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

\[ \text{call(you,movie,what)} \neq \text{call(you,police)} \]
Syntactic Structure

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

- **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
The cat sat on the mat

S
  /\  \
/    \ \
NP    VP
  |    /
  Det  V
   \  /  \
    Det N
     \ / \
      NP
       |  \
       Prep
        |  \
        Det
         |  \
         N
          |  \
          N

Det the

NP
  |  \
  N
   cat

VP
  |  \
  V
   sat

PP
  |  \
  Prep
   on

NP
  |  \
  Det
   the

N
  mat
Time flies like an arrow.

S

NP

NP

N

time

NP

Det

an

NP

N

arrow

V

flies

PP

Prep

like

VP


Time flies like an arrow.
| n   | m   |  |  |  |
|-----|-----|  |  |  |
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).
The old can can hold the water.
### Lexicon

*The old can can hold the water.*

<table>
<thead>
<tr>
<th>Noun(can, can)</th>
<th>Verb(hold, hold)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun(cans, can)</td>
<td>Verb(holds, hold)</td>
</tr>
<tr>
<td>Noun(water, water)</td>
<td>Aux(can, can)</td>
</tr>
<tr>
<td>Noun(hold, hold)</td>
<td>Adj(old, old)</td>
</tr>
<tr>
<td>Noun(holds, hold)</td>
<td></td>
</tr>
<tr>
<td>Det(the, the)</td>
<td></td>
</tr>
</tbody>
</table>
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
<table>
<thead>
<tr>
<th>Noun(can,can)</th>
<th>Verb(hold,hold)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun(cans,can)</td>
<td>Verb(holds,hold)</td>
</tr>
<tr>
<td>Noun(water,water)</td>
<td>Aux(can,can)</td>
</tr>
<tr>
<td>Noun(hold,hold)</td>
<td>Adj(old,old)</td>
</tr>
<tr>
<td>Noun(holds,hold)</td>
<td>Adj( ,  )</td>
</tr>
<tr>
<td>Det(the, the)</td>
<td></td>
</tr>
</tbody>
</table>
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
<table>
<thead>
<tr>
<th>Noun</th>
<th>Verb</th>
</tr>
</thead>
<tbody>
<tr>
<td>can, can</td>
<td>hold, hold</td>
</tr>
<tr>
<td>cans, can</td>
<td>holds, hold</td>
</tr>
<tr>
<td>water, water</td>
<td></td>
</tr>
<tr>
<td>hold, hold</td>
<td></td>
</tr>
<tr>
<td>holds, hold</td>
<td></td>
</tr>
<tr>
<td>the, the</td>
<td></td>
</tr>
<tr>
<td>old, old</td>
<td></td>
</tr>
<tr>
<td>old, old</td>
<td></td>
</tr>
</tbody>
</table>

**Lexicon**
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
  
  \[ S \rightarrow \text{NP VP} \]
  
  \[ S \rightarrow \text{Wh-word Aux NP VP} \]

- 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.
The old can can hold the water.

Bottom-up parse in progress
Bottom-up parse in progress

The old can can hold the water.
Bottom-up parse in progress

The old can can hold the water.
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, (1st, 2nd, 3rd)
  ➢ *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).
Machine Translation

• 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-Ken-Mul-cul} 103 {Conn-wi-Ci-wen-Tae-Tae-e-Ke} {Sa-Ryenng-Pu-Ka} {Cu-cernn-Ta}.). |
| Glosser             | additional supply₂ 103 FSB₁ headquarters₀ gave |
| OTS MT system       | Additional supply₂ 103 FSB₁ headquarters₀ which you bite gave. |
| Target              | Headquarters₀ gave 103rd FSB₁ additional supplies₂. |
| Penn/CGT            | Headquarters₀ gave an additional supply₂ to a 103 forward support battalion₁. |
Translation Issues:
Korean to English

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

KO: psuwlay-ka cenhang-ulol ka-sse-ta
EN: The unit went to the front line
Common Thread

- 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:  pswey-ka  cenpang-uio  ka-ss-ta
EN:  The unit went to the front line
Transfer lexicons – SMT?
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
Cornerstone: English lexical resource

- 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
    - Comlex 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%

Dang & Palmer, SIGLEX02
WordNet: - call, 28 senses

WN2, WN13, WN28
WN3, WN19
WN1, WN22
WN18, WN27
WN5, WN16
WN12
WN17, WN11

WN15, WN26
WN4, WN7, WN8, WN9
WN20, WN25
WN6, WN23
WN10, WN14, WN21, WN24
WordNet: - call, 28 senses, Senseval2 groups (engineering!)

- **Loud cry**
  - WN5, WN16, WN12
  - WN3, WN19
  - WN1, WN22

- **Label**
  - WN18, WN27
  - WN2, WN13
  - WN28
  - WN17, WN11

- **Bird or animal cry**
  - WN15, WN26
  - WN4, WN7, WN8, WN9
  - WN20, WN25

- **Request**
  - WN6, WN23

- **Call a loan/bond**
  - WN10, WN14, WN21, WN24

- **Visit**

- **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