DCS 530

SECTION ON NATURAL LANGUAGE UNDERSTANDING
JAMES ALLEN
FALL, 2017
THE HAPPY DOG RAN IN THE FIELD WITH ITS TONGUE HANGING OUT

LANGUAGE STRUCTURE AND FUNCTION
THE HAPPY DOG RAN IN THE FIELD WITH ITS TONGUE HANGING OUT

REFERENCE TO OBJECTS IS THE FUNCTION OF NOUN PHRASES

LANGUAGE STRUCTURE AND FUNCTION
THE HAPPY DOG RAN IN THE FIELD WITH ITS TONGUE HANGING OUT

REFERENCE TO OBJECTS IS THE FUNCTION OF NOUN PHRASES

LANGUAGE STRUCTURE AND FUNCTION
THE HAPPY DOG RAN IN THE FIELD WITH ITS TONGUE HANGING OUT

REFERENCE TO OBJECTS IS THE FUNCTION OF NOUN PHRASES

LANGUAGE STRUCTURE AND FUNCTION
THE HAPPY DOG RAN IN THE FIELD WITH ITS TONGUE HANGING OUT

REFERENCE TO OBJECTS IS THE FUNCTION OF NOUN PHRASES

PRONOUNS OFTEN REFER INDIRECTLY

LANGUAGE STRUCTURE AND FUNCTION

REFERENCE TO EVENTS IS THE FUNCTION OF VERB PHRASES

EVENTS DESCRIBE THE WORLD OVER TIME

LANGUAGE STRUCTURE AND FUNCTION
REFERENCE TO EVENTS IS THE FUNCTION OF VERB PHRASES

EVENTS DESCRIBE THE WORLD OVER TIME

LANGUAGE STRUCTURE AND FUNCTION

THE HAPPY DOG [ran] IN [the] FIELD WITH [[its] TONGUE] HANGING OUT

THE AGENT DOING THE RUNNING

THE OBJECT THAT IS HANGING OUT

EVENTS ARE STRUCTURED

EVENTS DESCRIBE THE WORLD OVER TIME

LANGUAGE STRUCTURE AND FUNCTION

THE RUNNING IS LOCATED WITHIN THE FIELD

THE RUNNING CO-OCCURS WITH THE TONGUE-HANGING-OUT EVENT

PREPOSITIONS (OR ADVERBS) RELATE EVENTS TO THEIR ARGUMENTS

EVENTS DESCRIBE THE WORLD OVER TIME

LANGUAGE STRUCTURE AND FUNCTION
THE HAPPY DOG RAN IN THE FIELD WITH ITS TONGUE HANGING OUT

THE RUNNING CO-OCCURS WITH THE TONGUE-HANGING-OUT EVENT

STRUCTURAL AMBIGUITY

TO DETERMINE INTENDED MEANING WE MUST DECIDE WHAT MODIFIES WHAT

LANGUAGE STRUCTURE AND FUNCTION

THE FIELD
CO-OCCURS WITH
TONGUE-HANGING-OUT
EVENT

THE RUNNING
CO-OCCURS WITH
THE TONGUE-
HANGING-OUT
EVENT

TO DETERMINE
INTENDED
MEANING
WE MUST DECIDE
WHAT MODIFIES
WHAT

STRUCTURAL AMBIGUITY

LANGUAGE STRUCTURE AND FUNCTION
Language structure and function

THE HAPPY DOG \text{ran} \text{in} \text{the field} \text{with} [[\text{its} \text{tongue}] \text{hanging out}}

\text{Compare: the dog ran in the field with the weeds growing tall}

\text{Structural ambiguity}

\text{To determine intended meaning we must decide what modifies what}

This sentence describes a proposition about the world.

Propositions are claims that can be true or false.
THE HAPPY DOG ran in the field with its tongue hanging out.

This sentence describes a proposition about the world.

Ran : agent [the happy dog]
: location [in [the field]]
: manner [with]
[ hanging-out : affected [[its] tongue]

Propositions are claims that can be true or false.

Language structure and function.
A SPEECH ACT INVOLVES A SPEAKER RELATING A PROPOSITION TO THE WORLD

SPEECH ACTS ARE ACTIONS AND MAY SUCCEED OR FAIL

LANGUAGE STRUCTURE AND FUNCTION
AN **INFORM** ACT CLAIMS A PROPOSITION IS TRUE: 
THE HAPPY DOG RAN IN THE FIELD WITH ITS TONGUE HANGING OUT

A SPEECH ACT INVOLVES A SPEAKER RELATING A PROPOSITION TO THE WORLD

SPEECH ACTS ARE ACTIONS AND MAY SUCCEED OR FAIL
AN **INFORM** ACT CLAIMS A PROPOSITION IS TRUE:
THE HAPPY DOG RAN IN THE FIELD WITH ITS TONGUE HANGING OUT

A **QUERY** ACT ASKS IF A PROPOSITION IS TRUE
DID THE HAPPY DOG RUN IN THE FIELD WITH ITS TONGUE HANGING OUT?

A **SPEECH ACT INVOLVES A SPEAKER RELATING A PROPOSITION TO THE WORLD**

**SPEECH ACTS ARE ACTIONS AND MAY SUCEED OR FAIL**
A **QUERY** ACT ASKS IF A PROPOSITION IS TRUE
DID THE HAPPY DOG RUN IN THE FIELD WITH ITS TONGUE HANGING OUT?

A **REQUEST/COMMAND** ACT TRIES TO MAKE A PROPOSITION TRUE
(TO FIDO) RUN IN THE FIELD WITH YOUR TONGUE HANGING OUT!

A **SPEECH ACT** INVOLVES A SPEAKER RELATING A PROPOSITION TO THE WORLD

SPEECH ACTS ARE ACTIONS AND MAY SUCCEED OR FAIL

LANGUAGE STRUCTURE AND FUNCTION
THE HAPPY DOG RAN IN THE FIELD WITH ITS TONGUE HANGING OUT

THE INDICATES THE OBJECT IS UNIQUE IN CONTEXT

DETERMINERS INDICATE QUANTITIES AND UNIQUENESS OF THE REFERRING EXPRESSION

DETAILS: DETERMINERS

LANGUAGE STRUCTURE AND FUNCTION
THE HAPPY DOG RAN IN THE FIELD WITH ITS TONGUE HANGING OUT

**THE** INDICATES THE OBJECT IS UNIQUE IN CONTEXT

POSSESSIVES INDICATES THE OBJECT IS UNIQUE WITH RESPECT TO ANOTHER NOUN PHRASE

**DETERMINERS** INDICATE QUANTITIES AND UNIQUENESS OF THE REFERRING EXPRESSION


DETAILS: DETERMINERS

LANGUAGE STRUCTURE AND FUNCTION
The happy dog ran in the field with its tongue hanging out.

Happy is an important property of the dog in this context.

Adjectives indicate properties of the referring expression.

Details: Adjectives

Language structure and function.
THE HAPPY DOG RAN IN THE FIELD WITH ITS TONGUE HANGING OUT

REPRESENTING STRUCTURE: CONTEXT FREE GRAMMAR

LANGUAGE STRUCTURE AND FUNCTION
STANFORD CORENLP TOOLS

THE HAPPY DOG RAN IN THE FIELD WITH ITS TONGUE HANGING OUT

REPRESENTING STRUCTURE: DEPENDENCY PARSING
(NLP.STANFORD.EDU:8080/CORENLP)

LANGUAGE STRUCTURE AND FUNCTION
STANFORD CORENLP TOOLS

THE HAPPY DOG RAN IN THE FIELD WITH ITS TONGUE HANGING OUT

Basic dependencies:

Collapsed dependencies:

BUT IT GETS THE COREFERENCE WRONG ....

Coreference:

REPRESENTING STRUCTURE: DEPENDENCY PARSES
(NLP.STANFORD.EDU:8080/CORENLP)

LANGUAGE STRUCTURE AND FUNCTION
THE HAPPY DOG RAN IN THE FIELD WITH ITS TONGUE HANGING OUT

COMPARE: THE DOG RAN IN THE FIELD WITH THE WEEDS GROWING TALL

IS IT

[RAN :AGENT [THE HAPPY DOG]
 :LOCATION [IN [THE FIELD]]
 :MANNER [WITH
  [HANGING-OUT
   :AFFECTED [[ITS] TONGUE]]]

OR

[RAN :AGENT [THE HAPPY DOG]
 :LOCATION [IN [THE FIELD]]
 :CONTAINS [WITH
  [HANGING-OUT
   :AFFECTED [[ITS] TONGUE]]]

DECISIONS AFFECTING AMBIGUITY

LANGUAGE STRUCTURE AND KNOWLEDGE
THE HAPPY DOG RAN IN THE FIELD WITH ITS TONGUE HANGING OUT

COMPARE: THE DOG RAN IN THE FIELD WITH THE WEEDS GROWING TALL

IS IT

[RAN :AGENT [THE HAPPY DOG]
:LOCATION [IN [THE FIELD]]
:MANNER [WITH]
[HANGING-OUT :AFFECTED [[ITS] TONGUE]]

OR

[RAN :AGENT [THE HAPPY DOG]
:LOCATION [IN [THE FIELD]]
:CONTAINS [WITH]
[HANGING-OUT :AFFECTED [[ITS] TONGUE]]

DECISIONS THAT AFFECT THIS

(1) STRUCTURE: DOES THE “WITH” ADVERBIAL MODIFY “RUN” OR “FIELD”
(2) REFERENCE: DOES “IT” REFER TO THE DOG OR THE FIELD?
(3) WORD SENSES: DOES “WITH” MEAN “MANNER” OR “CONTAINS”?

DECISIONS AFFECTING AMBIGUITY

LANGUAGE STRUCTURE AND KNOWLEDGE
THE HAPPY DOG RAN IN THE FIELD WITH ITS TONGUE HANGING OUT

COMPARE: THE DOG RAN IN THE FIELD WITH THE WEEDS GROWING TALL

DECISIONS AFFECTING INTERPRETATION
(1) DOES THE “WITH” ADVERBIAL MODIFY “RUN” OR “FIELD”
(2) DOES “IT” REFER TO THE DOG OR THE FIELD?
(3) DOES “FIELD” MEAN “A LOCATION” OR “AN ACADEMIC DISCIPLINE”?
(4) DOES “HANG OUT” MEAN “SUSPENDED” OR “GATHER SOCIALLY”?

WHAT KNOWLEDGE HELPS RESOLVE AMBIGUITY?

LANGUAGE STRUCTURE AND KNOWLEDGE
THE HAPPY DOG RAN IN THE FIELD WITH ITS TONGUE HANGING OUT

COMPARE: THE DOG RAN IN THE FIELD WITH THE WEEDS GROWING TALL

DECISIONS AFFECTING INTERPRETATION

(1) DOES THE “WITH” ADVERBIAL MODIFY “RUN” OR “FIELD”?
(2) DOES “IT” REFER TO THE DOG OR THE FIELD?
(3) DOES “FIELD” MEAN “A LOCATION” OR “AN ACADEMIC DISCIPLINE”?
(4) DOES “HANG OUT” MEAN “SUSPENDED” OR “GATHER SOCIA LLY”?

(1) DOGS HAVE TONGUES,
(2) FIELDS DON’T HAVE TONGUES
(3) TONGUES OFTEN HANG OUT OF DOG’S MOUTHS
(4) TONGUES CAN’T HANG OUT OF A FIELD
(5) TONGUES CAN’T HANG OUT SOCIA LLY! (ONLY PEOPLE CAN)
(6) RUNNING TYPICALLY HAPPENS IN LOCATIONS, AND NOT IN ACADEMIC DISCIPLINES (E.G., THE FIELD OF COMPUTER SCIENCE)

WHAT KNOWLEDGE HELPS RESOLVE AMBIGUITY?

LANGUAGE STRUCTURE AND KNOWLEDGE
I. The trophy would not fit in the brown suitcase because it was too big. What was too big?
   Answer 0: the trophy
   Answer 1: the suitcase

(1) IF SOMETHING FITS IN SOMETHING OF SIZE X, THEN IT WOULD FIT IN SOMETHING LARGER THAN X
(2) BEING TOO BIG IS A COMMON REASON WHY SOMETHING DOESN’T FIT
I. The trophy would not fit in the brown suitcase because it was too big. What was too big?
   Answer 0: the trophy
   Answer 1: the suitcase

   (1) IF SOMETHING FITS IN SOMETHING OF SIZE X, THEN IT WOULD FIT IN SOMETHING LARGER THAN X
   (2) BEING TOO BIG IS A COMMON REASON WHY SOMETHING DOESN’T FIT

II. The town councilors refused to give the demonstrators a permit because they feared violence. Who feared violence?
   Answer 0: the town councilors
   Answer 1: the angry demonstrators

   (1) TYPICALLY, A GOOD REASON TO REFUSE SOMETHING IS BECAUSE YOU FEAR SOME CONSEQUENCE
   (2)....
INTENTION EXAMPLES

in a supermarket...

customer. Black beans?
clerk: On aisle three

(1) CUSTOMERS ARE TYPICALLY TRYING FIND AND BUY PRODUCTS
(2) CLERK & CUSTOMER DON’T KNOW EACH OTHER

in a supermarket

customer. Black beans?
partner: No we had too many last week.

(1) WE HAD A LOT OF BLACK BEANS LAST WEEK
(2) WE HAVE NO BLACK BEANS IN THE CART YET

WHAT KNOWLEDGE HELPS RESOLVE AMBIGUITY?

LANGUAGE STRUCTURE AND KNOWLEDGE
BUT UNDERSTANDING REQUIRES CONTEXT!

At a grocery store ...
  Customer: *black beans?*
  clerk: *aisle 3.*

**BUT IN A HOME ENVIRONMENT...**

When arriving home ...
  Spouse: *black beans?*
  You: *Oh, sorry, I forget to get them.*

When exploring nutrition options ...
  Spouse: *black beans?*
  You: *227 calories in a cup*

When cooking ...
  Spouse: *black beans?*
  You: *in the cupboard.*

When cooking (adding black beans to a pot) ...
  Spouse: *black beans?*
  You: *don’t you like them.*
SYNTAX

THE STRUCTURE OF LANGUAGE
Figure 3.1 A tree representation of John ate the cat

1. $S \rightarrow NP \ VP$
2. $VP \rightarrow V \ NP$
3. $NP \rightarrow NAME$
4. $NP \rightarrow ART \ N$
5. $NAME \rightarrow John$
6. $V \rightarrow ate$
7. $ART \rightarrow the$
8. $N \rightarrow cat$
PARSING METHODS

1. $S \rightarrow NP\ VP$
2. $VP \rightarrow V\ NP$
3. $NP \rightarrow NAME$
4. $NP \rightarrow ART\ N$
5. $NAME \rightarrow John$
6. $V \rightarrow ate$
7. $ART \rightarrow the$
8. $N \rightarrow cat$

**TOP DOWN**

- $S \rightarrow NP\ VP$
- $S \rightarrow NAME\ VP$
- $S \rightarrow John\ VP$
- $S \rightarrow John\ V\ NP$
- $S \rightarrow John\ ate\ NP$
- $S \rightarrow John\ ate\ ART\ N$
- $S \rightarrow John\ ate\ the\ N$
- $S \rightarrow John\ ate\ the\ cat$

**BOTTOM UP**

- $S \rightarrow NP\ VP$
- $NAME\ VP \rightarrow NAME\ V\ NP$
- $NAME\ V\ NP \rightarrow NAME\ V\ ART\ N$
- $NAME\ V\ ART\ N \rightarrow NP\ V\ ART\ N$
- $NP\ V\ ART\ N \rightarrow NP\ V\ NP$
- $NP\ V\ NP \rightarrow NP\ VP$
- $NP\ VP \rightarrow S$

(rewriting $S$)
(rewriting NP)
(rewriting NAME)
(rewriting VP)
(rewriting V)
(rewriting NP)
(rewriting ART)
(rewriting N)

(rewriting John)
(rewriting ate)
(rewriting the)
(rewriting cat)
(rewriting NAME)
(rewriting ART N)
(rewriting V NP)
(rewriting NP VP)
TOP DOWN PARSE AS SEARCH

“THE OLD MAN CRIED”

<table>
<thead>
<tr>
<th>Step</th>
<th>Current state</th>
<th>Backup States</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>((S) 1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>((NP VP) 1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>((ART N VP) 1)</td>
<td></td>
<td>S rewritten to NP VP</td>
</tr>
<tr>
<td></td>
<td>((ART ADJ N VP) 1)</td>
<td></td>
<td>NP rewritten producing two new states</td>
</tr>
<tr>
<td>4.</td>
<td>((N VP) 2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>((VP) 3)</td>
<td></td>
<td>the backup state remains</td>
</tr>
<tr>
<td>6.</td>
<td>((V) 3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>(( ) 4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>((V NP) 3)</td>
<td></td>
<td>the first backup is chosen</td>
</tr>
<tr>
<td>9.</td>
<td>((NP) 4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>((ART N) 4)</td>
<td></td>
<td>looking for ART at 4 fails</td>
</tr>
<tr>
<td>11.</td>
<td>((ART ADJ N) 4)</td>
<td></td>
<td>fails again</td>
</tr>
<tr>
<td>12.</td>
<td>((ART ADJ N VP) 1)</td>
<td></td>
<td>now exploring backup state saved in step 3</td>
</tr>
<tr>
<td>13.</td>
<td>((ADJ N VP) 2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.</td>
<td>((N VP) 3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td>((VP) 4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.</td>
<td>((V) 4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.</td>
<td>(( ) 5)</td>
<td></td>
<td>success!</td>
</tr>
</tbody>
</table>

1. $S \rightarrow NP \ VP$
2. $NP \rightarrow ART \ N$
3. $NP \rightarrow ART \ ADJ \ N$
4. $VP \rightarrow V$
5. $VP \rightarrow V \ NP$

NEED TO GENERATE ALL POSSIBILITIES

NEED TO GENERATE ALL POSSIBILITIES

ALL TERMS ARE GONE BUT NOT AT END OF SENTENCE!

TAKING FIRST BACKUP STATE

SERIES OF FAILURES TO RESUME AT POSITION 4

STARTING AGAIN AT 1!
CHARTS: ELIMINATING REPEATING THE SAME WORK AGAIN AND AGAIN

LEXICON

- the: ART
- large: ADJ
- can: N, AUX, V
- hold: N, V
- water: N, V

GRAMMAR

1. S → NP VP
2. NP → ART ADJ N
3. NP → ART N
4. NP → ADJ N
5. VP → AUX VP
6. VP → V NP

“THE”

“LARGE”

NP → ART • ADJ N

NP → ART • N
CHARTS: ELIMINATING REPEATING THE SAME WORK AGAIN AND AGAIN

LEXICON

- the: ART
- large: ADJ
- can: N, AUX, V
- hold: N, V
- water: N, V

GRAMMAR

1. S → NP VP
2. NP → ART ADJ N
3. NP → ART N
4. NP → ADJ N
5. VP → AUX VP
6. VP → V NP

“THE”

“LARGE”

STARTING ARC

EXTENDING ARC
CHART EXAMPLE (2) 
ADDING NEXT WORD: “CAN”

NEW CONSTITUENTS
(from completing arcs)

NEW LEXICAL CONSTITUENTS

NEW ACTIVE ARCS
(extentions)
CHART EXAMPLE (2)
ADDING NEXT WORD: “CAN”

NEW CONSTITUENTS
(FROM COMPLETING ARCS)

NEW LEXICAL
CONSTITUENTS

NEW ACTIVE ARCS
(EXTENSIONS)

NEW ACTIVE ARCS
(FROM GRAMMAR)
“THE LARGE CAN CAN HOLD”
CHART EXAMPLE (3) “THE LARGE CAN CAN HOLD”

NEW LEXICAL CONSTITUENTS

NEW ARCS (FROM GRAMMAR)

1. $S \rightarrow NP \cdot VP$
2. $NP \rightarrow ART \cdot ADJ \cdot N$
3. $NP \rightarrow ART \cdot N$
4. $NP \rightarrow ADJ \cdot N$
5. $VP \rightarrow AUX \cdot VP$
6. $VP \rightarrow V \cdot NP$
CHART EXAMPLE (4)
“THE LARGE CAN CAN HOLD THE WATER”
CHART EXAMPLE (4)
“THE LARGE CAN CAN HOLD THE WATER”

NEW CONSTITUENTS
FOR NP “THE WATER”
CHART EXAMPLE (4)
“THE LARGE CAN CAN HOLD THE WATER”

NP2 (rule 4)

NP1 (rule 2)

VP1 (RULE 6 FROM V3 & NP3)

NP3 (rule 3)

ART1 | ADJ1 | AUX1 | AUX2 | N3 | ART2 | N4

1. the
2. large
3. can
4. can
5. hold
6. the
7. water
8. 

S → NP • VP

S → NP • VP

VP → AUX • VP

VP → AUX • VP

VP → V • NP
CHART EXAMPLE (4)
“THE LARGE CAN CAN HOLD THE WATER”
CHART EXAMPLE (4)
“THE LARGE CAN CAN HOLD THE WATER”

S1 (RULE 1 FROM NP1 & VP2)

NP2 (rule 4)

NP1 (rule 2)

NP3 (rule 3)

V1

V2

V3

V4

ART1

ADJ1

AUX1

AUX2

N3

ART2

N4

1  the  2  large  3  can  4  can  5  hold  6  the  7  water  8

S → NP • VP

S → NP • VP

VP → AUX • VP

VP → AUX • VP

VP → V • NP

ARC COMPLETES
ARC COMPLETES
ARC COMPLETES
NEW CONSTITUENTS
FOR NP “THE WATER”
THE COMPLETE CHART

<table>
<thead>
<tr>
<th>S1 (rule 1 with NP1 and VP2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S2 (rule 1 with NP2 and VP2)</td>
</tr>
<tr>
<td>VP3 (rule 5 with AUX1 and VP2)</td>
</tr>
<tr>
<td>NP2 (rule 4)</td>
</tr>
<tr>
<td>VP2 (rule 5)</td>
</tr>
<tr>
<td>NP1 (rule 2)</td>
</tr>
<tr>
<td>VP1 (rule 6)</td>
</tr>
<tr>
<td>N1</td>
</tr>
<tr>
<td>N2</td>
</tr>
<tr>
<td>V1</td>
</tr>
<tr>
<td>V2</td>
</tr>
<tr>
<td>V3</td>
</tr>
<tr>
<td>V4</td>
</tr>
<tr>
<td>N3</td>
</tr>
<tr>
<td>ART2 (rule 3)</td>
</tr>
<tr>
<td>ART1</td>
</tr>
<tr>
<td>ADJ1</td>
</tr>
<tr>
<td>AUX1</td>
</tr>
<tr>
<td>AUX2</td>
</tr>
<tr>
<td>N4</td>
</tr>
</tbody>
</table>

1. S → NP VP
2. NP → ART ADJ N
3. NP → ART N
4. NP → ADJ N
5. VP → AUX VP
6. VP → V NP

"THE LARGE CAN CAN HOLD THE WATER"
TOWARDS PRACTICAL PARSING

- DISAMBIGUATION
  - There may be 100s of "legal" syntactic parses of a sentence, which one is right?

- EXPRESSIVITY
  - On the face of it, natural language seems beyond the practical expressive power of context-free grammars
    - Agreement, "movement" (e.g., questions, relative clauses, ..), ...
DISAMBIGUATION: STATISTICAL PARSERS

LARGE CORPUS OF PARSED SENTENCES (E.G., PENN TREEBANK)

PROBABILITY ESTIMATION

PROBABILISTIC LEXICON

PROBABILISTIC GRAMMAR

<table>
<thead>
<tr>
<th>Rule</th>
<th>Count for LHS</th>
<th>Count for Rule</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. S → NP VP</td>
<td>300</td>
<td>300</td>
<td>1</td>
</tr>
<tr>
<td>2. VP → V</td>
<td>300</td>
<td>116</td>
<td>.386</td>
</tr>
<tr>
<td>3. VP → V NP</td>
<td>300</td>
<td>118</td>
<td>.393</td>
</tr>
<tr>
<td>4. VP → V NP PP</td>
<td>66</td>
<td>22</td>
<td>.22</td>
</tr>
<tr>
<td>5. NP → NP PP</td>
<td>1023</td>
<td>241</td>
<td>.24</td>
</tr>
<tr>
<td>6. NP → N N</td>
<td>1023</td>
<td>92</td>
<td>.09</td>
</tr>
<tr>
<td>7. NP → N</td>
<td>1023</td>
<td>141</td>
<td>.14</td>
</tr>
<tr>
<td>8. NP → ART N</td>
<td>1023</td>
<td>558</td>
<td>.55</td>
</tr>
<tr>
<td>9. PP → P NP</td>
<td>307</td>
<td>307</td>
<td>1</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\text{PROB}(\text{ART} \mid \text{the}) &= .99 \\
\text{PROB}(N \mid \text{flies}) &= .48 \\
\text{PROB}(V \mid \text{flies}) &= .52 \\
\text{PROB}(V \mid \text{like}) &= .49 \\
\text{PROB}(P \mid \text{like}) &= .34 \\
\text{PROB}(\text{ART} \mid a) &= .995 \\
\text{PROB}(N \mid a) &= .005 \\
\text{PROB}(V \mid \text{flower}) &= .78 \\
\text{PROB}(V \mid \text{flower}) &= .22
\end{align*}
\]
**DISAMBIGUATION: STATISTICAL PARSERS**

**PROBABILISTIC LEXICON**

<table>
<thead>
<tr>
<th>Rule</th>
<th>Count for LHS</th>
<th>Count for Rule</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>S → NP VP</td>
<td>300</td>
<td>300</td>
<td>1</td>
</tr>
<tr>
<td>VP → V</td>
<td>300</td>
<td>116</td>
<td>.386</td>
</tr>
<tr>
<td>VP → V NP</td>
<td>300</td>
<td>118</td>
<td>.383</td>
</tr>
<tr>
<td>VP → V NP PP</td>
<td>300</td>
<td>66</td>
<td>.22</td>
</tr>
<tr>
<td>NP → NP PP</td>
<td>1023</td>
<td>241</td>
<td>.24</td>
</tr>
<tr>
<td>NP → N N</td>
<td>1023</td>
<td>92</td>
<td>.09</td>
</tr>
<tr>
<td>NP → N</td>
<td>1023</td>
<td>141</td>
<td>.14</td>
</tr>
<tr>
<td>NP → ART N</td>
<td>1023</td>
<td>558</td>
<td>.55</td>
</tr>
<tr>
<td>PP → P NP</td>
<td>307</td>
<td>307</td>
<td>1</td>
</tr>
</tbody>
</table>

**PROBABILISTIC GRAMMAR**

\[
\text{PROB(CONSTITUENT)} = \text{PROB(RULE)} \times \text{PROB(SUBCONSTIT_1)} \times \ldots \times \text{PROB(SUBCONSTIT_N)}
\]

**PROBABILISTIC CHART**

**NOTE:** SORRY, THE PROBABILITIES IN THE CHART COME FROM A DIFFERENT MODEL SO ARE NOT COMPUTABLE FROM THIS GRAMMAR & LEXICON!
STATE OF THE ART IN STATISTICAL PARSING

- A pure probabilistic context free grammar (PCFG) does not perform well.
- By adding more context in the rule probabilities (e.g., NP rules as subject of an S, ....) we can produce high performance systems.
- Accuracy around 95% of constituents.

Sounds good, but note that for a 10 word sentence that is less than a 50% chance of a totally correct parse!

Check out Stanford parser online: NLP.STANFORD.EDU:8080/PARSER/
FOCUS OF THE COURSE

- Most applications involving language in data science involve statistical models.
- "Shallow" processing, little semantics or contextual interpretation.
- We will review the basic statistical models that are used in current applications.
- Information retrieval, machine translation, sentiment analysis.
COURSEWORK

☐ MOST LECTURES WILL START WITH A 15 MINUTE QUIZ BASED

☐ THERE WILL BE A QUIZ THIS THURSDAY ON THE READINGS:
  ☐ CHAPTER 2 & 3 FROM ALLEN, “NATURAL LANGUAGE UNDERSTANDING”