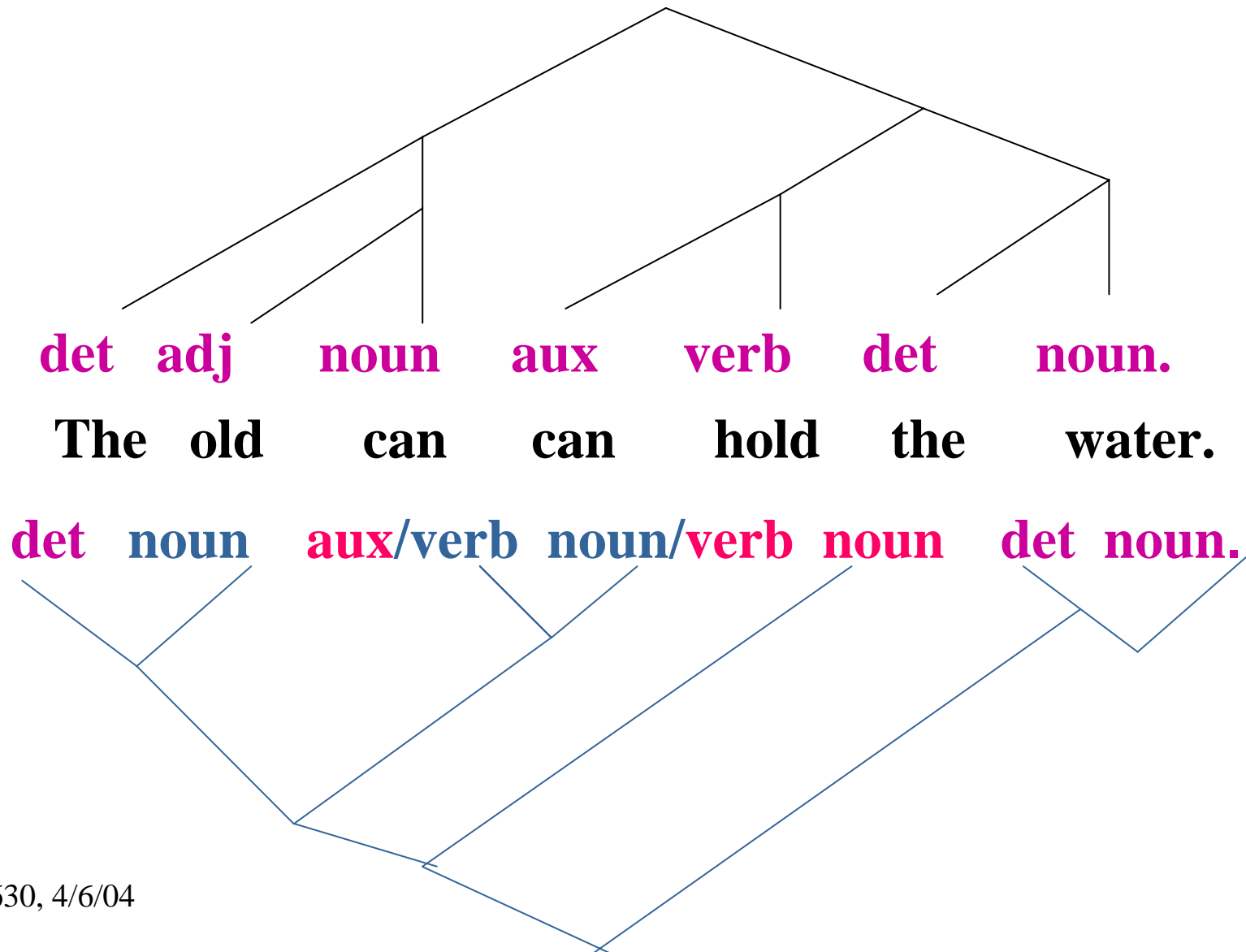
# Bottom-up approach

---

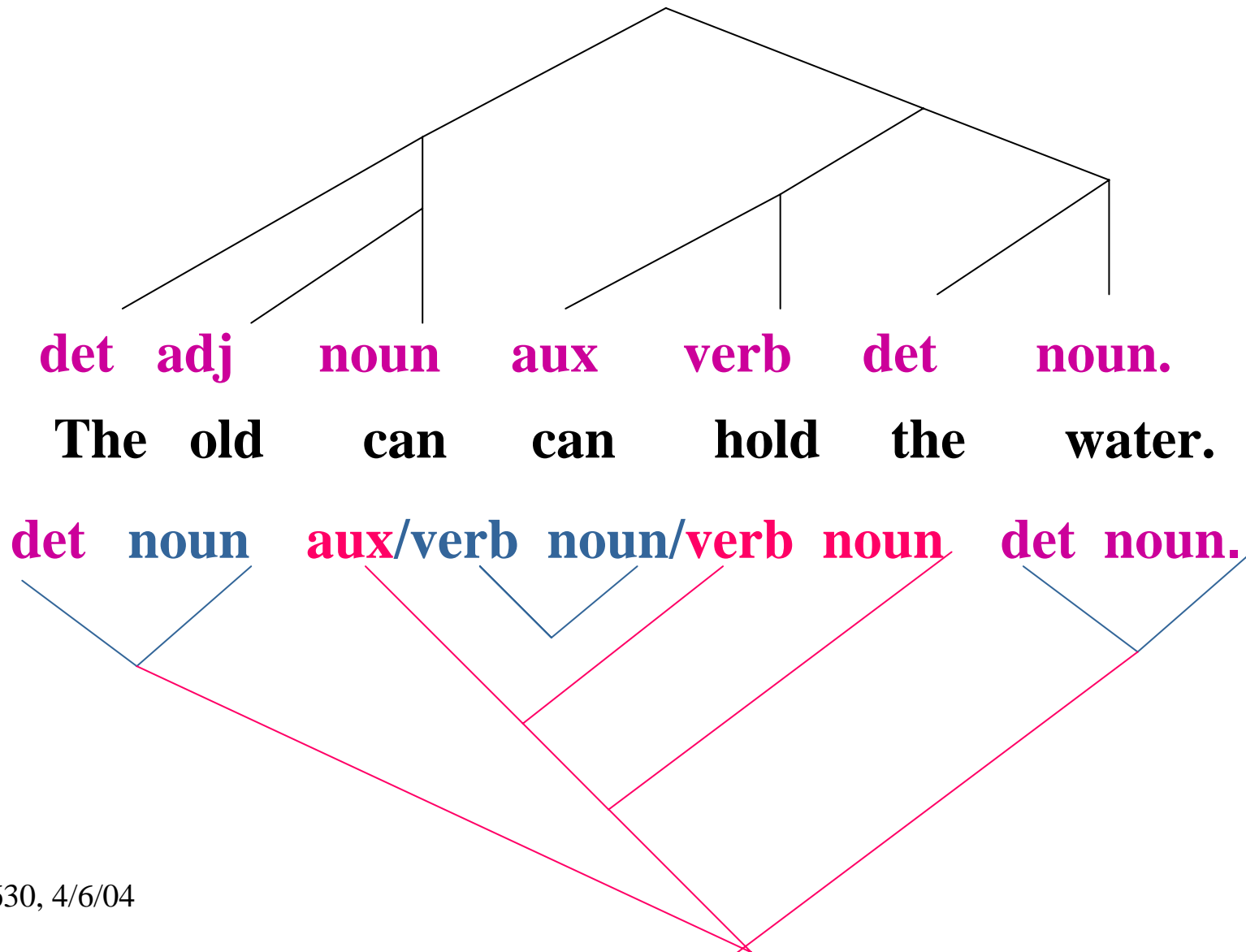


- Start with words in sentence.
- What structures do they correspond to?
- Once a structure is built, keep on a CHART.

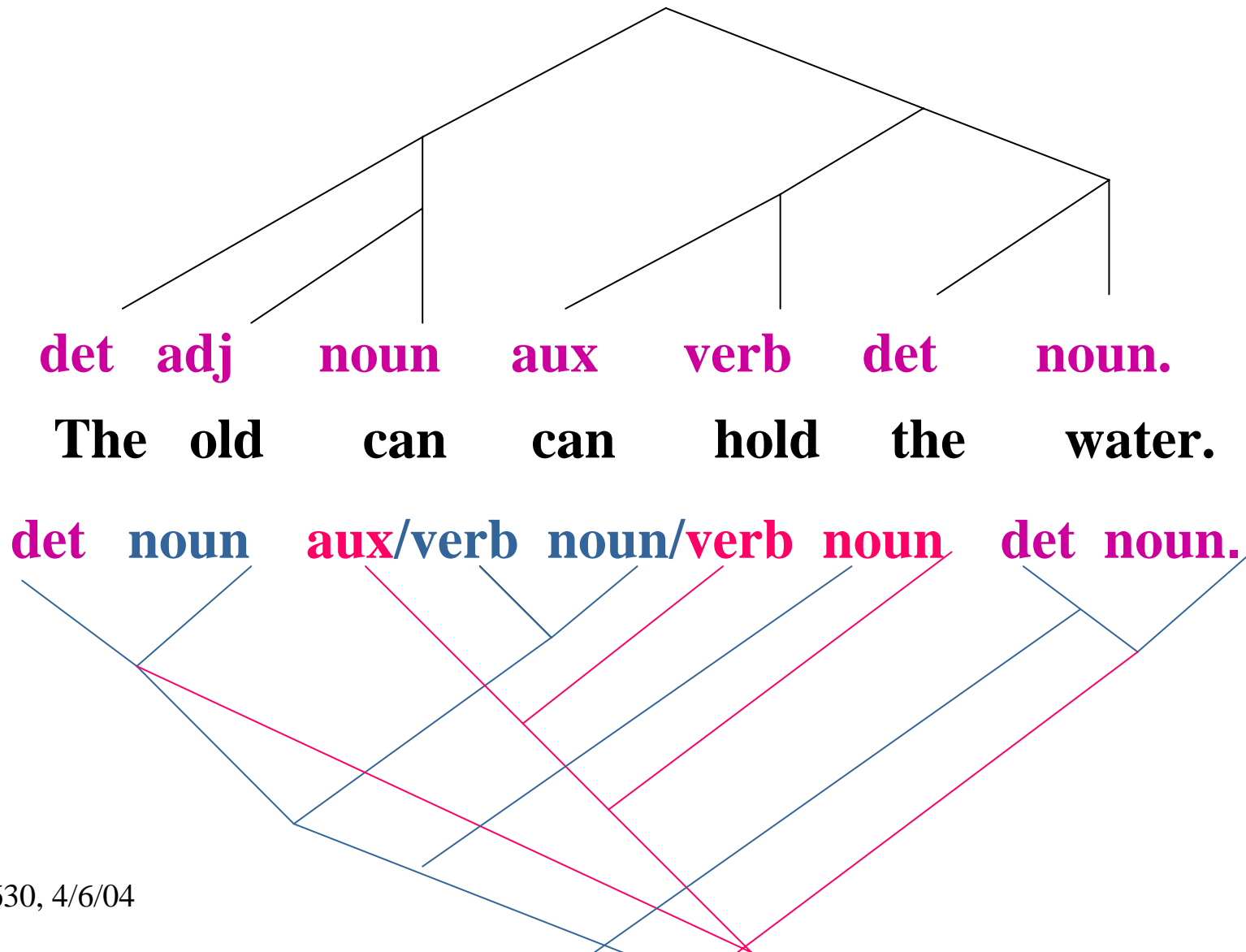
# Bottom-up parse in progress



# Bottom-up parse in progress



# Bottom-up parse in progress



# Top-down vs. Bottom-up

---



- Helps with POS ambiguities – only consider relevant POS
- Rebuilds the same structure repeatedly
- Spends a lot of time on impossible parses
- Has to consider every POS
- Builds each structure once
- Spends a lot of time on useless structures

**What would be better?**

# Hybrid approach

---



- Top-down with a chart
- Use look ahead and heuristics to pick most likely sentence type
- Use probabilities for pos tagging, pp attachments, etc.



# Headlines

---



- *Police Begin Campaign To Run Down Jaywalkers*
- *Iraqi Head Seeks Arms*
- *Teacher Strikes Idle Kids*
- *Miners Refuse To Work After Death*
- *Juvenile Court To Try Shooting Defendant*

# Syntactic Structure

---



- Syntactic categories and parsers
- Structural ambiguities in sentence interpretation
- **Features**
- Machine Translation

# Pronouns - Case

---



- She gave the book to her.
- She – subjective
- Her – objective

subjpronoun(she).

objpronoun(her).

# Features



- C for Case, Subjective/Objective
  - *She visited her.*
- P for Person agreement, (1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup>)
  - *I like him, You like him, He likes him,*
- N for Number agreement, Subject/Verb
  - *He likes him, They like him.*
- G for Gender agreement, Subject/Verb
  - English, reflexive pronouns *He washed himself.*
  - Romance languages, det/noun
- T for Tense,
  - auxiliaries, sentential complements, etc.
  - \* *will finished* is bad

# Example Lexicon Entries



## Using Features:

### Case, Number, Gender, Person

pronoun(subj, **sing, fem, third**, she, she).

pronoun(obj, **sing, fem, third**, her, her).

pronoun(obj, Num, Gender, **second**, you, you).

pronoun(subj, **sing**, Gender, **first**, I, I).

noun(Case, plural, Gender, third, flies, fly).

- One of the first applications for computers
  - bilingual dictionary > word-word translation
- Good translation requires *understanding!*
  - *War and Peace, The Sound and The Fury?*
- What can we do? Sublanguages.
  - technical domains, static vocabulary
  - Meteo in Canada, Caterpillar Tractor Manuals, Botanical descriptions, Military Messages

# Example translation

## Word Order and Scrambling

source	추가 공급물을 103 전위지원대대에게 사령부가 주었다. ({Chu-Ka} {Kong-Keup-Mul-eul} 103 {Cen-wi-Ci-wern-Tae-Tae-e-Ke} {Sa-Ryeong-Pu-Ka} {Cu- ren-Ta}.)
Glosser	additional supply <sub>2</sub> 103 FSB <sub>1</sub> headquarter <sub>0</sub> gave
OTS MT system	Additional supply <sub>2</sub> 103 FSB <sub>1</sub> headquarters <sub>0</sub> <u>which you bite</u> <span style="border: 1px solid black; padding: 2px;">gave</span> .
target	Headquarters <sub>0</sub> <span style="border: 1px solid black; padding: 2px;">gave</span> 103rd FSB <sub>1</sub> additional supplies <sub>2</sub> .
Penn/CGT	Headquarters <sub>0</sub> gave an additional supply <sub>2</sub> to a 103 forward support battalion <sub>1</sub> .

# Translation Issues: Korean to English

---



- Word order
- Dropped arguments
- Lexical ambiguities
- Structure vs morphology

KO: pswtay-ka

cehpang-ulo

ka-ss-la

EN: The unit

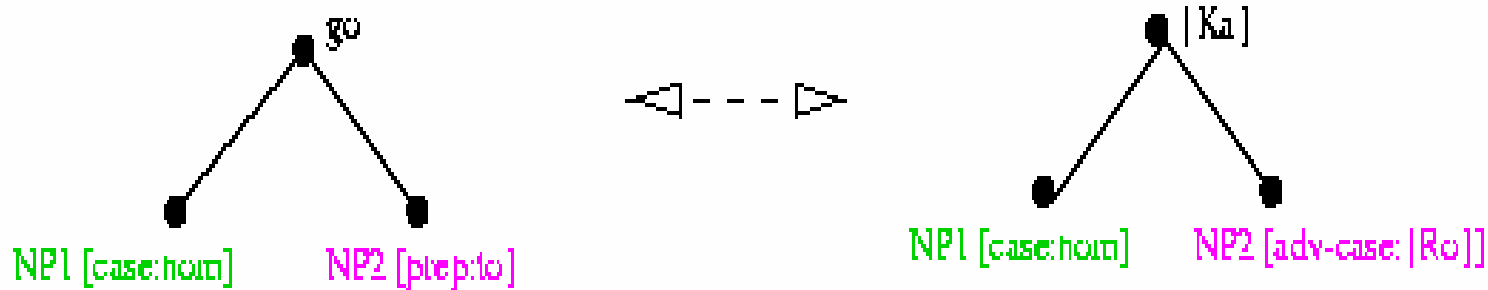
went

to the front line



- Predicate-argument structure
  - Basic *constituents* of the sentence and how they are *related* to each other
- Constituents
  - *John, Mary, the dog, pleasure, the store.*
- Relations
  - *Loves, feeds, go, to, bring*

# Abstracting away from surface structure



KO: pswtay-ka

cenpang-ulo

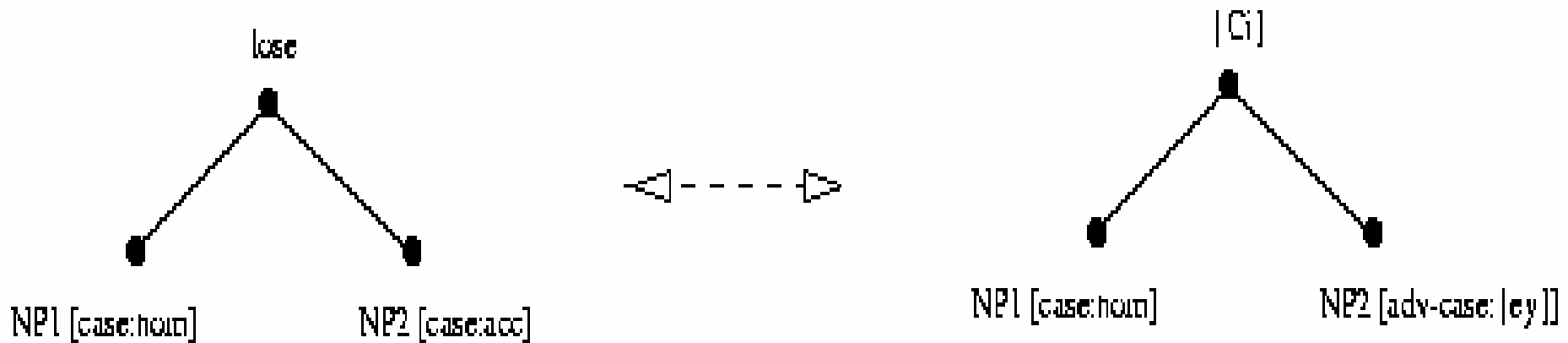
ka-ss-la

EN: The unit

went

to the front line

# Transfer lexicons – SMT?



# Machine Translation Lexical Choice- Word Sense Disambiguation

---



Iraq lost the battle.

*Ilakuka centwey ciessta.*

[Iraq] [battle] [lost].

John lost his computer.

*John-i computer-lul ilepelyessta.*

[John] [computer] [misplaced].

# Outline

---



- Introduction
- Syntactic Structure
- Syntactic ambiguities
- Semantic Structure
- Semantic ambiguities
- Conclusion

- That provides sets of possible syntactic frames for verbs.
- And provides clear, replicable sense distinctions.

AskJeeves: Who do you call for a good electronic lexical database for English?

# WordNet – Princeton

(Miller 1985, Fellbaum 1998)



## On-line lexical reference (dictionary)

- Nouns, verbs, adjectives, and adverbs grouped into synonym sets
- Other relations include hypernyms (ISA), antonyms, meronyms
- Typical top nodes - 5 out of 25
  - (*act, action, activity*)
  - (*animal, fauna*)
  - (*artifact*)
  - (*attribute, property*)
  - (*body, corpus*)

# WordNet – call, 28 senses



1. **name, call** -- (assign a specified, proper name to;  
*"They named their son David"; ...*)  
-> LABEL
2. **call, telephone, call up, phone, ring** -- (get or try to get into communication (with someone) by telephone;  
*"I tried to call you all night"; ...*)  
-> TELECOMMUNICATE
3. **call** -- (ascribe a quality to or give a name of a common noun that reflects a quality;  
*"He called me a bastard"; ...*)  
-> LABEL
4. **call, send for** -- (order, request, or command to come;  
*"She was called into the director's office"; "Call the police!"*)  
-> ORDER



- Limitations as a computational lexicon
  - Contains little syntactic information
    - Complex has syntax but no sense distinctions
  - No explicit lists of participants
  - Sense distinctions very fine-grained,
  - Definitions often vague
- Causes problems with creating training data for supervised Machine Learning – SENSEVAL2
  - Verbs > 16 senses (including *call*)
  - Inter-annotator Agreement ITA 73%,
  - Automatic Word Sense Disambiguation, WSD 60.2%

# WordNet: - call, 28 senses

---



WN2 , WN13,WN28

WN15 WN26

WN3 WN19

WN4 WN 7 WN8 WN9

WN1 WN22

WN20 WN25

WN18 WN27

WN5 WN 16

WN6 WN23

WN12

WN17 , WN 11

WN10, WN14, WN21, WN24

# WordNet: - call, 28 senses, Senseval2 groups (engineering!)



WN5, WN16, WN12

Loud cry

WN15 WN26

Bird or animal cry

WN3 WN19

WN1 WN22

Label

WN4 WN7 WN8 WN9

Request

WN20 WN25

Call a loan/bond

WN18 WN27

Challenge

WN2 WN13

Phone/radio  
WN28

WN6 WN23

Visit

WN17, WN11

WN10, WN14, WN21, WN24,

Bid

# Grouping improved scores:

ITA 82%, MaxEnt WSD 69%



- Call: 31% of errors due to confusion between senses within same group 1:
  - name, call -- (assign a specified, proper name to; *They named their son David*)
  - call -- (ascribe a quality to or give a name of a common noun that reflects a quality; *He called me a bastard*)
  - call -- (consider or regard as being; *I would not call her beautiful*)
  - 75% with training and testing on grouped senses vs.
  - 43% with training and testing on fine-grained senses

*Palmer, Dang, Fellbaum,, submitted, NLE*