

# 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

PRONOUNS  
OFTEN  
REFER  
INDIRECTLY



**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  
EVENTS  
IS THE FUNCTION  
OF VERB PHRASE**



**EVENTS DESCRIBE  
THE WORLD  
OVER TIME**

---

LANGUAGE STRUCTURE AND FUNCTION

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

**REFERENCE TO  
EVENTS  
IS THE FUNCTION  
OF VERB PHRASE**



**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 HAPPY DOG] **RAN IN** [THE FIELD] **WITH** [[ITS] TONGUE] **HANGING OUT**

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 HAPPY DOG] **RAN** IN [THE FIELD] **WITH** [[ITS] TONGUE] **HANGING OUT**

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

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 HAPPY DOG] **RAN** IN [THE FIELD] **WITH** [[ITS] TONGUE] **HANGING OUT**

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

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

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

**STRUCTURAL  
AMBIGUITY**



TO DETERMINE  
INTENDED  
MEANING  
WE MUST DECIDE  
WHAT MODIFIES  
WHAT

---

LANGUAGE STRUCTURE AND FUNCTION

[THE HAPPY DOG] **RAN IN** [THE FIELD] **WITH** [[ITS] TONGUE] **HANGING OUT**

**THIS SENTENCE  
DESCRIBES  
A PROPOSITION  
ABOUT THE  
WORLD**



**PROPOSITIONS ARE  
CLAIMS THAT  
CAN BE  
TRUE OR FALSE**

---

LANGUAGE STRUCTURE AND FUNCTION

[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

---

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 **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 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 **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



**“THE”, “A”, “SOME”,  
“MANY”, “A FEW”,  
“BOTH”, “NO”,  
“SEVERAL”, “TWO”,  
“A HUNDRED”**

**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

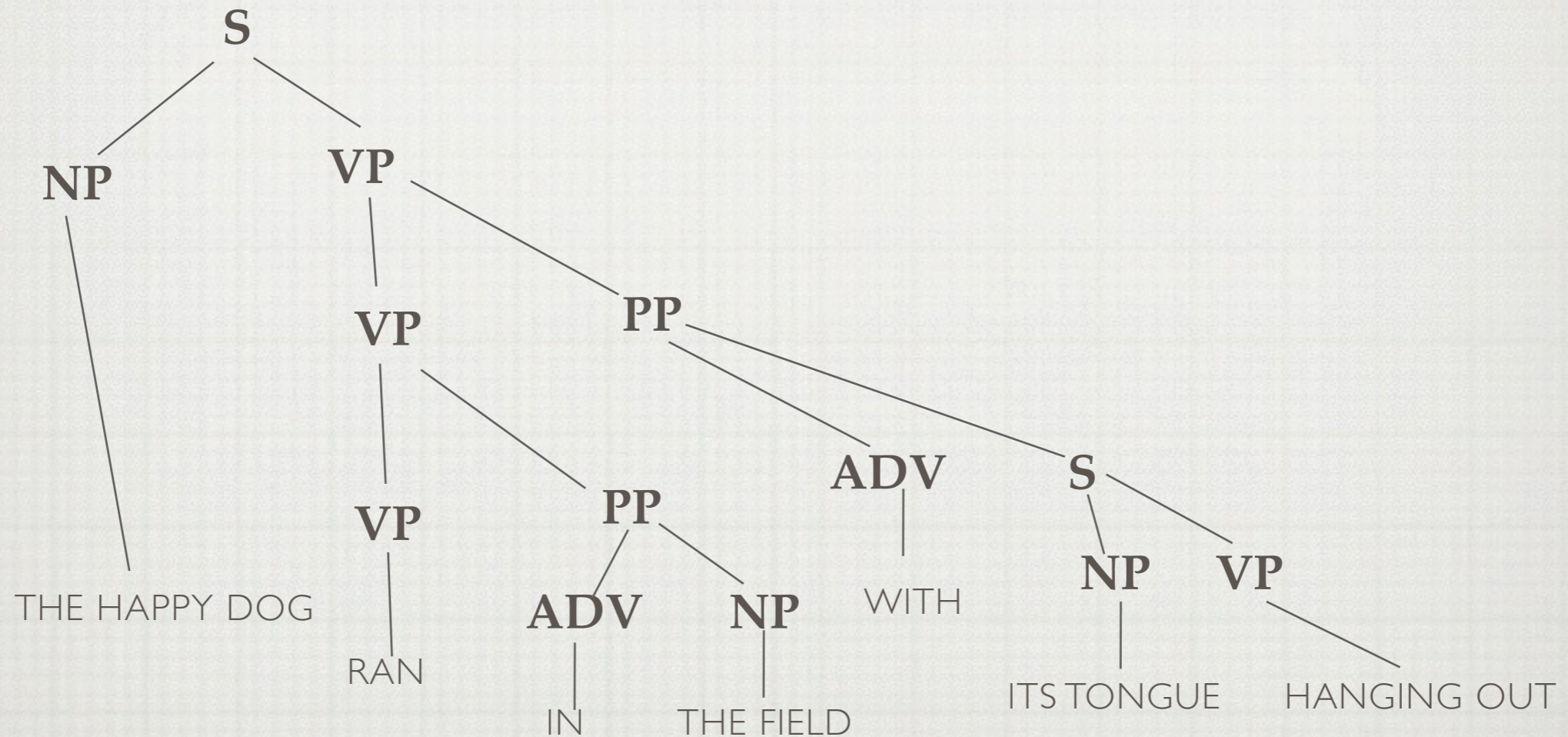


**DETAILS: ADJECTIVES**

---

LANGUAGE STRUCTURE AND FUNCTION

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



**REPRESENTING STRUCTURE: CONTEXT FREE GRAMMAR**

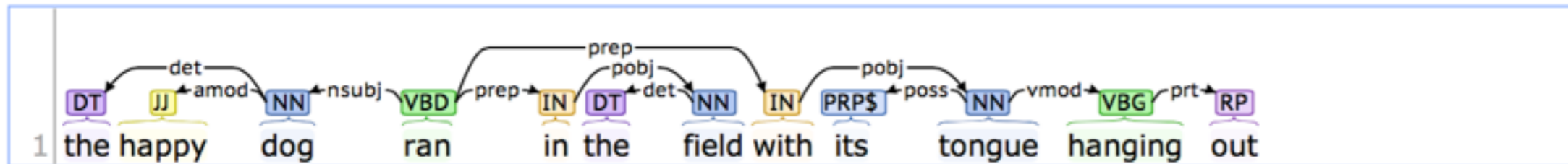
LANGUAGE STRUCTURE AND FUNCTION

# STANFORD CORENLP TOOLS

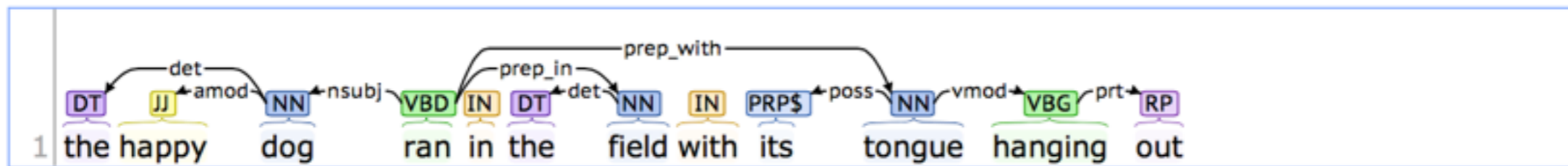
THE HAPPY DOG RAN IN THE FIELD  
WITH ITS TONGUE HANGING OUT



## Basic dependencies:



## Collapsed dependencies:



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

LANGUAGE STRUCTURE AND FUNCTION

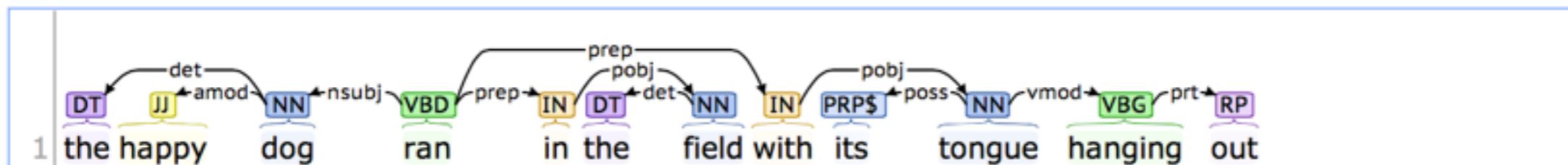


# STANFORD CORENLP TOOLS

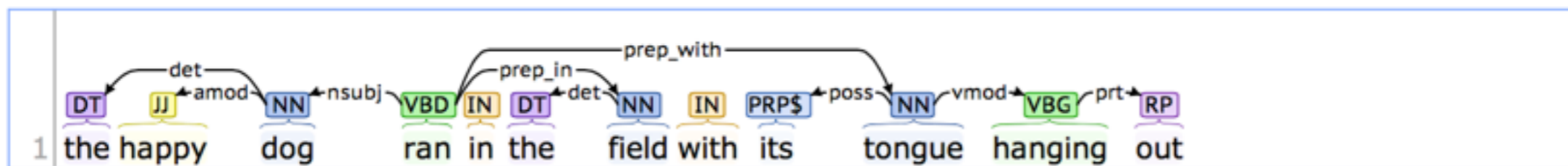
THE HAPPY DOG RAN IN THE FIELD  
WITH ITS TONGUE HANGING OUT



## Basic dependencies:

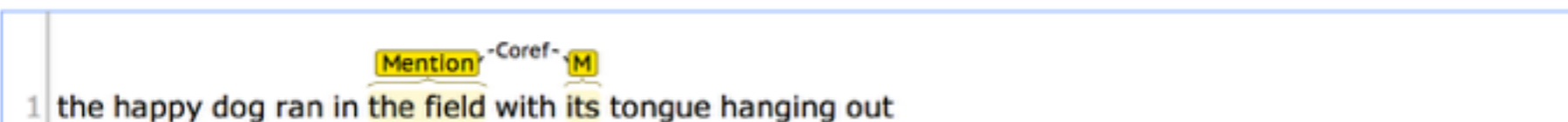


## Collapsed dependencies:



BUT IT GETS THE COREFERENCE WRONG ....

## Coreference:



**REPRESENTING STRUCTURE: DEPENDENCY PARSSES  
(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 SOCIALLY”?
  - (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 SOCIALLY! (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

# THE WINOGRAD CHALLENGE

([COMMONSENSE REASONING.ORG/WINOGRAD.HTML](http://COMMONSENSE REASONING.ORG/WINOGRAD.HTML))

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

**WHAT KNOWLEDGE HELPS RESOLVE AMBIGUITY?**

---

LANGUAGE STRUCTURE AND KNOWLEDGE

# THE WINOGRAD CHALLENGE

(COMMONSENSE REASONING.ORG/WINOGRAD.HTML)

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).....

**WHAT KNOWLEDGE HELPS RESOLVE AMBIGUITY?**

---

LANGUAGE STRUCTURE AND KNOWLEDGE

# 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.*

**DEEP UNDERSTANDING  
REQUIRES INTENTION  
RECOGNITION IN CONTEXT**

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

When arriving home ...

Spouse: *black beans?*

You: *Oh, sorry, I forget to get them.*

When cooking ...

Spouse: *black beans?*

You: *in the cupboard.*

When exploring nutrition options ...

Spouse: *black beans?*

You: *227 calories in a cup*

When cooking (adding black beans to a pot) ...

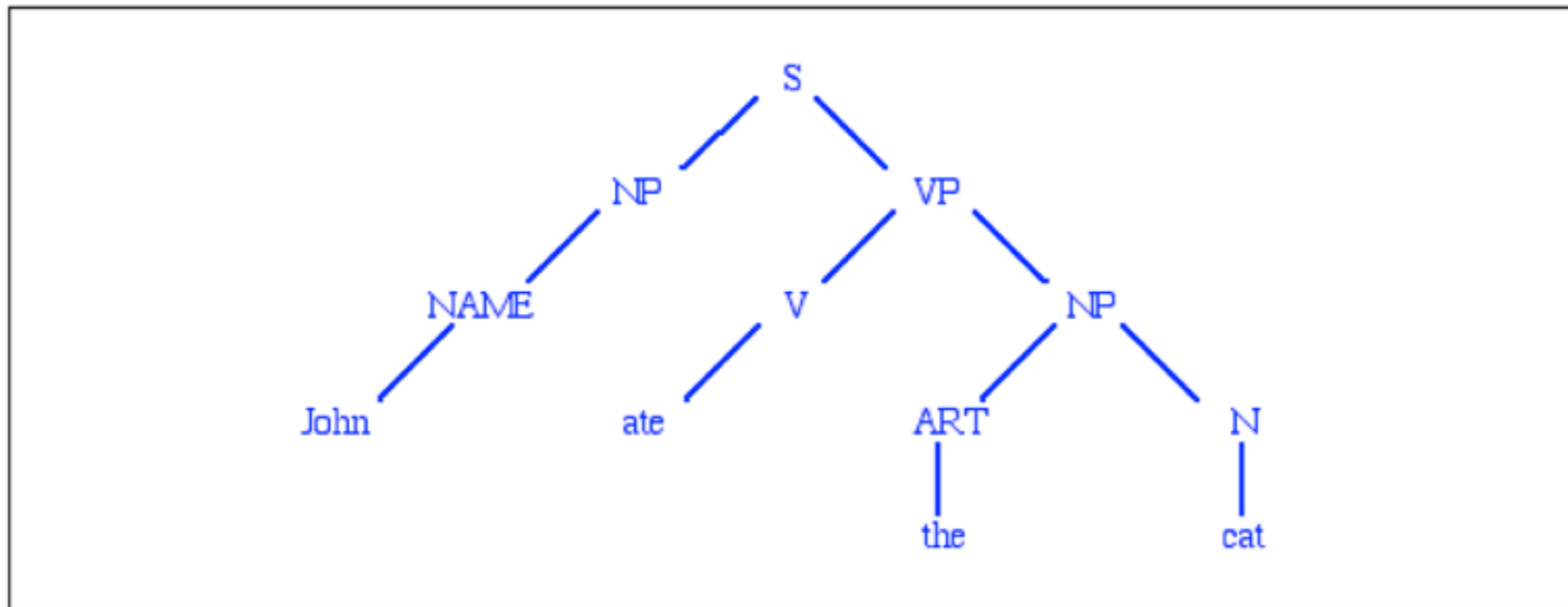
Spouse: *black beans?*

You: *don't you like them.*

# SYNTAX

THE STRUCTURE OF LANGUAGE

# CONTEXT FREE GRAMMARS



**Figure 3.1** A tree representation of *John ate the cat*

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

**S**  
⇒ NP VP (rewriting S)  
⇒ NAME VP (rewriting NP)  
⇒ John VP (rewriting NAME)  
⇒ John V NP (rewriting VP)  
⇒ John ate NP (rewriting V)  
⇒ John ate ART N (rewriting NP)  
⇒ John ate the N (rewriting ART)  
⇒ John ate the cat (rewriting N)

**TOP DOWN**

**BOTTOM UP**

⇒ NAME ate the cat (rewriting John)  
⇒ NAME V the cat (rewriting ate)  
⇒ NAME V ART cat (rewriting the)  
⇒ NAME V ART N (rewriting cat)  
⇒ NP V ART N (rewriting NAME)  
⇒ NP V NP (rewriting ART N)  
⇒ NP VP (rewriting V NP)  
⇒ S (rewriting NP VP)

# TOP DOWN PARSE AS SEARCH

“<sub>1</sub>THE<sub>2</sub>OLD<sub>3</sub>MAN<sub>3</sub>CRIED<sub>5</sub>”

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

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

**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.  $S \rightarrow NP VP$
2.  $NP \rightarrow ART ADJ N$
3.  $NP \rightarrow ART N$
4.  $NP \rightarrow ADJ N$
5.  $VP \rightarrow AUX VP$
6.  $VP \rightarrow V NP$

"THE"

"LARGE"



$NP \rightarrow ART \circ ADJ N$  →

$NP \rightarrow ART \circ N$  →

# 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.  $S \rightarrow NP VP$
2.  $NP \rightarrow ART ADJ N$
3.  $NP \rightarrow ART N$
4.  $NP \rightarrow ADJ N$
5.  $VP \rightarrow AUX VP$
6.  $VP \rightarrow V NP$

"THE"

"LARGE"



$NP \rightarrow ART \circ ADJ N$

$NP \rightarrow ART \circ N$

$NP \rightarrow ADJ \circ N$

$NP \rightarrow ART ADJ \circ N$

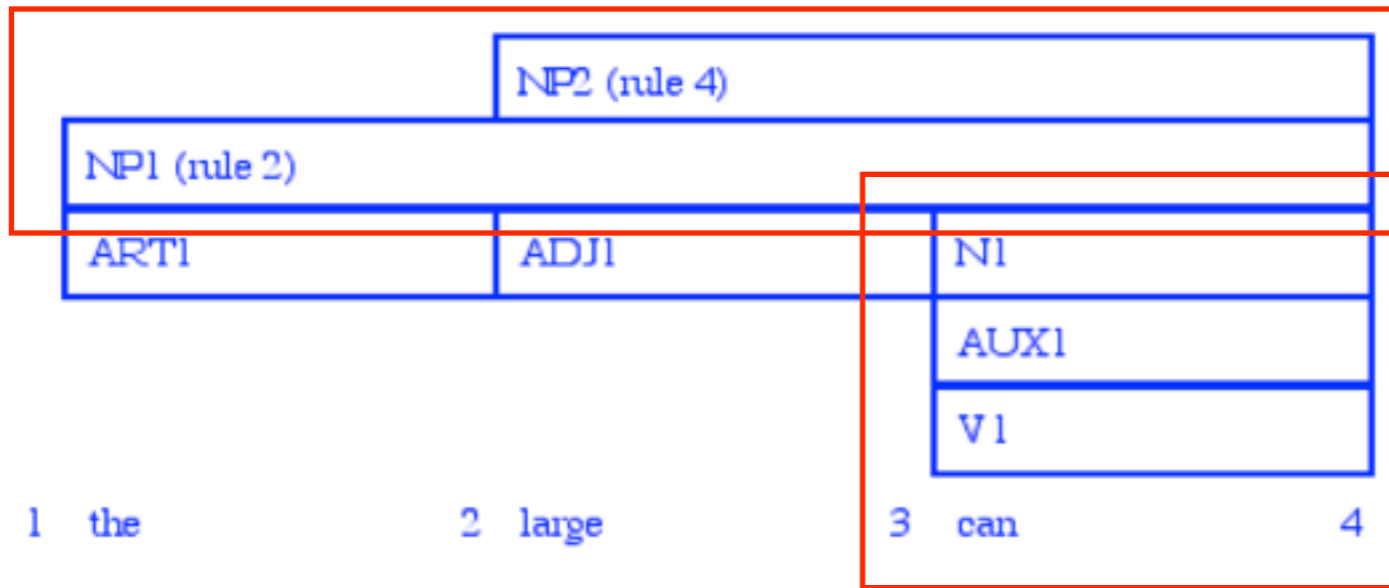
STARTING ARC

EXTENDING ARC  
ARC

# CHART EXAMPLE (2)

## ADDING NEXT WORD: "CAN"

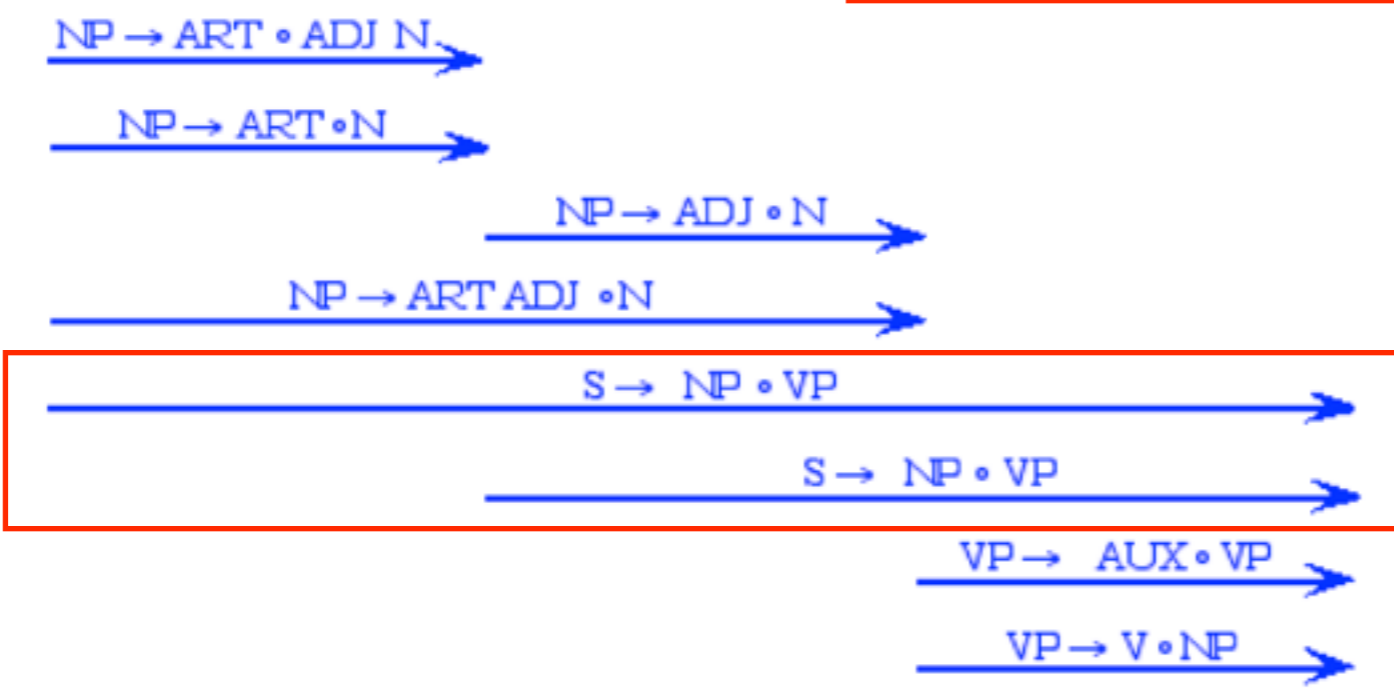
1.  $S \rightarrow NP VP$
2.  $NP \rightarrow ART ADJ N$
3.  $NP \rightarrow ART N$
4.  $NP \rightarrow ADJ N$
5.  $VP \rightarrow AUX VP$
6.  $VP \rightarrow V NP$



**NEW CONSTITUENTS  
(FROM COMPLETING ARCS)**

**NEW LEXICAL  
CONSTITUENTS**

**NEW ACTIVE ARCS  
(EXTENTIONS)**

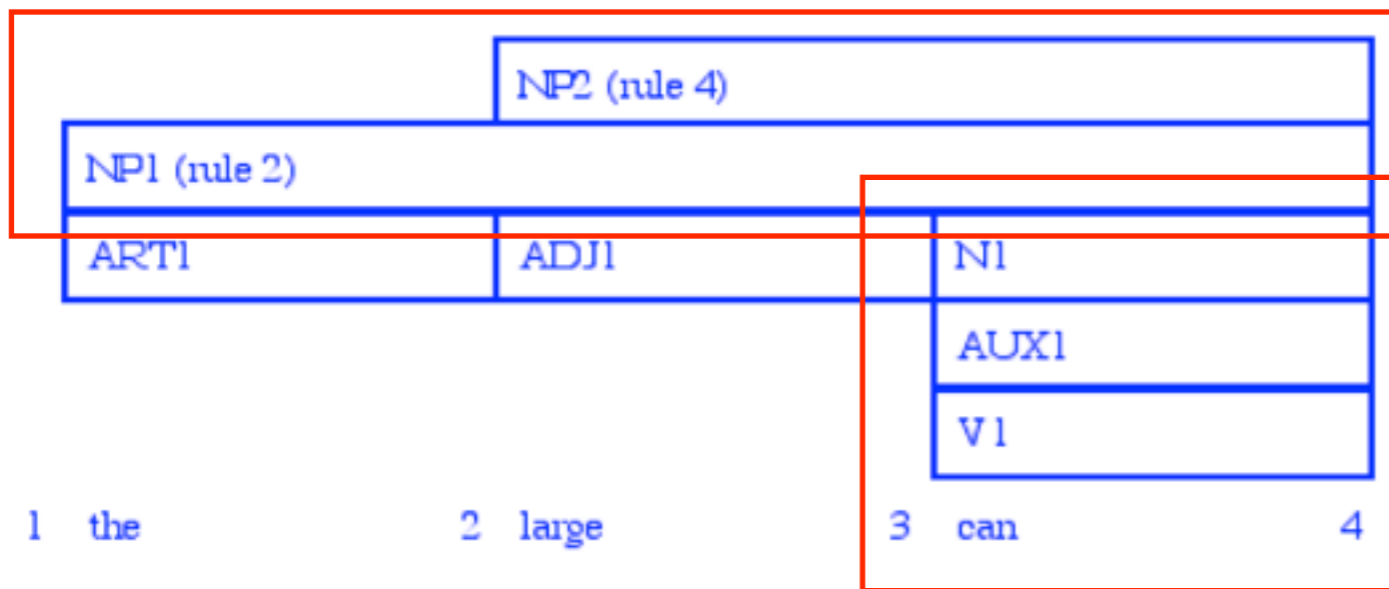




# CHART EXAMPLE (2)

## ADDING NEXT WORD: "CAN"

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



**NEW CONSTITUENTS  
(FROM COMPLETING ARCS)**

**NEW LEXICAL  
CONSTITUENTS**

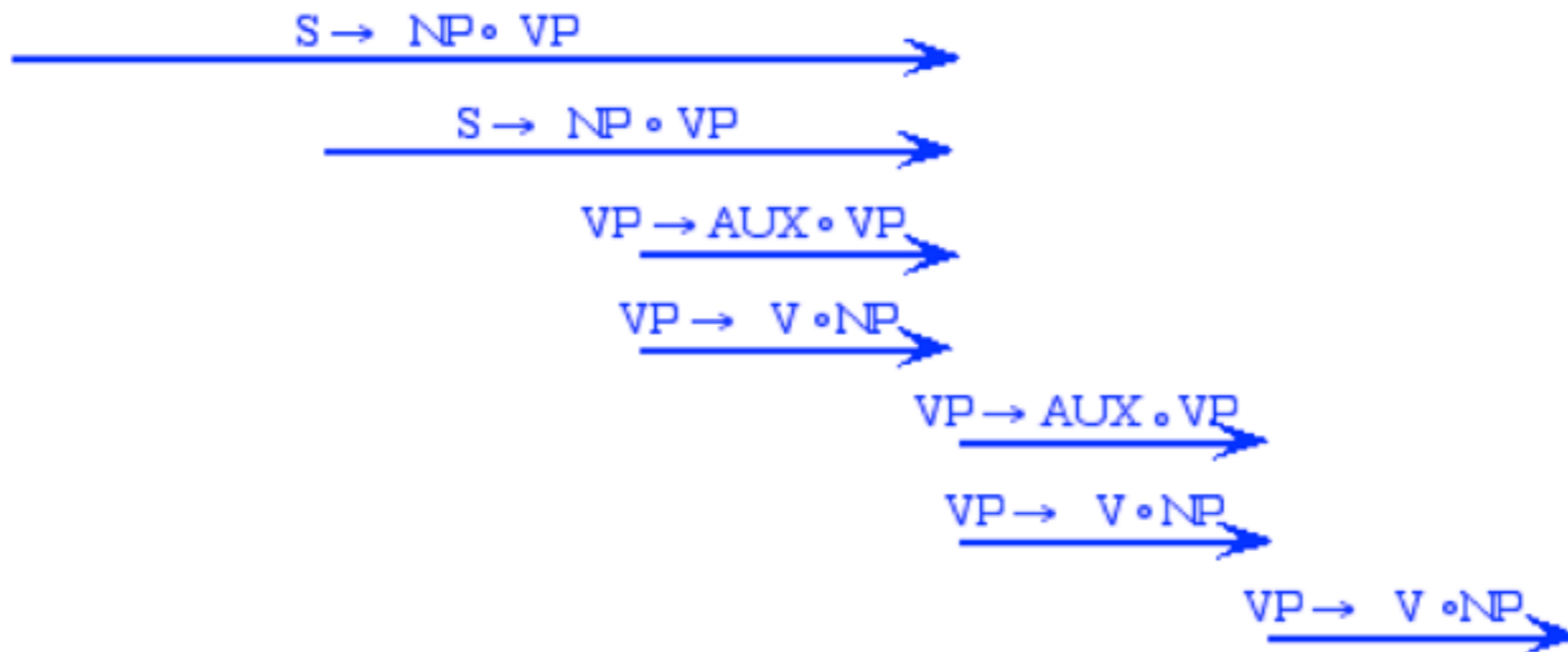
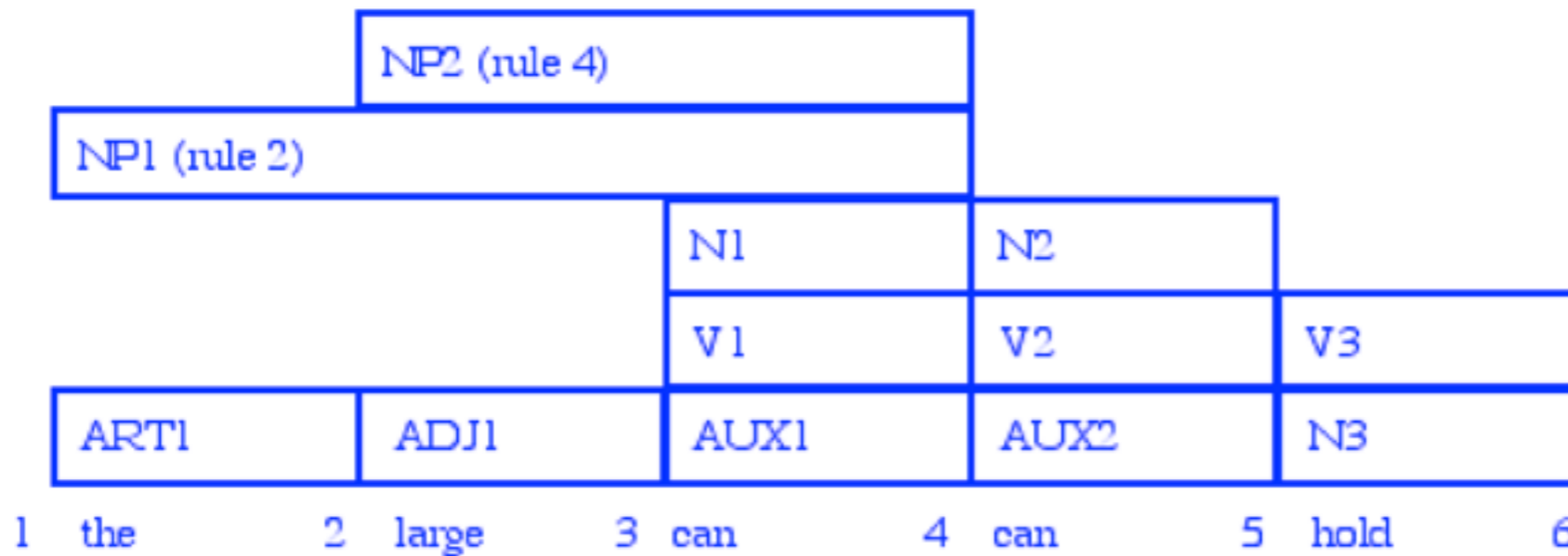
**NEW ACTIVE ARCS  
(EXTENTIONS)**

**NEW ACTIVE ARCS  
(FROM GRAMMAR)**

# CHART EXAMPLE (3)

“THE LARGE CAN CAN HOLD”

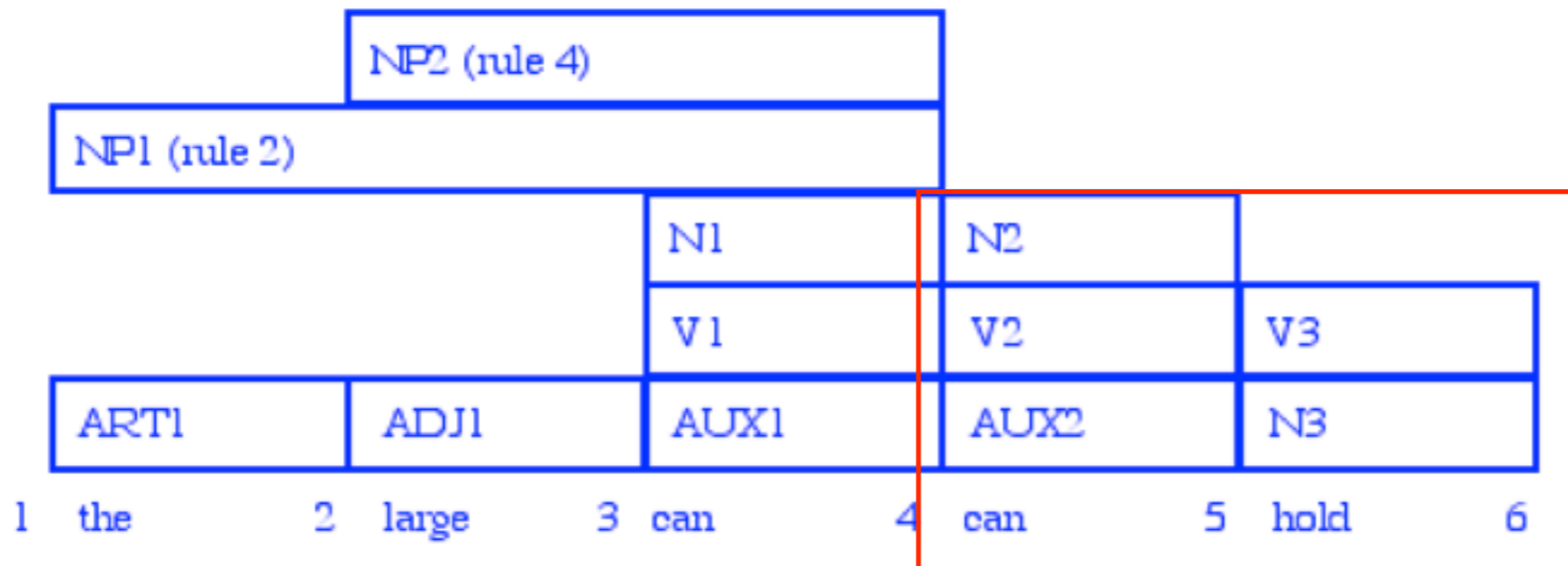
1.  $S \rightarrow NP VP$
2.  $NP \rightarrow ART ADJ N$
3.  $NP \rightarrow ART N$
4.  $NP \rightarrow ADJ N$
5.  $VP \rightarrow AUX VP$
6.  $VP \rightarrow V NP$



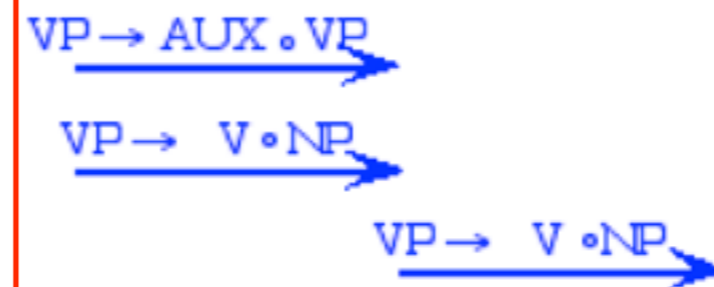
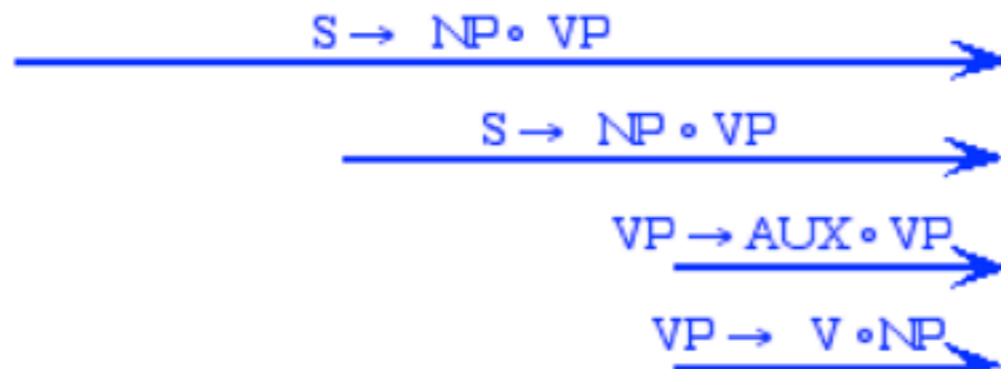
# CHART EXAMPLE (3)

“THE LARGE CAN CAN HOLD”

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



NEW LEXICAL  
CONSTITUENTS

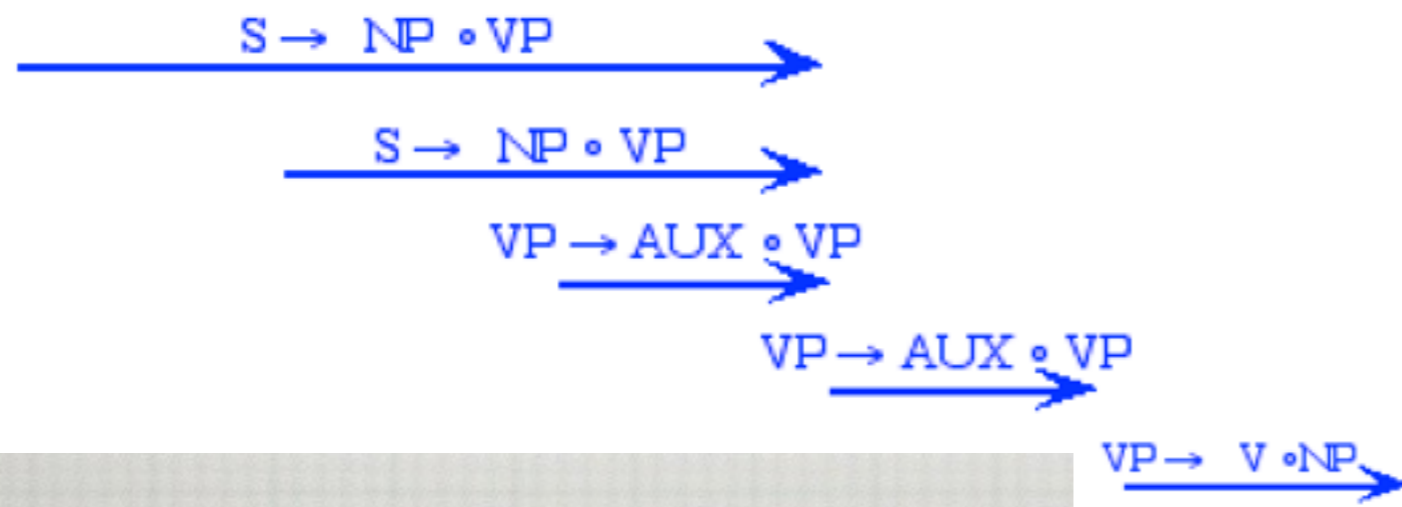


NEW ARCS  
(FROM GRAMMAR)

# CHART EXAMPLE (4)

“THE LARGE CAN CAN HOLD THE WATER”

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



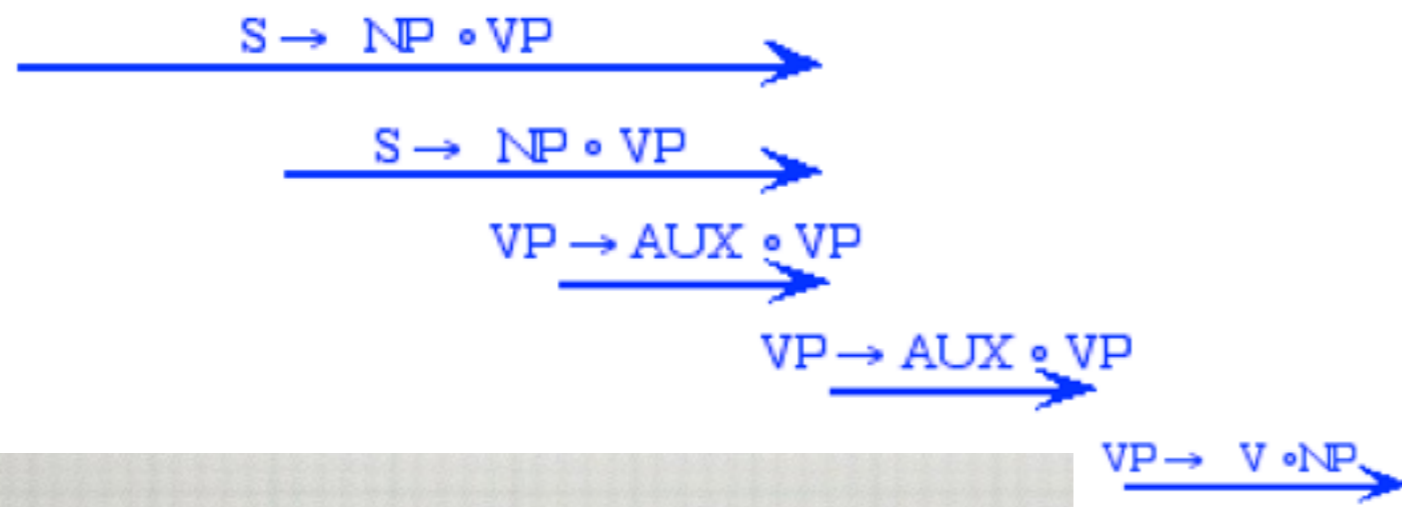
# CHART EXAMPLE (4)

“THE LARGE CAN CAN HOLD THE WATER”

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



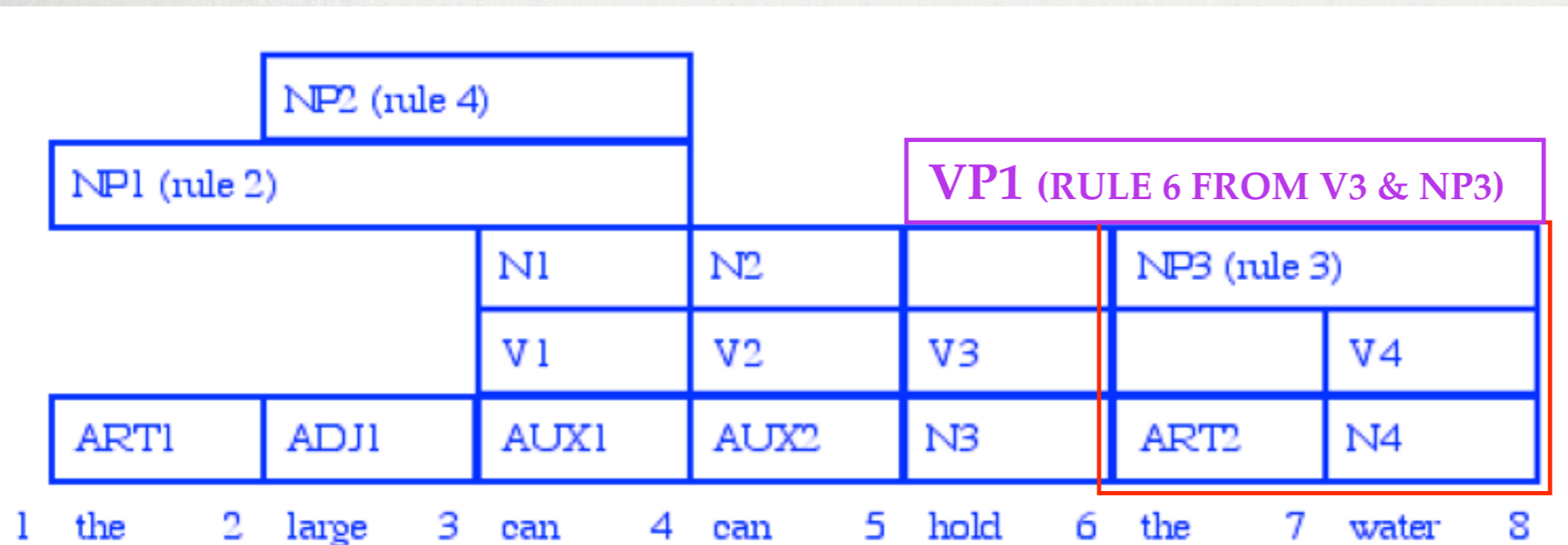
**NEW CONSITUENTS FOR NP “THE WATER”**



# CHART EXAMPLE (4)

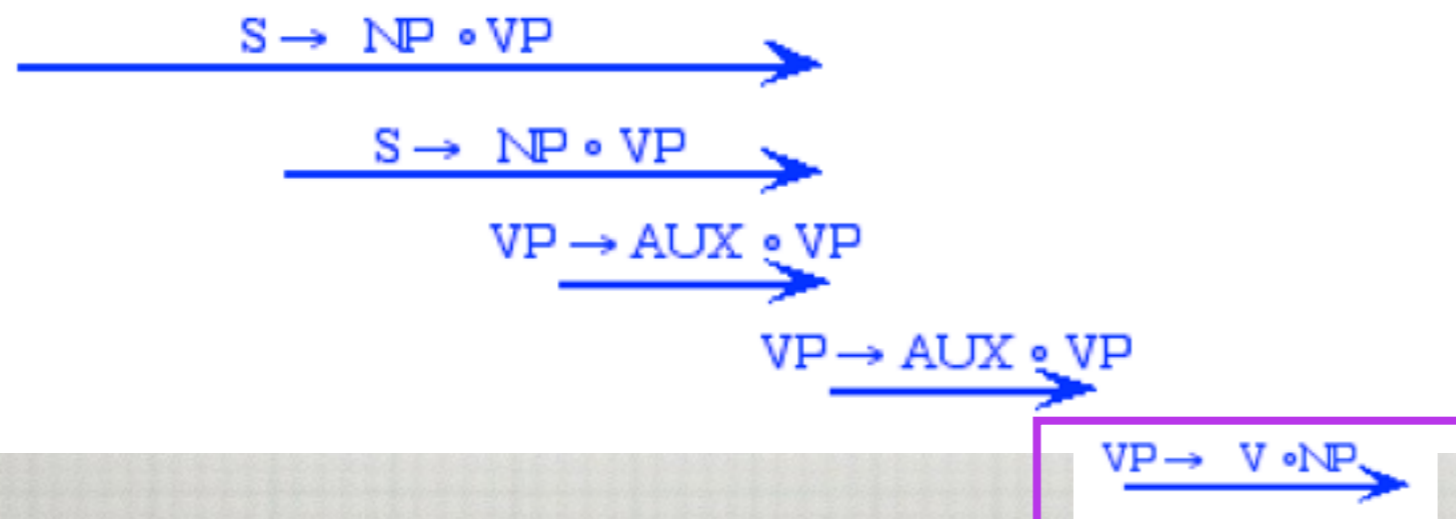
“THE LARGE CAN CAN HOLD THE WATER”

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



ARC COMPLETES

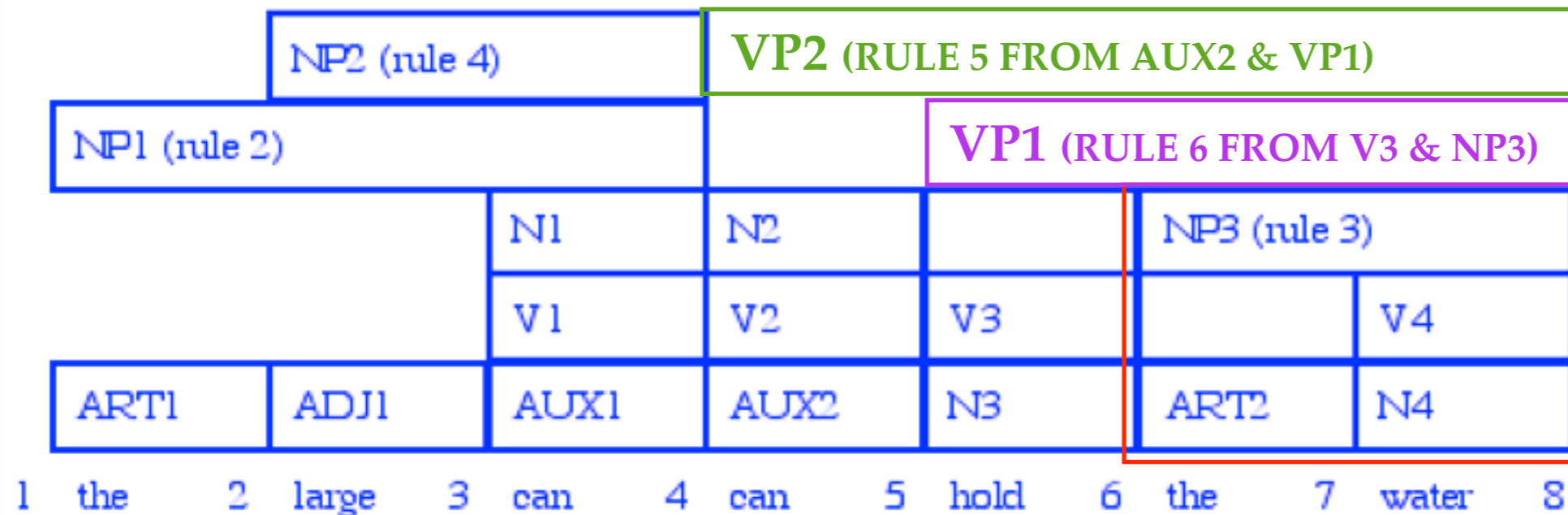
NEW CONSITUENTS FOR NP "THE WATER"



# CHART EXAMPLE (4)

“THE LARGE CAN CAN HOLD THE WATER”

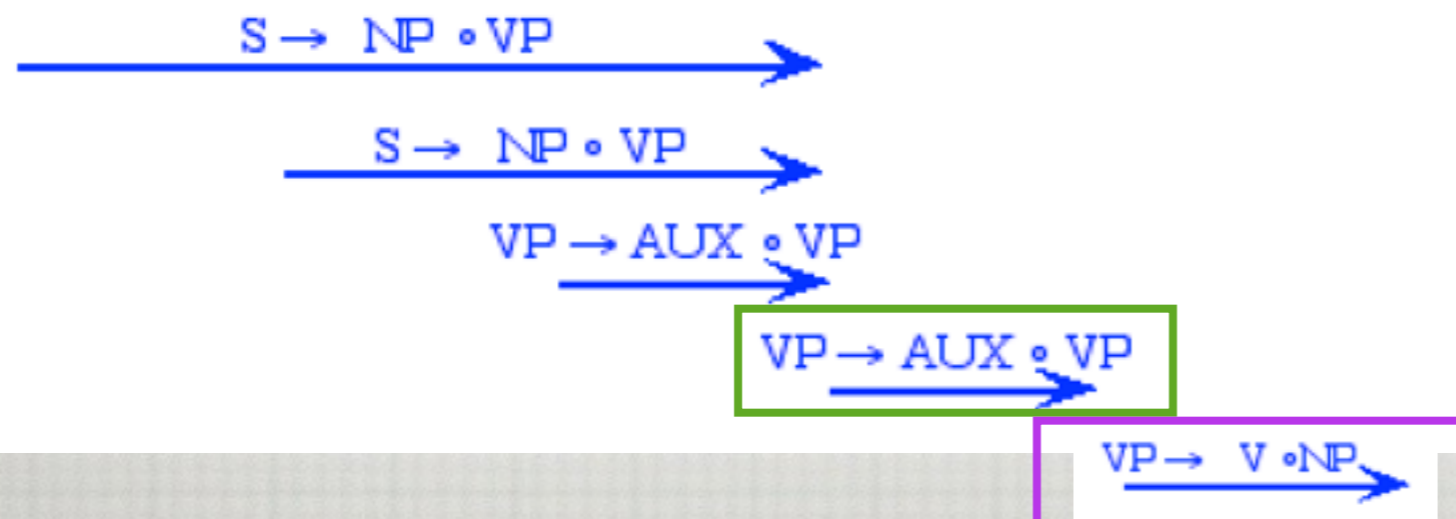
1.  $S \rightarrow NP VP$
2.  $NP \rightarrow ART ADJ N$
3.  $NP \rightarrow ART N$
4.  $NP \rightarrow ADJ N$
5.  $VP \rightarrow AUX VP$
6.  $VP \rightarrow V NP$



ARC COMPLETES

ARC COMPLETES

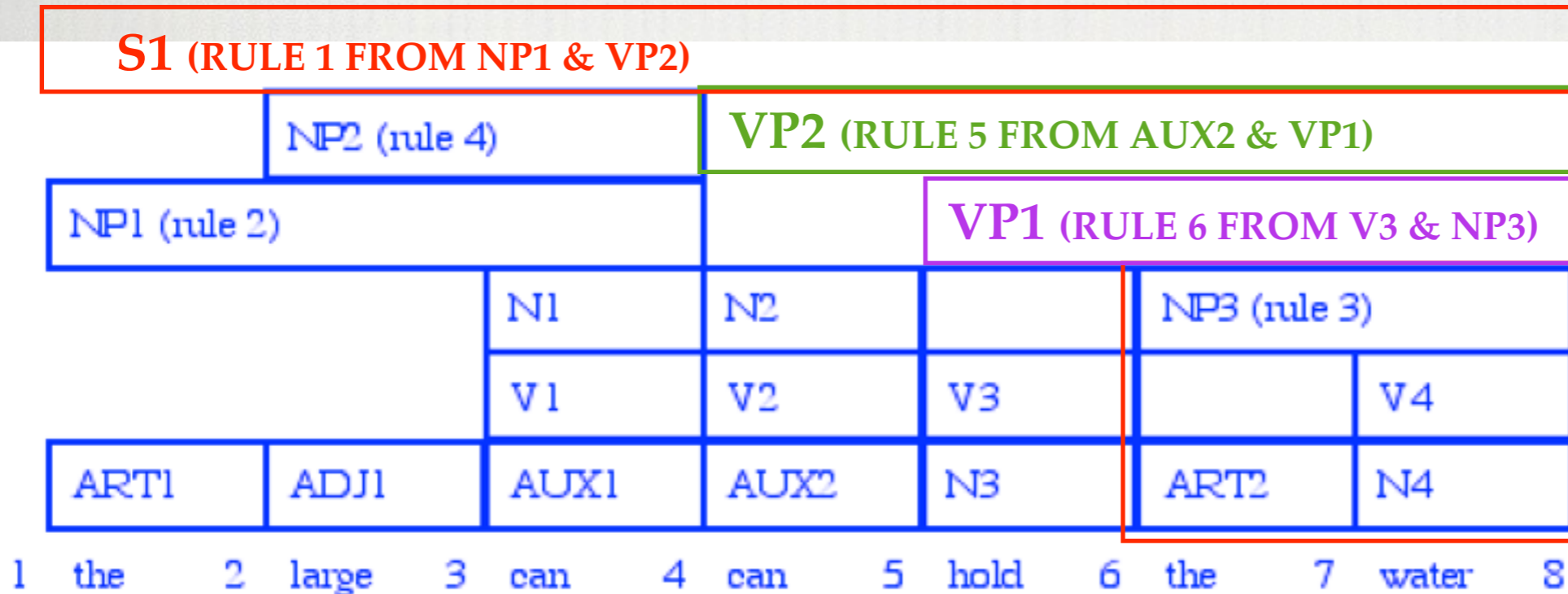
NEW CONSITUENTS FOR NP “THE WATER”



# CHART EXAMPLE (4)

“THE LARGE CAN CAN HOLD THE WATER”

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

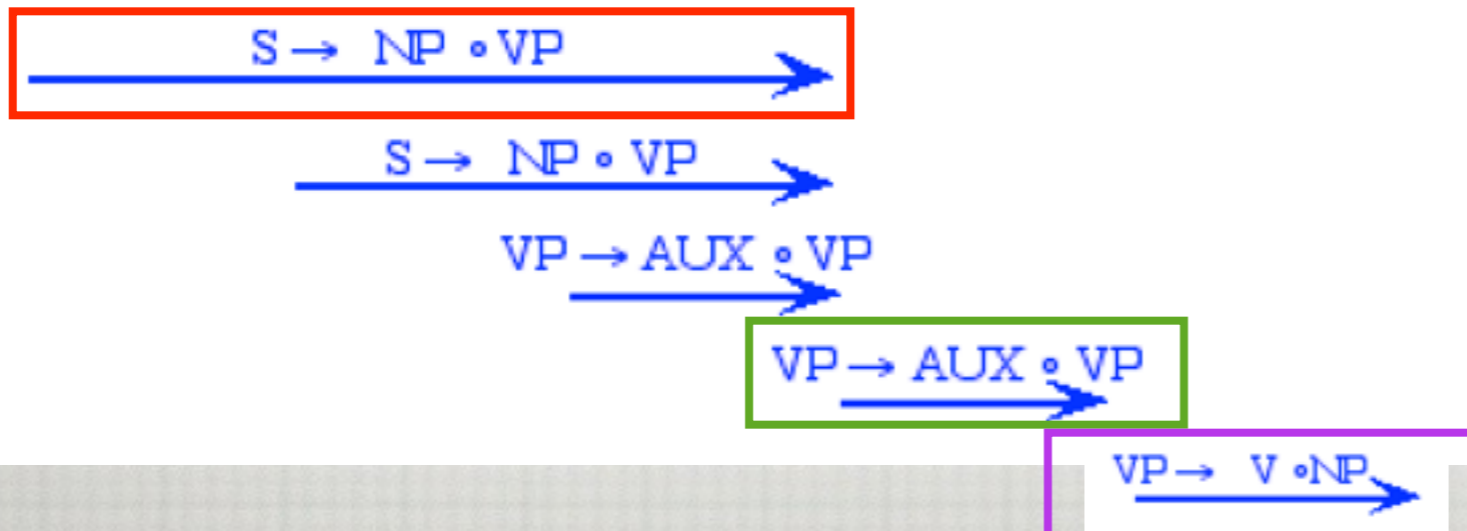


ARC COMPLETES

ARC COMPLETES

ARC COMPLETES

NEW CONSITUENTS FOR NP “THE WATER”





# CHART EXAMPLE (5)

“THE LARGE CAN CAN HOLD THE WATER”

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 COMPLETE CHART

S1 (rule 1 with NP1 and VP2)						
S2 (rule 1 with NP2 and VP2)						
		VP3 (rule 5 with AUX1 and VP2)				
NP2 (rule 4)		VP2 (rule 5)				
NP1 (rule 2)				VP1 (rule 6)		
		N1	N2			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

# TOWARDS PRACTICAL PARSING

---

- **DISAMBIGUATION**

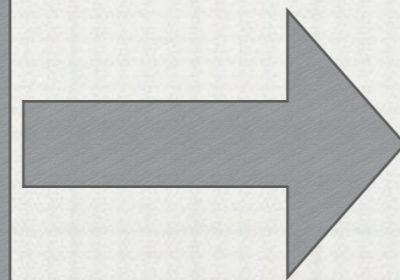
- **THERE MAY BE 100S OF “LEGAL” SYNTACTIC PARSES OF A SENTENCE, WHICH ONE IS RIGHT?**

- **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

$PROB(ART \mid the) \approx .99$	$PROB(N \mid like) \approx .16$
$PROB(N \mid flies) \approx .48$	$PROB(ART \mid a) \approx .995$
$PROB(V \mid flies) \approx .52$	$PROB(N \mid a) \approx .005$
$PROB(V \mid like) \approx .49$	$PROB(N \mid flower) \approx .78$
$PROB(P \mid like) \approx .34$	$PROB(V \mid flower) \approx .22$

**PROBABILISTIC LEXICON**

Rule	Count for LHS	Count for Rule	Probability
1. $S \rightarrow NP VP$	300	300	1
2. $VP \rightarrow V$	300	116	.386
3. $VP \rightarrow V NP$	300	118	.393
4. $VP \rightarrow V NP PP$	300	66	.22
5. $NP \rightarrow NP PP$	1023	241	.24
6. $NP \rightarrow N N$	1023	92	.09
7. $NP \rightarrow N$	1023	141	.14
8. $NP \rightarrow ART N$	1023	558	.55
9. $PP \rightarrow P NP$	307	307	1

**PROBABILISTIC GRAMMAR**

# DISAMBIGUATION: STATISTICAL PARSERS

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

## PROBABILISTIC LEXICON

$PROB(ART   the) \approx .99$	$PROB(N   like) \approx .16$
$PROB(N   flies) \approx .48$	$PROB(ART   a) \approx .995$
$PROB(V   flies) \approx .52$	$PROB(N   a) \approx .005$
$PROB(V   like) \approx .49$	$PROB(N   flower) \approx .78$
$PROB(P   like) \approx .34$	$PROB(V   flower) \approx .22$

Rule	Count for LHS	Count for Rule	Probability
1. $S \rightarrow NP VP$	300	300	1
2. $VP \rightarrow V$	300	116	.386
3. $VP \rightarrow V NP$	300	118	.393
4. $VP \rightarrow V NP PP$	300	66	.22
5. $NP \rightarrow NP PP$	1023	241	.24
6. $NP \rightarrow N N$	1023	92	.09
7. $NP \rightarrow N$	1023	141	.14
8. $NP \rightarrow ART N$	1023	558	.55
9. $PP \rightarrow P NP$	307	307	1

## PROBABILISTIC GRAMMAR

NP425		.14	
1 N422			
NP424			
1 N417			
2 N422		.00011	
NP423			
1 ART416			
2 N422		.54	
S421			
1 NP418			
2 VP420		$3.2 \times 10^{-8}$	
NP418		VP420	
1 N417		.00018	.00018
N417	"A"	.001	N422 "FLOWER" .999
ART416	"A"	.99	V419 "FLOWER" .00047

## PROBABILISTIC CHART

$$\begin{aligned}
 & \text{PROB(CONSTITUENT)} = \\
 & \text{PROB(RULE)} * \text{PROB(SUBCONSTITIT}_1) * \dots * \text{PROB(SUBCONSTITIT}_N)
 \end{aligned}$$

# 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/](http://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"