PCFGs

Scott Farrar
CLMA, University of Washington
farrar@u.washington.edu

January 20, 2010

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Today's lecture

1. Uncertainty
2. Probabilistic Context-Free Grammars
   - Uses of PCFGs: Disambiguation
   - Benefits of PCFGs
   - Issues
3. Probabilistic parsing
   - Probabilistic CKY
4. PCKY Example
5. Homework 3

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Uncertain language

CFGs are fine when the input is 100% predictable. But what about when the input data have never been seen?

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Uncertain language

CFGs are fine when the input is 100% predictable. But what about when the input data have never been seen?

Example

To get out of recession, we must **obamafy** *our economy.*

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Uncertain language

CFGs are fine when the input is 100% predictable. But what about when the input data have never been seen?

Example

To get out of recession, we must **obamafy** *our economy.*

When the input is unknown, then you're dealing with a **random variable**.

Uncertain language

CFGs are fine when the input is 100% predictable. But what about when the input data have never been seen?

Example

To get out of recession, we must **obamafy** *our economy.*

When the input is unknown, then you're dealing with a **random variable**.

- the weather

Uncertain language

CFGs are fine when the input is 100% predictable. But what about when the input data have never been seen?

Example

To get out of recession, we must **obamafy** *our economy.*

When the input is unknown, then you're dealing with a **random variable**.

- the weather
- the stock market

Uncertain language

CFGs are fine when the input is 100% predictable. But what about when the input data have never been seen?

Example

To get out of recession, we must **obamafy** *our economy.*

When the input is unknown, then you're dealing with a **random variable**.

- the weather
- the stock market
- what someone will say next to a speech recognition system

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Uncertain but regular

Strategy for leveraging probability theory for NLP: exploit regularities in known data to predict the structure of unseen data.

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Uncertain but regular

Strategy for leveraging probability theory for NLP: exploit regularities in known data to predict the structure of unseen data.

Example

To get out of recession, we must ????? our economy.

- We can exploit syntax.

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Uncertain but regular

Strategy for leveraging probability theory for NLP: exploit regularities in known data to predict the structure of unseen data.

Example

To get out of recession, we must ????? our economy.

- We can exploit syntax.
- We can exploit morphology.

Uncertain but regular

Strategy for leveraging probability theory for NLP: exploit regularities in known data to predict the structure of unseen data.

Example

To get out of recession, we must ?????fy our economy.

- We can exploit syntax.
- We can exploit morphology.

Today's lecture

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

What is a PCFG?

Definition

Also known as a *weighted* grammar, a probabilistic context-free grammar (PCFG) is one that assigns a probability to each production rule.

What is a PCFG?

Definition

Also known as a *weighted* grammar, a probabilistic context-free grammar (PCFG) is one that assigns a probability to each production rule.

It can be used to assign a probability to every string in the language (language model) and to every structure in the language.

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

What is a PCFG?

Definition

Also known as a *weighted* grammar, a probabilistic context-free grammar (PCFG) is one that assigns a probability to each production rule.

It can be used to assign a probability to every string in the language (language model) and to every structure in the language.

$\mathbf{S}$ = *To get out of recession, we must obamafy the economy.*
$P(\mathbf{S}) = 0.000000004538$

Formal definition of a PCFG

Definition

A CFG grammar is formally defined as $G = \langle N, \Sigma, P, S \rangle$ where:

- N is a set of non-terminal symbols, typically $S, A, B, \ldots$
- S is the starting or goal symbol from N, i.e., $S \in N$
- Σ is a set of terminal symbols, typically $x, y, z, \ldots$ disjoint from N
- P is a set of production rules with attached probabilities.

Formal definition of a PCFG (cont)

Definition

The productions P are of the form: $A \rightarrow \beta \;\; [p]$, where:

- A is a non-terminal $A \in N$
- β is a string of symbols from $(\Sigma \cup N)$
- $[p]$ is the probability (from 0 to 1) that A will have β as its constituent.

Conditional probability

- Thus, p is a **conditional probability**: the probability of some event e_1 given the occurrence of some other event e_2: $P(e_1|e_2)$, e.g., $P(raining|wetroads)$.

Conditional probability

- Thus, p is a **conditional probability**: the probability of some event e_1 given the occurrence of some other event e_2: $P(e_1|e_2)$, e.g., $P(raining|wetroads)$.
- In general then, we can express the conditional probability for any context-free rule in the following way: $P(A \rightarrow \beta)$ or $P(A \rightarrow \beta|A)$ or simply $P(\beta|A)$.

Conditional probability

- Thus, p is a **conditional probability**: the probability of some event e_1 given the occurrence of some other event e_2: $P(e_1|e_2)$, e.g., $P(raining|wetroads)$.
- In general then, we can express the conditional probability for any context-free rule in the following way: $P(A \rightarrow \beta)$ or $P(A \rightarrow \beta|A)$ or simply $P(\beta|A)$.
- In general then: conditional probability for any context-free rule: $P(RHS|LHS)$

Conditional probability

- Thus, p is a **conditional probability**: the probability of some event e_1 given the occurrence of some other event e_2: $P(e_1|e_2)$, e.g., $P(raining|wetroads)$.
- In general then, we can express the conditional probability for any context-free rule in the following way: $P(A \rightarrow \beta)$ or $P(A \rightarrow \beta|A)$ or simply $P(\beta|A)$.
- In general then: conditional probability for any context-free rule: $P(RHS|LHS)$

PCFGs are defined according to a generative perspective: top-down

Conditional probability

- Thus, p is a **conditional probability**: the probability of some event e_1 given the occurrence of some other event e_2: $P(e_1|e_2)$, e.g., $P(raining|wetroads)$.
- In general then, we can express the conditional probability for any context-free rule in the following way: $P(A \rightarrow \beta)$ or $P(A \rightarrow \beta|A)$ or simply $P(\beta|A)$.
- In general then: conditional probability for any context-free rule: $P(RHS|LHS)$

PCFGs are defined according to a generative perspective: top-down

It turns out that the non-generative approach, i.e., $P(LHS|RHS)$, gives a poor parsing model.

Calculating probabilities

Treebank as grammar

A rule probability is calculated by summing the occurrences of rules in a labeled, bracketed treebank. That is, the treebank acts a kind of implicit grammar, and the explicit grammar can be "induced" from labeled and bracketed trees.

$$P(A \rightarrow \beta | A) = \frac{Count(A \rightarrow \beta)}{\sum_{\gamma} Count(A \rightarrow \gamma)} = \frac{Count(A \rightarrow \beta)}{Count(A)}$$

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Probability Distribution

The total probability for a given Nonterminal must sum to one:

$$\sum_{\beta} P(A \rightarrow \beta) = 1$$

In a given corpus:

$S \rightarrow \beta$	occurs 212 times
$S \rightarrow NP\ VP$	occurs 97 times
$S \rightarrow AUX\ NP\ VP$	occurs 43
$S \rightarrow VP$	occurs 58
$S \rightarrow ADJP$	occurs 14 times

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Probability Distribution

The total probability for a given Nonterminal must sum to one:

$$\sum_{\beta} P(A \rightarrow \beta) = 1$$

In a given corpus:

$S \rightarrow \beta$	occurs 212 times	
$S \rightarrow NP\ VP$	occurs 97 times	P=0.4575
$S \rightarrow AUX\ NP\ VP$	occurs 43	
$S \rightarrow VP$	occurs 58	
$S \rightarrow ADJP$	occurs 14 times	

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Probability Distribution

The total probability for a given Nonterminal must sum to one:

$$\sum_{\beta} P(A \rightarrow \beta) = 1$$

In a given corpus:

$S \rightarrow \beta$	occurs 212 times	
$S \rightarrow NP\ VP$	occurs 97 times	P=0.4575
$S \rightarrow AUX\ NP\ VP$	occurs 43	P=0.2028
$S \rightarrow VP$	occurs 58	
$S \rightarrow ADJP$	occurs 14 times	

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Probability Distribution

The total probability for a given Nonterminal must sum to one:

$$\sum_{\beta} P(A \rightarrow \beta) = 1$$

In a given corpus:

$S \rightarrow \beta$	occurs 212 times	
$S \rightarrow NP\ VP$	occurs 97 times	P=0.4575
$S \rightarrow AUX\ NP\ VP$	occurs 43	P=0.2028
$S \rightarrow VP$	occurs 58	P=0.2736
$S \rightarrow ADJP$	occurs 14 times	

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Probability Distribution

The total probability for a given Nonterminal must sum to one:

$$\sum_{\beta} P(A \rightarrow \beta) = 1$$

In a given corpus:

$S \rightarrow \beta$	occurs 212 times	
$S \rightarrow NP\ VP$	occurs 97 times	P=0.4575
$S \rightarrow AUX\ NP\ VP$	occurs 43	P=0.2028
$S \rightarrow VP$	occurs 58	P=0.2736
$S \rightarrow ADJP$	occurs 14 times	P=0.0660

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Probability Distribution

The total probability for a given Nonterminal must sum to one:

$$\sum_{\beta} P(A \rightarrow \beta) = 1$$

In a given corpus:

$S \rightarrow \beta$	occurs 212 times	
$S \rightarrow NP\ VP$	occurs 97 times	P=0.4575
$S \rightarrow AUX\ NP\ VP$	occurs 43	P=0.2028
$S \rightarrow VP$	occurs 58	P=0.2736
$S \rightarrow ADJP$	occurs 14 times	P=0.0660
		$0.9999 \approx 1$

Using PCFGs for disambiguation

Probability of a particular parse tree (its derivation) is defined as:

$$P(T) = \prod_{i=1}^{n} P(RHS_i | LHS_i)$$

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Using PCFGs for disambiguation

Probability of a particular parse tree (its derivation) is defined as:

$$P(T) = \prod_{i=1}^{n} P(RHS_i | LHS_i)$$

Using PCFGs for disambiguation

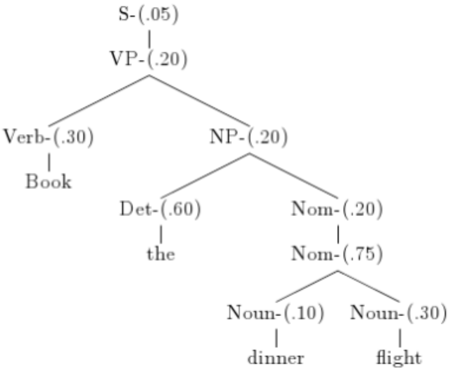
Probability of a particular parse tree (its derivation) is defined as:

$$P(T) = \prod_{i=1}^{n} P(RHS_i | LHS_i)$$

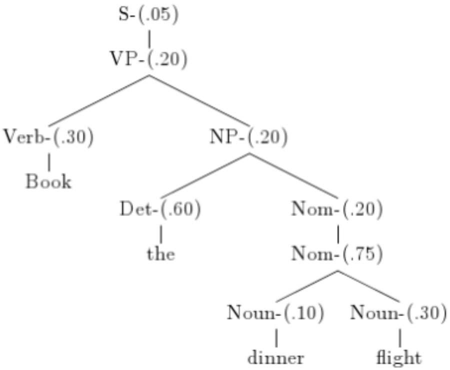
That is, we multiply together all rule expansion probabilities to obtain the probability of the tree.

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Probability of a parse tree 1

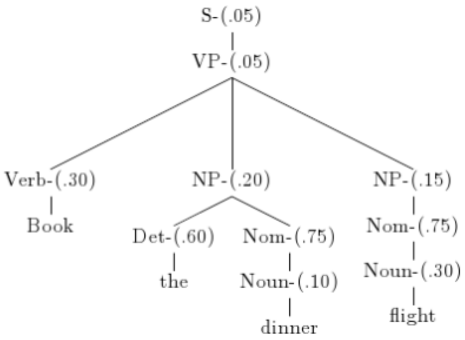


Probability of a parse tree 1

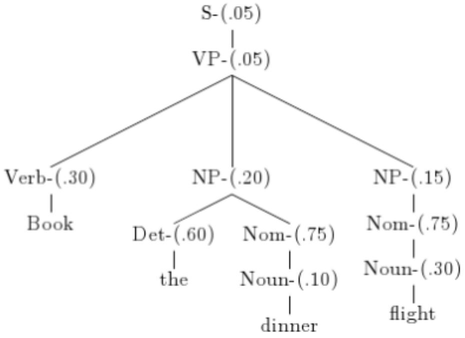


$$.05 \times .20 \times .30 \times .20 \times .60 \times .20 \times .75 \times .10 \times .30 = 1.62 \times 10^{-6}$$

Probability of a parse tree 2

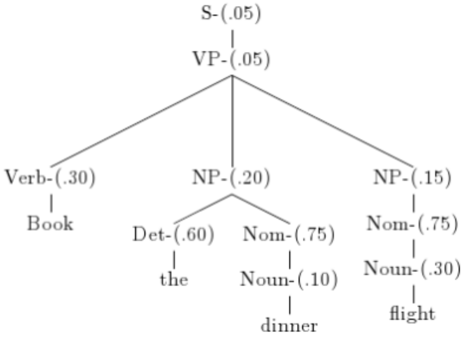


Probability of a parse tree 2



$$.05 * .05 * .30 * .20 * .60 * .75 * .10 * .15 * .75 * .30 = 2.28 \times 10^{-7}$$

Probability of a parse tree 2



$.05 * .05 * .30 * .20 * .60 * .75 * .10 * .15 * .75 * .30 = 2.28 \times 10^{-7}$

Parse tree 1 was greater: $1.62 \times 10^{-6} > 2.28 \times 10^{-7}$

Disambiguation strategy

A disambiguation strategy is to choose the parse with the highest probability:

$$\hat{T}(S) = \arg \max_{yield(T)}$$

- Choose the most likely parse $\hat{T}$ given the $yield(T)$, or the string sequence S
- What exactly is being maximized?
- The probability of the parse tree T given input sentence S.

Benefits of using PCFGs

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Benefits of using PCFGs

- As the number of possible trees for a given input grows, a PCFG gives some idea of the plausibility of a particular parse.

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Benefits of using PCFGs

- As the number of possible trees for a given input grows, a PCFG gives some idea of the plausibility of a particular parse.
- PCFGs help smooth out disfluencies and other errors in the training set.

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Benefits of using PCFGs

- As the number of possible trees for a given input grows, a PCFG gives some idea of the plausibility of a particular parse.
- PCFGs help smooth out disfluencies and other errors in the training set.
- PCFGs are good for inducing grammars from positive data (from parsed treebanks).

Issues

Place invariance

probability of a subtree does not depend on the position of its terminals

Scott Farrar CLMA, University of Washington farrar@u.washington.ed

Issues

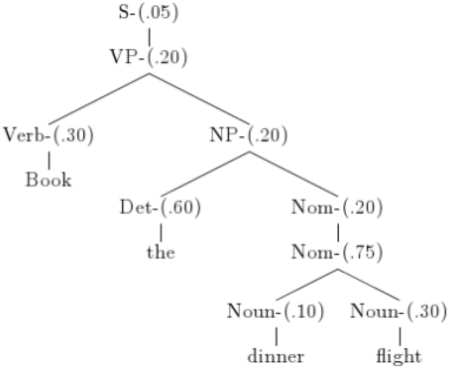
Place invariance

probability of a subtree does not depend on the position of its terminals

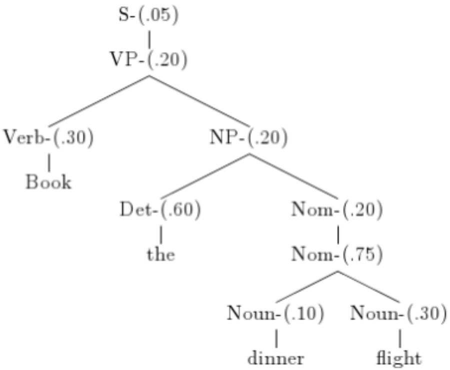
Recall example tree:

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Probability of a parse tree



Probability of a parse tree



$$P(NP) = .20 \times .60 \times .20 \times .75 \times .10 \times .30 = 0.00054$$

Some assumptions with PCFGs

- **place invariance**: probability of a subtree does not depend on the position of its terminals

Some assumptions with PCFGs

- **place invariance**: probability of a subtree does not depend on the position of its terminals
  P(*the dog bit the boy*) = P(*the boy bit the dog*)

Some assumptions with PCFGs

- **place invariance**: probability of a subtree does not depend on the position of its terminals
  P(*the dog bit the boy*) = P(*the boy bit the dog*)
- **context-free**: probability of a subtree does not depend on words not dominated by the subtree

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Some assumptions with PCFGs

- **place invariance**: probability of a subtree does not depend on the position of its terminals
  P(*the dog bit the boy*) = P(*the boy bit the dog*)
- **context-free**: probability of a subtree does not depend on words not dominated by the subtree
  (the bird) swims ..., (the bird) flies ...

Context sensitivity of certain rules

Are subjects usually pronouns or full NPs? what about objects?

Context sensitivity of certain rules

Are subjects usually pronouns or full NPs? what about objects?

	Pronoun	Non-pronoun
Subject	91%	9%
Object	34%	66%

Context sensitivity of certain rules

Are subjects usually pronouns or full NPs? what about objects?

	Pronoun	Non-pronoun
Subject	91%	9%
Object	34%	66%

Example

He saw the accident involving the red car.

Scott Farrar
CLMA, University of Washington far-rar@u.washington.ed

Context sensitivity of certain rules

Are subjects usually pronouns or full NPs? what about objects?

	Pronoun	Non-pronoun
Subject	91%	9%
Object	34%	66%

Example

He saw the accident involving the red car.

should be more likely than:

Example

The man in the street near the bank saw it

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Today's lecture

1. Uncertainty
2. Probabilistic Context-Free Grammars
   - Uses of PCFGs: Disambiguation
   - Benefits of PCFGs
   - Issues
3. Probabilistic parsing
   - Probabilistic CKY
4. PCKY Example
5. Homework 3

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Deterministic parsing

Using deterministic methods, it's difficult to tell which parses are correct.

Scott Farrar
CLMA, University of Washington far-rar@u.washington.ed

Deterministic parsing

Using deterministic methods, it's difficult to tell which parses are correct.

Strategy: prune search space by eliminating suboptimal or improbable ones

Deterministic parsing

Using deterministic methods, it's difficult to tell which parses are correct.

Strategy: prune search space by eliminating suboptimal or improbable ones

Use a PCFG to guide the pruning process; chose the best parse, or *n* best parses.

CKY vs. Prob-CKY

- In the non-probabilistic version, what's contained in a CKY cell?

Scott Farrar
CLMA, University of Washington far-rar@u.washington.ed

CKY vs. Prob-CKY

- In the non-probabilistic version, what's contained in a CKY cell?
- All possible structures for a given span of input; in other words, all possible syntactic interpretations for a given substring.
  ...(time flies)... can be a VP, S, NP, ...

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

CKY vs. Prob-CKY

- In the non-probabilistic version, what's contained in a CKY cell?
- All possible structures for a given span of input; in other words, all possible syntactic interpretations for a given substring.
  ...(time flies)... can be a VP, S, NP, ...
- What if we only need the most likely parse (or top 10 most likely parses) ?

Probabilistic-CKY

function PROBABILISTIC-CKY(*words*, *grammar*) **returns** most probable *parse*, probability
for j $\leftarrow$ **from** 1 **to** LENGTH(*words*) **do**
for all $\{A | A \rightarrow words[j] \in grammar\}$
$table[j-1, j, A] \leftarrow P(A \rightarrow words[j])$
for $i \leftarrow$ **from** $j-2$ **downto** 0 **do**
for $k \leftarrow i+1$ **to** $j-1$ **do**
for all $\{A | A \rightarrow B \ \ C \in grammar,$
and $table[i,k,B] > 0$ and $table[k,j,C] > 0\}$
if $table[i,j,A] < P(A \rightarrow B \ \ C) \times table[i,k,B] \times table[k,j,C])$
$table[i,j,A] \leftarrow P(A \rightarrow B \ \ C) \times table[i,k,B] \times table[k,j,C])$
$back[i,j,A] \leftarrow \{k, B, C\}$
return BUILD_TREE(*back*[1,LENGTH(*words*), *S*]), *table*[1,LENGTH(*words*), *S*]

See pcky_eg.pdf.

Today's lecture

1. Uncertainty
2. Probabilistic Context-Free Grammars
   - Uses of PCFGs: Disambiguation
   - Benefits of PCFGs
   - Issues
3. Probabilistic parsing
   - Probabilistic CKY
4. PCKY Example
5. Homework 3

CKY example

a 1	pilot 2	likes 3	flying 4	planes 5

$S \longrightarrow NP\ VP$
$VP \longrightarrow VBG\ NNS$
$VP \longrightarrow VBZ\ VP$
$VP \longrightarrow VBZ\ NP$
$NP \longrightarrow DT\ NN$
$NP \longrightarrow JJ\ NNS$
$DT \longrightarrow a$
$NN \longrightarrow pilot$
$VBZ \longrightarrow likes$
$VBG \longrightarrow flying$
$JJ \longrightarrow flying$
$NNS \longrightarrow planes$

CKY example

a	pilot	likes	flying	planes
1	2	3	4	5
DT				

#

$S \longrightarrow NP\ VP$
$VP \longrightarrow VBG\ NNS$
$VP \longrightarrow VBZ\ VP$
$VP \longrightarrow VBZ\ NP$
$NP \longrightarrow DT\ NN$
$NP \longrightarrow JJ\ NNS$
$DT \longrightarrow a$
$NN \longrightarrow pilot$
$VBZ \longrightarrow likes$
$VBG \longrightarrow flying$
$JJ \longrightarrow flying$
$NNS \longrightarrow planes$

CKY example

a 1	**pilot** 2	likes 3	flying 4	planes 5
DT				
	NN			

$S \longrightarrow NP\ VP$
$VP \longrightarrow VBG\ NNS$
$VP \longrightarrow VBZ\ VP$
$VP \longrightarrow VBZ\ NP$
$NP \longrightarrow DT\ NN$
$NP \longrightarrow JJ\ NNS$
$DT \longrightarrow a$
\# $NN \longrightarrow pilot$
$VBZ \longrightarrow likes$
$VBG \longrightarrow flying$
$JJ \longrightarrow flying$
$NNS \longrightarrow planes$

CKY example

a 1	**pilot** 2	likes 3	flying 4	planes 5
DT	NP			
	NN			

#

$S \longrightarrow NP\ VP$
$VP \longrightarrow VBG\ NNS$
$VP \longrightarrow VBZ\ VP$
$VP \longrightarrow VBZ\ NP$
$NP \longrightarrow DT\ NN$
$NP \longrightarrow JJ\ NNS$
$DT \longrightarrow a$
$NN \longrightarrow pilot$
$VBZ \longrightarrow likes$
$VBG \longrightarrow flying$
$JJ \longrightarrow flying$
$NNS \longrightarrow planes$

CKY example

a 1	pilot 2	likes 3	flying 4	planes 5
DT	NP			
	NN			
		VBZ		

#

$S \longrightarrow NP\ VP$
$VP \longrightarrow VBG\ NNS$
$VP \longrightarrow VBZ\ VP$
$VP \longrightarrow VBZ\ NP$
$NP \longrightarrow DT\ NN$
$NP \longrightarrow JJ\ NNS$
$DT \longrightarrow a$
$NN \longrightarrow pilot$
$VBZ \longrightarrow likes$
$VBG \longrightarrow flying$
$JJ \longrightarrow flying$
$NNS \longrightarrow planes$

CKY example

a 1	pilot 2	likes 3	flying 4	planes 5
DT	NP			
	NN	-		
		VBZ		

$S \longrightarrow NP\ VP$
$VP \longrightarrow VBG\ NNS$
$VP \longrightarrow VBZ\ VP$
$VP \longrightarrow VBZ\ NP$
$NP \longrightarrow DT\ NN$
$NP \longrightarrow JJ\ NNS$
$DT \longrightarrow a$
$NN \longrightarrow pilot$
$VBZ \longrightarrow likes$
$VBG \longrightarrow flying$
$JJ \longrightarrow flying$
$NNS \longrightarrow planes$

CKY example

a 1	pilot 2	likes 3	flying 4	planes 5
DT	NP	-		
	NN	-		
		VBZ		

$S \longrightarrow NP\ VP$
$VP \longrightarrow VBG\ NNS$
$VP \longrightarrow VBZ\ VP$
$VP \longrightarrow VBZ\ NP$
$NP \longrightarrow DT\ NN$
$NP \longrightarrow JJ\ NNS$
$DT \longrightarrow a$
$NN \longrightarrow pilot$
$VBZ \longrightarrow likes$
$VBG \longrightarrow flying$
$JJ \longrightarrow flying$
$NNS \longrightarrow planes$

CKY example

a 1	pilot 2	likes 3	flying 4	planes 5
DT	NP	-		
	NN	-		
		VBZ		
			JJ	

$S \longrightarrow NP\ VP$
$VP \longrightarrow VBG\ NNS$
$VP \longrightarrow VBZ\ VP$
$VP \longrightarrow VBZ\ NP$
$NP \longrightarrow DT\ NN$
$NP \longrightarrow JJ\ NNS$
$DT \longrightarrow a$
$NN \longrightarrow pilot$
$VBZ \longrightarrow likes$
$VBG \longrightarrow flying$
\# $JJ \longrightarrow flying$
$NNS \longrightarrow planes$

CKY example

a 1	pilot 2	likes 3	flying 4	planes 5
DT	NP	-		
	NN	-		
		VBZ		
			JJ VBG	

#

$S \longrightarrow NP\ VP$
$VP \longrightarrow VBG\ NNS$
$VP \longrightarrow VBZ\ VP$
$VP \longrightarrow VBZ\ NP$
$NP \longrightarrow DT\ NN$
$NP \longrightarrow JJ\ NNS$
$DT \longrightarrow a$
$NN \longrightarrow pilot$
$VBZ \longrightarrow likes$
$VBG \longrightarrow flying$
$JJ \longrightarrow flying$
$NNS \longrightarrow planes$

CKY example

a 1	pilot 2	likes 3	flying 4	planes 5
DT	NP	-		
	NN	-		
		VBZ	-	
			JJ VBG	

$S \longrightarrow NP\ VP$
$VP \longrightarrow VBG\ NNS$
$VP \longrightarrow VBZ\ VP$
$VP \longrightarrow VBZ\ NP$
$NP \longrightarrow DT\ NN$
$NP \longrightarrow JJ\ NNS$
$DT \longrightarrow a$
$NN \longrightarrow pilot$
$VBZ \longrightarrow likes$
$VBG \longrightarrow flying$
$JJ \longrightarrow flying$
$NNS \longrightarrow planes$

CKY example

a 1	pilot 2	likes 3	flying 4	planes 5
DT	NP	-		
	NN	-	-	
		VBZ	-	
			JJ VBG	

$S \longrightarrow NP\ VP$
$VP \longrightarrow VBG\ NNS$
$VP \longrightarrow VBZ\ VP$
$VP \longrightarrow VBZ\ NP$
$NP \longrightarrow DT\ NN$
$NP \longrightarrow JJ\ NNS$
$DT \longrightarrow a$
$NN \longrightarrow pilot$
$VBZ \longrightarrow likes$
$VBG \longrightarrow flying$
$JJ \longrightarrow flying$
$NNS \longrightarrow planes$

CKY example

a 1	pilot 2	likes 3	flying 4	planes 5
DT	NP	-	-	
	NN	-	-	
		VBZ	-	
			JJ VBG	

$S \longrightarrow NP\ VP$
$VP \longrightarrow VBG\ NNS$
$VP \longrightarrow VBZ\ VP$
$VP \longrightarrow VBZ\ NP$
$NP \longrightarrow DT\ NN$
$NP \longrightarrow JJ\ NNS$
$DT \longrightarrow a$
$NN \longrightarrow pilot$
$VBZ \longrightarrow likes$
$VBG \longrightarrow flying$
$JJ \longrightarrow flying$
$NNS \longrightarrow planes$

CKY example

a 1	pilot 2	likes 3	flying 4	planes 5
DT	NP	-	-	
	NN	-	-	
		VBZ	-	
			JJ VBG	
				NNS

S → *NP VP*
VP → *VBG NNS*
VP → *VBZ VP*
VP → *VBZ NP*
NP → *DT NN*
NP → *JJ NNS*
DT → *a*
NN → *pilot*
VBZ → *likes*
VBG → *flying*
JJ → *flying*
NNS → *planes*

CKY example

a 1	pilot 2	likes 3	flying 4	planes 5
DT	NP	-	-	
	NN	-	-	
		VBZ	-	
			JJ VBG	NP
				NNS

#

$S \longrightarrow NP\ VP$
$VP \longrightarrow VBG\ NNS$
$VP \longrightarrow VBZ\ VP$
$VP \longrightarrow VBZ\ NP$
$NP \longrightarrow DT\ NN$
$NP \longrightarrow JJ\ NNS$
$DT \longrightarrow a$
$NN \longrightarrow pilot$
$VBZ \longrightarrow likes$
$VBG \longrightarrow flying$
$JJ \longrightarrow flying$
$NNS \longrightarrow planes$

CKY example

a 1	pilot 2	likes 3	flying 4	planes 5
DT	NP	-	-	
	NN	-	-	
		VBZ	-	
			JJ VBG	NP VP
				NNS

- $S \longrightarrow NP\ VP$
- \# $VP \longrightarrow VBG\ NNS$
- $VP \longrightarrow VBZ\ VP$
- $VP \longrightarrow VBZ\ NP$
- $NP \longrightarrow DT\ NN$
- $NP \longrightarrow JJ\ NNS$
- $DT \longrightarrow a$
- $NN \longrightarrow pilot$
- $VBZ \longrightarrow likes$
- $VBG \longrightarrow flying$
- $JJ \longrightarrow flying$
- $NNS \longrightarrow planes$

CKY example

a 1	pilot 2	likes 3	flying 4	planes 5
DT	NP	-	-	
	NN	-	-	
		VBZ	-	VP
			JJ VBG	NP VP
				NNS

#

$S \longrightarrow NP\ VP$
$VP \longrightarrow VBG\ NNS$
$VP \longrightarrow VBZ\ VP$
$VP \longrightarrow VBZ\ NP$
$NP \longrightarrow DT\ NN$
$NP \longrightarrow JJ\ NNS$
$DT \longrightarrow a$
$NN \longrightarrow pilot$
$VBZ \longrightarrow likes$
$VBG \longrightarrow flying$
$JJ \longrightarrow flying$
$NNS \longrightarrow planes$

CKY example

a 1	pilot 2	likes 3	flying 4	planes 5
DT	NP	-	-	
	NN	-	-	
		VBZ	-	VP VP
			JJ VBG	NP VP
				NNS

#

$S \longrightarrow NP\ VP$
$VP \longrightarrow VBG\ NNS$
$VP \longrightarrow VBZ\ VP$
$VP \longrightarrow VBZ\ NP$
$NP \longrightarrow DT\ NN$
$NP \longrightarrow JJ\ NNS$
$DT \longrightarrow a$
$NN \longrightarrow pilot$
$VBZ \longrightarrow likes$
$VBG \longrightarrow flying$
$JJ \longrightarrow flying$
$NNS \longrightarrow planes$

CKY example

a 1	pilot 2	likes 3	flying 4	planes 5
DT	NP	-	-	
	NN	-	-	-
		VBZ	-	VP VP
			JJ VBG	NP VP
				NNS

$S \longrightarrow NP\ VP$
$VP \longrightarrow VBG\ NNS$
$VP \longrightarrow VBZ\ VP$
$VP \longrightarrow VBZ\ NP$
$NP \longrightarrow DT\ NN$
$NP \longrightarrow JJ\ NNS$
$DT \longrightarrow a$
$NN \longrightarrow pilot$
$VBZ \longrightarrow likes$
$VBG \longrightarrow flying$
$JJ \longrightarrow flying$
$NNS \longrightarrow planes$

CKY example

a 1	pilot 2	likes 3	flying 4	planes 5
DT	NP	-	-	S
	NN	-	-	-
		VBZ	-	VP VP
			JJ VBG	NP VP
				NNS

\# $S \longrightarrow NP\ VP$
$VP \longrightarrow VBG\ NNS$
$VP \longrightarrow VBZ\ VP$
$VP \longrightarrow VBZ\ NP$
$NP \longrightarrow DT\ NN$
$NP \longrightarrow JJ\ NNS$
$DT \longrightarrow a$
$NN \longrightarrow pilot$
$VBZ \longrightarrow likes$
$VBG \longrightarrow flying$
$JJ \longrightarrow flying$
$NNS \longrightarrow planes$

CKY example

a 1	pilot 2	likes 3	flying 4	planes 5
DT	NP	-	-	S S
	NN	-	-	-
		VBZ	-	VP VP
			JJ VBG	NP VP
				NNS

$S \longrightarrow NP\ VP$
$VP \longrightarrow VBG\ NNS$
$VP \longrightarrow VBZ\ VP$
$VP \longrightarrow VBZ\ NP$
$NP \longrightarrow DT\ NN$
$NP \longrightarrow JJ\ NNS$
$DT \longrightarrow a$
$NN \longrightarrow pilot$
$VBZ \longrightarrow likes$
$VBG \longrightarrow flying$
$JJ \longrightarrow flying$
$NNS \longrightarrow planes$

Probabilistic CKY: all parses

a 1	pilot 2	likes 3	flying 4	planes 5

$S \longrightarrow NP\ VP$	[1.0]
$VP \longrightarrow VBG\ NNS$	[0.1]
$VP \longrightarrow VBZ\ VP$	[0.1]
$VP \longrightarrow VBZ\ NP$	[0.3]
$NP \longrightarrow DT\ NN$	[0.3]
$NP \longrightarrow JJ\ NNS$	[0.4]
$DT \longrightarrow a$	[0.3]
$NN \longrightarrow pilot$	[0.1]
$VBZ \longrightarrow likes$	[0.4]
$VBG \longrightarrow flying$	[0.5]
$JJ \longrightarrow flying$	[0.1]
$NNS \longrightarrow planes$	[.34]

Probabilistic CKY: all parses

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]				

	$S \longrightarrow NP\ VP$	[1.0]
	$VP \longrightarrow VBG\ NNS$	[0.1]
	$VP \longrightarrow VBZ\ VP$	[0.1]
	$VP \longrightarrow VBZ\ NP$	[0.3]
	$NP \longrightarrow DT\ NN$	[0.3]
	$NP \longrightarrow JJ\ NNS$	[0.4]
#	$DT \longrightarrow a$	[0.3]
	$NN \longrightarrow pilot$	[0.1]
	$VBZ \longrightarrow likes$	[0.4]
	$VBG \longrightarrow flying$	[0.5]
	$JJ \longrightarrow flying$	[0.1]
	$NNS \longrightarrow planes$	[.34]

Probabilistic CKY: all parses

a 1	**pilot** 2	likes 3	flying 4	planes 5
DT [0.3]				
	NN [0.1]			

	$S \longrightarrow NP\ VP$		[1.0]
	$VP \longrightarrow VBG\ NNS$		[0.1]
	$VP \longrightarrow VBZ\ VP$		[0.1]
	$VP \longrightarrow VBZ\ NP$		[0.3]
	$NP \longrightarrow DT\ NN$		[0.3]
	$NP \longrightarrow JJ\ NNS$		[0.4]
	$DT \longrightarrow a$		[0.3]
#	$NN \longrightarrow pilot$		[0.1]
	$VBZ \longrightarrow likes$		[0.4]
	$VBG \longrightarrow flying$		[0.5]
	$JJ \longrightarrow flying$		[0.1]
	$NNS \longrightarrow planes$		[.34]

Probabilistic CKY: all parses

a 1	**pilot** 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]			
	NN [0.1]			

	Rule	Prob
	$S \longrightarrow NP\ VP$	[1.0]
	$VP \longrightarrow VBG\ NNS$	[0.1]
	$VP \longrightarrow VBZ\ VP$	[0.1]
	$VP \longrightarrow VBZ\ NP$	[0.3]
#	$NP \longrightarrow DT\ NN$	[0.3]
	$NP \longrightarrow JJ\ NNS$	[0.4]
	$DT \longrightarrow a$	[0.3]
	$NN \longrightarrow pilot$	[0.1]
	$VBZ \longrightarrow likes$	[0.4]
	$VBG \longrightarrow flying$	[0.5]
	$JJ \longrightarrow flying$	[0.1]
	$NNS \longrightarrow planes$	[.34]

0.3 x 0.1 x 0.3 = 0.009

Probabilistic CKY: all parses

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]			
	NN [0.1]			
		VBZ [0.4]		

$S \longrightarrow NP\ VP$	[1.0]
$VP \longrightarrow VBG\ NNS$	[0.1]
$VP \longrightarrow VBZ\ VP$	[0.1]
$VP \longrightarrow VBZ\ NP$	[0.3]
$NP \longrightarrow DT\ NN$	[0.3]
$NP \longrightarrow JJ\ NNS$	[0.4]
$DT \longrightarrow a$	[0.3]
$NN \longrightarrow pilot$	[0.1]
#	$VBZ \longrightarrow likes$	[0.4]
$VBG \longrightarrow flying$	[0.5]
$JJ \longrightarrow flying$	[0.1]
$NNS \longrightarrow planes$	[.34]

Probabilistic CKY: all parses

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]			
	NN [0.1]	-		
		VBZ [0.4]		

$S \longrightarrow NP\ VP$	[1.0]
$VP \longrightarrow VBG\ NNS$	[0.1]
$VP \longrightarrow VBZ\ VP$	[0.1]
$VP \longrightarrow VBZ\ NP$	[0.3]
$NP \longrightarrow DT\ NN$	[0.3]
$NP \longrightarrow JJ\ NNS$	[0.4]
$DT \longrightarrow a$	[0.3]
$NN \longrightarrow pilot$	[0.1]
$VBZ \longrightarrow likes$	[0.4]
$VBG \longrightarrow flying$	[0.5]
$JJ \longrightarrow flying$	[0.1]
$NNS \longrightarrow planes$	[.34]

Probabilistic CKY: all parses

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]	-		
	NN [0.1]	-		
		VBZ [0.4]		

S	$\longrightarrow NP\ VP$	[1.0]
VP	$\longrightarrow VBG\ NNS$	[0.1]
VP	$\longrightarrow VBZ\ VP$	[0.1]
VP	$\longrightarrow VBZ\ NP$	[0.3]
NP	$\longrightarrow DT\ NN$	[0.3]
NP	$\longrightarrow JJ\ NNS$	[0.4]
DT	$\longrightarrow a$	[0.3]
NN	$\longrightarrow pilot$	[0.1]
VBZ	$\longrightarrow likes$	[0.4]
VBG	$\longrightarrow flying$	[0.5]
JJ	$\longrightarrow flying$	[0.1]
NNS	$\longrightarrow planes$	[.34]

Probabilistic CKY: all parses

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]	-		
	NN [0.1]	-		
		VBZ [0.4]		
			JJ [0.1]	

	$S \longrightarrow NP\ VP$	[1.0]
	$VP \longrightarrow VBG\ NNS$	[0.1]
	$VP \longrightarrow VBZ\ VP$	[0.1]
	$VP \longrightarrow VBZ\ NP$	[0.3]
	$NP \longrightarrow DT\ NN$	[0.3]
	$NP \longrightarrow JJ\ NNS$	[0.4]
	$DT \longrightarrow a$	[0.3]
	$NN \longrightarrow pilot$	[0.1]
	$VBZ \longrightarrow likes$	[0.4]
	$VBG \longrightarrow flying$	[0.5]
#	$JJ \longrightarrow flying$	[0.1]
	$NNS \longrightarrow planes$	[.34]

Probabilistic CKY: all parses

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]	-		
	NN [0.1]	-		
		VBZ [0.4]		
			JJ [0.1] VBG [0.5]	

	$S \longrightarrow NP\ VP$		[1.0]
	$VP \longrightarrow VBG\ NNS$		[0.1]
	$VP \longrightarrow VBZ\ VP$		[0.1]
	$VP \longrightarrow VBZ\ NP$		[0.3]
	$NP \longrightarrow DT\ NN$		[0.3]
	$NP \longrightarrow JJ\ NNS$		[0.4]
	$DT \longrightarrow a$		[0.3]
	$NN \longrightarrow pilot$		[0.1]
	$VBZ \longrightarrow likes$		[0.4]
#	$VBG \longrightarrow flying$		[0.5]
	$JJ \longrightarrow flying$		[0.1]
	$NNS \longrightarrow planes$		[.34]

Probabilistic CKY: all parses

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]	-		
	NN [0.1]	-		
		VBZ [0.4]	-	
			JJ [0.1] VBG [0.5]	

$S \longrightarrow NP\ VP$	[1.0]
$VP \longrightarrow VBG\ NNS$	[0.1]
$VP \longrightarrow VBZ\ VP$	[0.1]
$VP \longrightarrow VBZ\ NP$	[0.3]
$NP \longrightarrow DT\ NN$	[0.3]
$NP \longrightarrow JJ\ NNS$	[0.4]
$DT \longrightarrow a$	[0.3]
$NN \longrightarrow pilot$	[0.1]
$VBZ \longrightarrow likes$	[0.4]
$VBG \longrightarrow flying$	[0.5]
$JJ \longrightarrow flying$	[0.1]
$NNS \longrightarrow planes$	[.34]

Probabilistic CKY: all parses

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]	-		
	NN [0.1]	-	-	
		VBZ [0.4]	-	
			JJ [0.1] VBG [0.5]	

$S \longrightarrow NP\ VP$		[1.0]
$VP \longrightarrow VBG\ NNS$		[0.1]
$VP \longrightarrow VBZ\ VP$		[0.1]
$VP \longrightarrow VBZ\ NP$		[0.3]
$NP \longrightarrow DT\ NN$		[0.3]
$NP \longrightarrow JJ\ NNS$		[0.4]
$DT \longrightarrow a$		[0.3]
$NN \longrightarrow pilot$		[0.1]
$VBZ \longrightarrow likes$		[0.4]
$VBG \longrightarrow flying$		[0.5]
$JJ \longrightarrow flying$		[0.1]
$NNS \longrightarrow planes$		[.34]

Probabilistic CKY: all parses

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]	-	-	
	NN [0.1]	-	-	
		VBZ [0.4]	-	
			JJ [0.1] VBG [0.5]	

S → *NP VP*	[1.0]
VP → *VBG NNS*	[0.1]
VP → *VBZ VP*	[0.1]
VP → *VBZ NP*	[0.3]
NP → *DT NN*	[0.3]
NP → *JJ NNS*	[0.4]
DT → *a*	[0.3]
NN → *pilot*	[0.1]
VBZ → *likes*	[0.4]
VBG → *flying*	[0.5]
JJ → *flying*	[0.1]
NNS → *planes*	[.34]

Probabilistic CKY: all parses

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]	-	-	
	NN [0.1]	-	-	
		VBZ [0.4]	-	
			JJ [0.1] VBG [0.5]	
				NNS [.34]

	$S \longrightarrow NP\ VP$	[1.0]
	$VP \longrightarrow VBG\ NNS$	[0.1]
	$VP \longrightarrow VBZ\ VP$	[0.1]
	$VP \longrightarrow VBZ\ NP$	[0.3]
	$NP \longrightarrow DT\ NN$	[0.3]
	$NP \longrightarrow JJ\ NNS$	[0.4]
	$DT \longrightarrow a$	[0.3]
	$NN \longrightarrow pilot$	[0.1]
	$VBZ \longrightarrow likes$	[0.4]
	$VBG \longrightarrow flying$	[0.5]
	$JJ \longrightarrow flying$	[0.1]
#	$NNS \longrightarrow planes$	[.34]

Probabilistic CKY: all parses

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]	-	-	
	NN [0.1]	-	-	
		VBZ [0.4]	-	
			JJ [0.1] VBG [0.5]	NP [.0136]
				NNS [.34]

	$S \longrightarrow NP\ VP$	[1.0]
	$VP \longrightarrow VBG\ NNS$	[0.1]
	$VP \longrightarrow VBZ\ VP$	[0.1]
	$VP \longrightarrow VBZ\ NP$	[0.3]
	$NP \longrightarrow DT\ NN$	[0.3]
#	$NP \longrightarrow JJ\ NNS$	[0.4]
	$DT \longrightarrow a$	[0.3]
	$NN \longrightarrow pilot$	[0.1]
	$VBZ \longrightarrow likes$	[0.4]
	$VBG \longrightarrow flying$	[0.5]
	$JJ \longrightarrow flying$	[0.1]
	$NNS \longrightarrow planes$	[.34]

$0.1 \times 0.34 \times 0.4 =$
0.0136

Probabilistic CKY: all parses

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]	-	-	
	NN [0.1]	-	-	
		VBZ [0.4]	-	
			JJ [0.1] VBG [0.5]	NP [.0136] VP [.017]
				NNS [.34]

	$S \longrightarrow NP\ VP$	[1.0]
#	$VP \longrightarrow VBG\ NNS$	[0.1]
	$VP \longrightarrow VBZ\ VP$	[0.1]
	$VP \longrightarrow VBZ\ NP$	[0.3]
	$NP \longrightarrow DT\ NN$	[0.3]
	$NP \longrightarrow JJ\ NNS$	[0.4]
	$DT \longrightarrow a$	[0.3]
	$NN \longrightarrow pilot$	[0.1]
	$VBZ \longrightarrow likes$	[0.4]
	$VBG \longrightarrow flying$	[0.5]
	$JJ \longrightarrow flying$	[0.1]
	$NNS \longrightarrow planes$	[.34]

$0.5 \times 0.34 \times 0.1 = 0.017$

Probabilistic CKY: all parses

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]	-	-	
	NN [0.1]	-	-	
		VBZ [0.4]	-	VP [.001632]
			JJ [0.1] VBG [0.5]	NP [.0136] VP [.017]
				NNS [.34]

	$S \longrightarrow NP\ VP$	[1.0]
	$VP \longrightarrow VBG\ NNS$	[0.1]
	$VP \longrightarrow VBZ\ VP$	[0.1]
#	$VP \longrightarrow VBZ\ NP$	[0.3]
	$NP \longrightarrow DT\ NN$	[0.3]
	$NP \longrightarrow JJ\ NNS$	[0.4]
	$DT \longrightarrow a$	[0.3]
	$NN \longrightarrow pilot$	[0.1]
	$VBZ \longrightarrow likes$	[0.4]
	$VBG \longrightarrow flying$	[0.5]
	$JJ \longrightarrow flying$	[0.1]
	$NNS \longrightarrow planes$	[.34]

$0.4 \times 0.0136 \times 0.3 =$
0.001632

Probabilistic CKY: all parses

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]	-	-	
	NN [0.1]	-	-	
		VBZ [0.4]	-	VP [.001632] VP [.00068]
			JJ [0.1] VBG [0.5]	NP [.0136] VP [.017]
				NNS [.34]

	Rule	Prob.
	$S \longrightarrow NP\ VP$	[1.0]
	$VP \longrightarrow VBG\ NNS$	[0.1]
#	$VP \longrightarrow VBZ\ VP$	[0.1]
	$VP \longrightarrow VBZ\ NP$	[0.3]
	$NP \longrightarrow DT\ NN$	[0.3]
	$NP \longrightarrow JJ\ NNS$	[0.4]
	$DT \longrightarrow a$	[0.3]
	$NN \longrightarrow pilot$	[0.1]
	$VBZ \longrightarrow likes$	[0.4]
	$VBG \longrightarrow flying$	[0.5]
	$JJ \longrightarrow flying$	[0.1]
	$NNS \longrightarrow planes$	[.34]

$0.4 \times 0.017 \times 0.1 =$
0.00068

Probabilistic CKY: all parses

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]	-	-	
	NN [0.1]	-	-	-
		VBZ [0.4]	-	VP [.001632] VP [.00068]
			JJ [0.1] VBG [0.5]	NP [.0136] VP [.017]
				NNS [.34]

$S \longrightarrow NP\ VP$	[1.0]
$VP \longrightarrow VBG\ NNS$	[0.1]
$VP \longrightarrow VBZ\ VP$	[0.1]
$VP \longrightarrow VBZ\ NP$	[0.3]
$NP \longrightarrow DT\ NN$	[0.3]
$NP \longrightarrow JJ\ NNS$	[0.4]
$DT \longrightarrow a$	[0.3]
$NN \longrightarrow pilot$	[0.1]
$VBZ \longrightarrow likes$	[0.4]
$VBG \longrightarrow flying$	[0.5]
$JJ \longrightarrow flying$	[0.1]
$NNS \longrightarrow planes$	[.34]

Probabilistic CKY: all parses

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]	-	-	S [$1.4688x10^{-5}$]
	NN [0.1]	-	-	-
		VBZ [0.4]	-	VP [.001632] VP [.00068]
			JJ [0.1] VBG [0.5]	NP [.0136] VP [.017]
				NNS [.34]

#	Rule	Prob
	$S \longrightarrow NP\ VP$	[1.0]
	$VP \longrightarrow VBG\ NNS$	[0.1]
	$VP \longrightarrow VBZ\ VP$	[0.1]
	$VP \longrightarrow VBZ\ NP$	[0.3]
	$NP \longrightarrow DT\ NN$	[0.3]
	$NP \longrightarrow JJ\ NNS$	[0.4]
	$DT \longrightarrow a$	[0.3]
	$NN \longrightarrow pilot$	[0.1]
	$VBZ \longrightarrow likes$	[0.4]
	$VBG \longrightarrow flying$	[0.5]
	$JJ \longrightarrow flying$	[0.1]
	$NNS \longrightarrow planes$	[.34]

$0.009 \times 0.001632 \times 1.0 = 1.4688x10^{-5}$

Probabilistic CKY: all parses

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]	-	-	S [$1.4688x10^{-5}$] S [$6.12x10^{-6}$]
	NN [0.1]	-	-	-
		VBZ [0.4]	-	VP [.001632] VP [.00068]
			JJ [0.1] VBG [0.5]	NP [.0136] VP [.017]
				NNS [.34]

$S \longrightarrow NP\ VP$	[1.0]
$VP \longrightarrow VBG\ NNS$	[0.1]
$VP \longrightarrow VBZ\ VP$	[0.1]
$VP \longrightarrow VBZ\ NP$	[0.3]
$NP \longrightarrow DT\ NN$	[0.3]
$NP \longrightarrow JJ\ NNS$	[0.4]
$DT \longrightarrow a$	[0.3]
$NN \longrightarrow pilot$	[0.1]
$VBZ \longrightarrow likes$	[0.4]
$VBG \longrightarrow flying$	[0.5]
$JJ \longrightarrow flying$	[0.1]
$NNS \longrightarrow planes$	[.34]

$0.009 \times 0.00068 \times 1.0 = 6.12x10^{-6}$

Probabilistic CKY: best parse

a 1	pilot 2	likes 3	flying 4	planes 5

$S \longrightarrow NP\ VP$		[1.0]
$VP \longrightarrow VBG\ NNS$		[0.1]
$VP \longrightarrow VBZ\ VP$		[0.1]
$VP \longrightarrow VBZ\ NP$		[0.3]
$NP \longrightarrow DT\ NN$		[0.3]
$NP \longrightarrow JJ\ NNS$		[0.4]
$DT \longrightarrow a$		[0.3]
$NN \longrightarrow pilot$		[0.1]
$VBZ \longrightarrow likes$		[0.4]
$VBG \longrightarrow flying$		[0.5]
$JJ \longrightarrow flying$		[0.1]
$NNS \longrightarrow planes$		[.34]

Probabilistic CKY: best parse

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]				

	$S \longrightarrow NP\ VP$	[1.0]
	$VP \longrightarrow VBG\ NNS$	[0.1]
	$VP \longrightarrow VBZ\ VP$	[0.1]
	$VP \longrightarrow VBZ\ NP$	[0.3]
	$NP \longrightarrow DT\ NN$	[0.3]
	$NP \longrightarrow JJ\ NNS$	[0.4]
#	$DT \longrightarrow a$	[0.3]
	$NN \longrightarrow pilot$	[0.1]
	$VBZ \longrightarrow likes$	[0.4]
	$VBG \longrightarrow flying$	[0.5]
	$JJ \longrightarrow flying$	[0.1]
	$NNS \longrightarrow planes$	[.34]

Probabilistic CKY: best parse

a 1	**pilot** 2	likes 3	flying 4	planes 5
DT [0.3]				
	NN [0.1]			

	$S \longrightarrow NP\ VP$	[1.0]
	$VP \longrightarrow VBG\ NNS$	[0.1]
	$VP \longrightarrow VBZ\ VP$	[0.1]
	$VP \longrightarrow VBZ\ NP$	[0.3]
	$NP \longrightarrow DT\ NN$	[0.3]
	$NP \longrightarrow JJ\ NNS$	[0.4]
	$DT \longrightarrow a$	[0.3]
#	$NN \longrightarrow pilot$	[0.1]
	$VBZ \longrightarrow likes$	[0.4]
	$VBG \longrightarrow flying$	[0.5]
	$JJ \longrightarrow flying$	[0.1]
	$NNS \longrightarrow planes$	[.34]

Probabilistic CKY: best parse

a 1	**pilot** 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]			
	NN [0.1]			

	S → *NP VP*	[1.0]
	VP → *VBG NNS*	[0.1]
	VP → *VBZ VP*	[0.1]
	VP → *VBZ NP*	[0.3]
#	*NP* → *DT NN*	[0.3]
	NP → *JJ NNS*	[0.4]
	DT → *a*	[0.3]
	NN → *pilot*	[0.1]
	VBZ → *likes*	[0.4]
	VBG → *flying*	[0.5]
	JJ → *flying*	[0.1]
	NNS → *planes*	[.34]

$0.3 \times 0.1 \times 0.3 =$
0.009

Probabilistic CKY: best parse

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]			
	NN [0.1]			
		VBZ [0.4]		

#

$S \longrightarrow NP\ VP$	[1.0]
$VP \longrightarrow VBG\ NNS$	[0.1]
$VP \longrightarrow VBZ\ VP$	[0.1]
$VP \longrightarrow VBZ\ NP$	[0.3]
$NP \longrightarrow DT\ NN$	[0.3]
$NP \longrightarrow JJ\ NNS$	[0.4]
$DT \longrightarrow a$	[0.3]
$NN \longrightarrow pilot$	[0.1]
$VBZ \longrightarrow likes$	[0.4]
$VBG \longrightarrow flying$	[0.5]
$JJ \longrightarrow flying$	[0.1]
$NNS \longrightarrow planes$	[.34]

Probabilistic CKY: best parse

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]			
	NN [0.1]	-		
		VBZ [0.4]		

S	$\longrightarrow$ NP VP	[1.0]
VP	$\longrightarrow$ VBG NNS	[0.1]
VP	$\longrightarrow$ VBZ VP	[0.1]
VP	$\longrightarrow$ VBZ NP	[0.3]
NP	$\longrightarrow$ DT NN	[0.3]
NP	$\longrightarrow$ JJ NNS	[0.4]
DT	$\longrightarrow$ a	[0.3]
NN	$\longrightarrow$ $pilot$	[0.1]
VBZ	$\longrightarrow$ $likes$	[0.4]
VBG	$\longrightarrow$ $flying$	[0.5]
JJ	$\longrightarrow$ $flying$	[0.1]
NNS	$\longrightarrow$ $planes$	[.34]

Probabilistic CKY: best parse

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]	-		
	NN [0.1]	-		
		VBZ [0.4]		

$S \longrightarrow NP\ VP$	[1.0]
$VP \longrightarrow VBG\ NNS$	[0.1]
$VP \longrightarrow VBZ\ VP$	[0.1]
$VP \longrightarrow VBZ\ NP$	[0.3]
$NP \longrightarrow DT\ NN$	[0.3]
$NP \longrightarrow JJ\ NNS$	[0.4]
$DT \longrightarrow a$	[0.3]
$NN \longrightarrow pilot$	[0.1]
$VBZ \longrightarrow likes$	[0.4]
$VBG \longrightarrow flying$	[0.5]
$JJ \longrightarrow flying$	[0.1]
$NNS \longrightarrow planes$	[.34]

Probabilistic CKY: best parse

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]	-		
	NN [0.1]	-		
		VBZ [0.4]		
			JJ [0.1]	

	$S \longrightarrow NP\ VP$	[1.0]
	$VP \longrightarrow VBG\ NNS$	[0.1]
	$VP \longrightarrow VBZ\ VP$	[0.1]
	$VP \longrightarrow VBZ\ NP$	[0.3]
	$NP \longrightarrow DT\ NN$	[0.3]
	$NP \longrightarrow JJ\ NNS$	[0.4]
	$DT \longrightarrow a$	[0.3]
	$NN \longrightarrow pilot$	[0.1]
	$VBZ \longrightarrow likes$	[0.4]
	$VBG \longrightarrow flying$	[0.5]
#	$JJ \longrightarrow flying$	[0.1]
	$NNS \longrightarrow planes$	[.34]

Probabilistic CKY: best parse

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]	-		
	NN [0.1]	-		
		VBZ [0.4]		
			JJ [0.1] VBG [0.5]	

	$S \longrightarrow NP\ VP$	[1.0]
	$VP \longrightarrow VBG\ NNS$	[0.1]
	$VP \longrightarrow VBZ\ VP$	[0.1]
	$VP \longrightarrow VBZ\ NP$	[0.3]
	$NP \longrightarrow DT\ NN$	[0.3]
	$NP \longrightarrow JJ\ NNS$	[0.4]
	$DT \longrightarrow a$	[0.3]
	$NN \longrightarrow pilot$	[0.1]
	$VBZ \longrightarrow likes$	[0.4]
#	$VBG \longrightarrow flying$	[0.5]
	$JJ \longrightarrow flying$	[0.1]
	$NNS \longrightarrow planes$	[.34]

Probabilistic CKY: best parse

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]	-		
	NN [0.1]	-		
		VBZ [0.4]	-	
			JJ [0.1] VBG [0.5]	

$S \longrightarrow NP\ VP$	[1.0]
$VP \longrightarrow VBG\ NNS$	[0.1]
$VP \longrightarrow VBZ\ VP$	[0.1]
$VP \longrightarrow VBZ\ NP$	[0.3]
$NP \longrightarrow DT\ NN$	[0.3]
$NP \longrightarrow JJ\ NNS$	[0.4]
$DT \longrightarrow a$	[0.3]
$NN \longrightarrow pilot$	[0.1]
$VBZ \longrightarrow likes$	[0.4]
$VBG \longrightarrow flying$	[0.5]
$JJ \longrightarrow flying$	[0.1]
$NNS \longrightarrow planes$	[.34]

Probabilistic CKY: best parse

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]	-		
	NN [0.1]	-	-	
		VBZ [0.4]	-	
			JJ [0.1] VBG [0.5]	

$S \longrightarrow NP\ VP$	[1.0]
$VP \longrightarrow VBG\ NNS$	[0.1]
$VP \longrightarrow VBZ\ VP$	[0.1]
$VP \longrightarrow VBZ\ NP$	[0.3]
$NP \longrightarrow DT\ NN$	[0.3]
$NP \longrightarrow JJ\ NNS$	[0.4]
$DT \longrightarrow a$	[0.3]
$NN \longrightarrow pilot$	[0.1]
$VBZ \longrightarrow likes$	[0.4]
$VBG \longrightarrow flying$	[0.5]
$JJ \longrightarrow flying$	[0.1]
$NNS \longrightarrow planes$	[.34]

Probabilistic CKY: best parse

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]	-	-	
	NN [0.1]	-	-	
		VBZ [0.4]	-	
			JJ [0.1] VBG [0.5]	

$S \longrightarrow NP\ VP$	[1.0]
$VP \longrightarrow VBG\ NNS$	[0.1]
$VP \longrightarrow VBZ\ VP$	[0.1]
$VP \longrightarrow VBZ\ NP$	[0.3]
$NP \longrightarrow DT\ NN$	[0.3]
$NP \longrightarrow JJ\ NNS$	[0.4]
$DT \longrightarrow a$	[0.3]
$NN \longrightarrow pilot$	[0.1]
$VBZ \longrightarrow likes$	[0.4]
$VBG \longrightarrow flying$	[0.5]
$JJ \longrightarrow flying$	[0.1]
$NNS \longrightarrow planes$	[.34]

Probabilistic CKY: best parse

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]	-	-	
	NN [0.1]	-	-	
		VBZ [0.4]	-	
			JJ [0.1] VBG [0.5]	
				NNS [.34]

	$S \longrightarrow NP\ VP$	[1.0]
	$VP \longrightarrow VBG\ NNS$	[0.1]
	$VP \longrightarrow VBZ\ VP$	[0.1]
	$VP \longrightarrow VBZ\ NP$	[0.3]
	$NP \longrightarrow DT\ NN$	[0.3]
	$NP \longrightarrow JJ\ NNS$	[0.4]
	$DT \longrightarrow a$	[0.3]
	$NN \longrightarrow pilot$	[0.1]
	$VBZ \longrightarrow likes$	[0.4]
	$VBG \longrightarrow flying$	[0.5]
	$JJ \longrightarrow flying$	[0.1]
#	$NNS \longrightarrow planes$	[.34]

Probabilistic CKY: best parse

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]	-	-	
	NN [0.1]	-	-	
		VBZ [0.4]	-	
			JJ [0.1] VBG [0.5]	NP [.0136]
				NNS [.34]

	$S \longrightarrow NP\ VP$	[1.0]
	$VP \longrightarrow VBG\ NNS$	[0.1]
	$VP \longrightarrow VBZ\ VP$	[0.1]
	$VP \longrightarrow VBZ\ NP$	[0.3]
	$NP \longrightarrow DT\ NN$	[0.3]
#	$NP \longrightarrow JJ\ NNS$	[0.4]
	$DT \longrightarrow a$	[0.3]
	$NN \longrightarrow pilot$	[0.1]
	$VBZ \longrightarrow likes$	[0.4]
	$VBG \longrightarrow flying$	[0.5]
	$JJ \longrightarrow flying$	[0.1]
	$NNS \longrightarrow planes$	[.34]

$0.1 \times 0.34 \times 0.4 =$
0.0136

Probabilistic CKY: best parse

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]	-	-	
	NN [0.1]	-	-	
		VBZ [0.4]	-	
			JJ [0.1] VBG [0.5]	NP [.0136] VP [.017]
				NNS [.34]

	Rule	Prob.
	$S \longrightarrow NP\ VP$	[1.0]
#	$VP \longrightarrow VBG\ NNS$	[0.1]
	$VP \longrightarrow VBZ\ VP$	[0.1]
	$VP \longrightarrow VBZ\ NP$	[0.3]
	$NP \longrightarrow DT\ NN$	[0.3]
	$NP \longrightarrow JJ\ NNS$	[0.4]
	$DT \longrightarrow a$	[0.3]
	$NN \longrightarrow pilot$	[0.1]
	$VBZ \longrightarrow likes$	[0.4]
	$VBG \longrightarrow flying$	[0.5]
	$JJ \longrightarrow flying$	[0.1]
	$NNS \longrightarrow planes$	[.34]

$0.5 \times 0.34 \times 0.1 =$
0.017

Probabilistic CKY: best parse

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]	-	-	
	NN [0.1]	-	-	
		VBZ [0.4]	-	VP [.001632]
			JJ [0.1] VBG [0.5]	NP [.0136] VP [.017]
				NNS [.34]

	Rule	Prob
	$S \longrightarrow NP\ VP$	[1.0]
	$VP \longrightarrow VBG\ NNS$	[0.1]
	$VP \longrightarrow VBZ\ VP$	[0.1]
#	$VP \longrightarrow VBZ\ NP$	[0.3]
	$NP \longrightarrow DT\ NN$	[0.3]
	$NP \longrightarrow JJ\ NNS$	[0.4]
	$DT \longrightarrow a$	[0.3]
	$NN \longrightarrow pilot$	[0.1]
	$VBZ \longrightarrow likes$	[0.4]
	$VBG \longrightarrow flying$	[0.5]
	$JJ \longrightarrow flying$	[0.1]
	$NNS \longrightarrow planes$	[.34]

$0.4 \times 0.0136 \times 0.3 = 0.001632$

Probabilistic CKY: best parse

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]	-	-	
	NN [0.1]	-	-	
		VBZ [0.4]	-	VP [.001632]
			JJ [0.1] VBG [0.5]	NP [.0136] VP [.017]
				NNS [.34]

	$S \longrightarrow NP\ VP$	[1.0]
	$VP \longrightarrow VBG\ NNS$	[0.1]
#	$VP \longrightarrow VBZ\ VP$	[0.1]
	$VP \longrightarrow VBZ\ NP$	[0.3]
	$NP \longrightarrow DT\ NN$	[0.3]
	$NP \longrightarrow JJ\ NNS$	[0.4]
	$DT \longrightarrow a$	[0.3]
	$NN \longrightarrow pilot$	[0.1]
	$VBZ \longrightarrow likes$	[0.4]
	$VBG \longrightarrow flying$	[0.5]
	$JJ \longrightarrow flying$	[0.1]
	$NNS \longrightarrow planes$	[.34]

Move second VP to chart?
0.4 x 0.017 x 0.1 = 0.00068

Probabilistic CKY: best parse

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]	-	-	
	NN [0.1]	-	-	
		VBZ [0.4]	-	VP [.001632]
			JJ [0.1] VBG [0.5]	NP [.0136] VP [.017]
				NNS [.34]

$S \longrightarrow NP\ VP$		[1.0]
$VP \longrightarrow VBG\ NNS$		[0.1]
$VP \longrightarrow VBZ\ VP$		[0.1]
$VP \longrightarrow VBZ\ NP$		[0.3]
$NP \longrightarrow DT\ NN$		[0.3]
$NP \longrightarrow JJ\ NNS$		[0.4]
$DT \longrightarrow a$		[0.3]
$NN \longrightarrow pilot$		[0.1]
$VBZ \longrightarrow likes$		[0.4]
$VBG \longrightarrow flying$		[0.5]
$JJ \longrightarrow flying$		[0.1]
$NNS \longrightarrow planes$		[.34]

Is 0.001632
less than 0.00068 ?

Probabilistic CKY: best parse

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]	-	-	
	NN [0.1]	-	-	-
		VBZ [0.4]	-	VP [.001632]
			JJ [0.1] VBG [0.5]	NP [.0136] VP [.017]
				NNS [.34]

S	→ *NP VP*	[1.0]
VP	→ *VBG NNS*	[0.1]
VP	→ *VBZ VP*	[0.1]
VP	→ *VBZ NP*	[0.3]
NP	→ *DT NN*	[0.3]
NP	→ *JJ NNS*	[0.4]
DT	→ *a*	[0.3]
NN	→ *pilot*	[0.1]
VBZ	→ *likes*	[0.4]
VBG	→ *flying*	[0.5]
JJ	→ *flying*	[0.1]
NNS	→ *planes*	[.34]

Probabilistic CKY: best parse

a 1	pilot 2	likes 3	flying 4	planes 5
DT [0.3]	NP [.009]	-	-	S [$1.4688x10^{-5}$]
	NN [0.1]	-	-	-
		VBZ [0.4]	-	VP [.001632]
			JJ [0.1] VBG [0.5]	NP [.0136] VP [.017]
				NNS [.34]

#		
	S → *NP VP*	[1.0]
	VP → *VBG NNS*	[0.1]
	VP → *VBZ VP*	[0.1]
	VP → *VBZ NP*	[0.3]
	NP → *DT NN*	[0.3]
	NP → *JJ NNS*	[0.4]
	DT → *a*	[0.3]
	NN → *pilot*	[0.1]
	VBZ → *likes*	[0.4]
	VBG → *flying*	[0.5]
	JJ → *flying*	[0.1]
	NNS → *planes*	[.34]

$0.009 \times 0.001632 \times 1.0 = 1.4688x10^{-5}$

Today's lecture

1. Uncertainty
2. Probabilistic Context-Free Grammars
   - Uses of PCFGs: Disambiguation
   - Benefits of PCFGs
   - Issues
3. Probabilistic parsing
   - Probabilistic CKY
4. PCKY Example
5. Homework 3

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Homework 3

1. work with a real PCFG
2. build a probabistic parser (CKY)
3. evaluate the results

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Homework 3

1. work with a real PCFG
2. build a probabistic parser (CKY)
3. evaluate the results

$$P(\textit{Hw3 is easy.}) = 0.0000001$$
$$P(\textit{Hw3 is hard.}) = 0.004$$

Parsing: dev/train/test paradigm

The Wall Street Journal (WSJ) section of the Penn Treebank (PTB), for all its faults, provides a very useful resource for comparing parser performance.

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Parsing: dev/train/test paradigm

The Wall Street Journal (WSJ) section of the Penn Treebank (PTB), for all its faults, provides a very useful resource for comparing parser performance.

In building a probabilistic parser, there are four kinds of resources that are commonly used esp. in the ACL related literature:

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Parsing: dev/train/test paradigm

The Wall Street Journal (WSJ) section of the Penn Treebank (PTB), for all its faults, provides a very useful resource for comparing parser performance.

In building a probabilistic parser, there are four kinds of resources that are commonly used esp. in the ACL related literature:

1. **training data**: large number of annotated sentences (sec. 2–21 of PTB has 39,830 sentences)

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Parsing: dev/train/test paradigm

The Wall Street Journal (WSJ) section of the Penn Treebank (PTB), for all its faults, provides a very useful resource for comparing parser performance.

In building a probabilistic parser, there are four kinds of resources that are commonly used esp. in the ACL related literature:

1. **training data**: large number of annotated sentences (sec. 2–21 of PTB has 39,830 sentences)
2. **development data**: small number of annotated sentences used to "tweak" parser (sec. 22, of PTB)

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed

Parsing: dev/train/test paradigm

The Wall Street Journal (WSJ) section of the Penn Treebank (PTB), for all its faults, provides a very useful resource for comparing parser performance.

In building a probabilistic parser, there are four kinds of resources that are commonly used esp. in the ACL related literature:

1. **training data**: large number of annotated sentences (sec. 2–21 of PTB has 39,830 sentences)
2. **development data**: small number of annotated sentences used to "tweak" parser (sec. 22, of PTB)
3. **test data**: small-medium number of un-annotated sentences used as input to parser (sec. 23 of PTB has 2416 sentences, $\sim$ 6% of training set)

Parsing: dev/train/test paradigm

The Wall Street Journal (WSJ) section of the Penn Treebank (PTB), for all its faults, provides a very useful resource for comparing parser performance.

In building a probabilistic parser, there are four kinds of resources that are commonly used esp. in the ACL related literature:

1. **training data**: large number of annotated sentences (sec. 2–21 of PTB has 39,830 sentences)
2. **development data**: small number of annotated sentences used to "tweak" parser (sec. 22, of PTB)
3. **test data**: small-medium number of un-annotated sentences used as input to parser (sec. 23 of PTB has 2416 sentences, $\sim 6\%$ of training set)
4. **gold standard**: annotated version of test data, with no errors (hidden till parser is developed)

Scott Farrar
CLMA, University of Washington farrar@u.washington.ed