

SPECIAL BONUS LECTURE!

Context-Free Grammars and Translation of Human Languages

Greg Hanneman

11-453: FLAC

Tuesday, March 20, 2012

**“Automatically Improved Category
Labels for SCFG-Based Statistical
Machine Translation”**

“Automatically Improved Category
Labels for **SCFG**-Based Statistical
Machine Translation”

Context-Free Grammars

$$S \rightarrow A \mid B$$

$$A \rightarrow 0 A 1 \mid \varepsilon$$

$$B \rightarrow 1 B 0 \mid \varepsilon$$

Context-Free Grammars

$$S \rightarrow A \mid B$$

$$A \rightarrow 0 A 1 \mid \varepsilon$$

$$B \rightarrow 1 B 0 \mid \varepsilon$$

- How would we get a CFG for **English**?
 - Start symbol S = “sentence”
 - Nonterminals = English syntactic phrases
 - Terminals = English words

A CFG for English?

$S \rightarrow NP VP \mid AUX NP Vbar$

$NP \rightarrow D Nbar \mid Nbar \mid NP PP \mid NP C NP$

$Nbar \rightarrow ADJ N \mid N$

$VP \rightarrow AUX Vbar \mid Vbar$

$Vbar \rightarrow V \mid V NP \mid V PP \mid V NP PP \mid V NP NP$

$V \rightarrow V P$

$PP \rightarrow P NP \mid PP C PP$

$V \rightarrow \text{is} \mid \text{do} \mid \text{reads} \mid \text{climbing} \mid \text{saw} \mid \dots$

$N \rightarrow \text{book} \mid \text{trees} \mid \text{she} \mid \dots$ $D \rightarrow \text{a} \mid \text{the} \mid \text{some} \mid \dots$

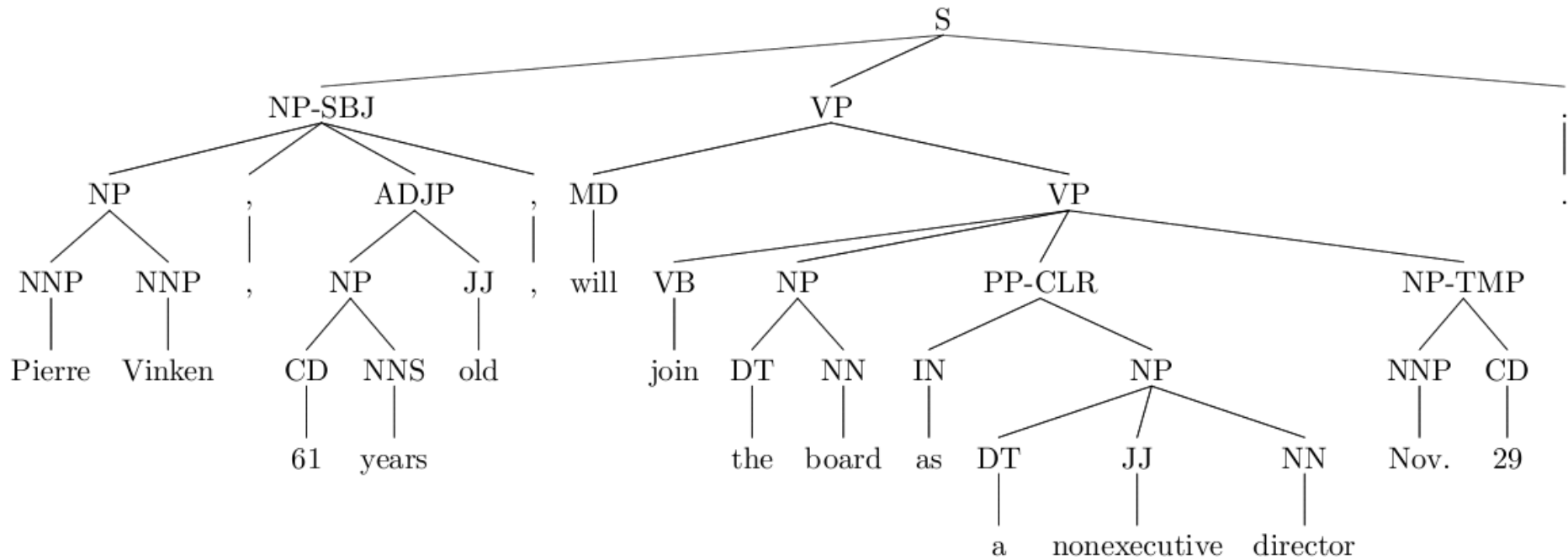
$ADJ \rightarrow \text{bad} \mid \text{good} \mid \dots$ $P \rightarrow \text{to} \mid \text{in} \mid \text{at} \mid \dots$

$AUX \rightarrow \text{were} \mid \text{do} \mid \dots$ $C \rightarrow \text{and} \mid \text{or} \mid \dots$

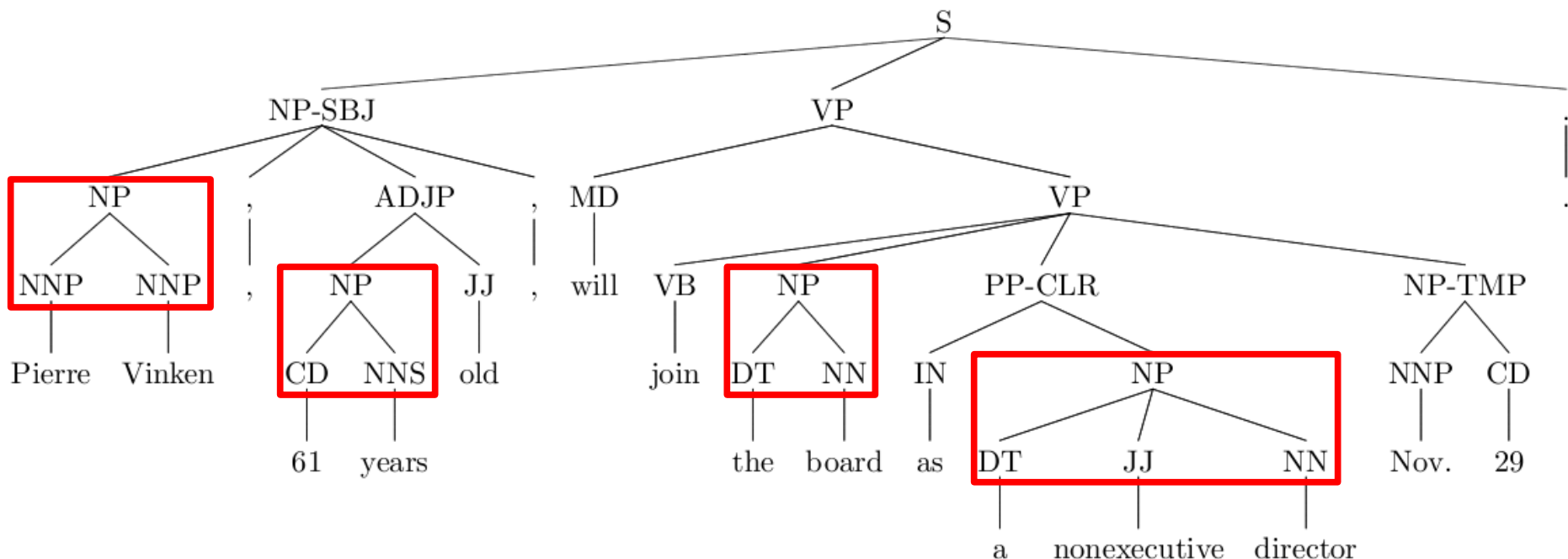
A CFG for English?

- Two key principles in modern natural-language processing:
 - Collect **training data** (examples) for the problem you want to solve
 - Use **statistics** from the data to learn a model to create new examples

A CFG for English?



A CFG for English?



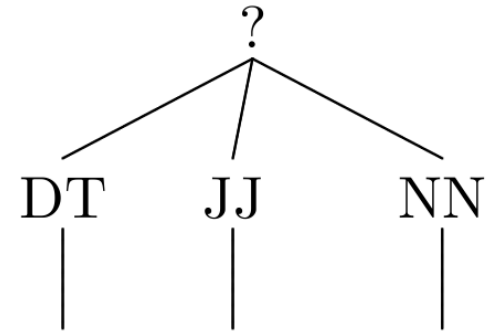
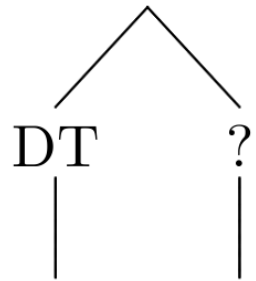
count(NP → NNP NNP)++;

count(NP → CD NNS)++;

count(NP → DT NN)++;

count(NP → DT JJ NN)++;

A CFG for English?



The board elected Mr. Vinken after a short debate .

$P(\text{NN} \mid \text{board}) ?$

$P(\text{VB} \mid \text{board}) ?$

$P(\text{NN} \rightarrow \text{board} \mid \text{DT} \rightarrow \text{The}) ?$

$P(\text{NP} \mid \text{DT JJ NN}) ?$

$P(\text{DT JJ NN} \mid \text{NP}) ?$

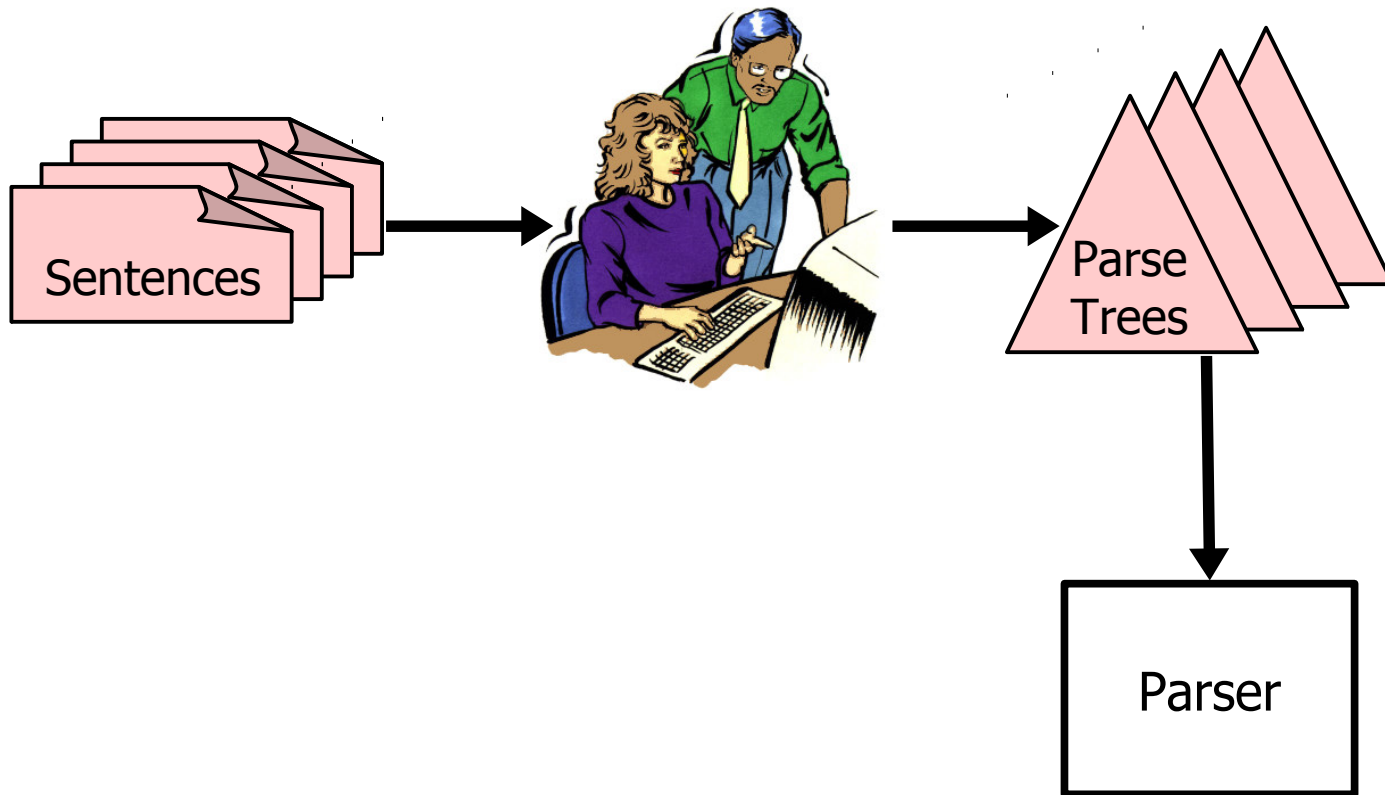
A CFG for English?

- Result: a **statistical parser** for English



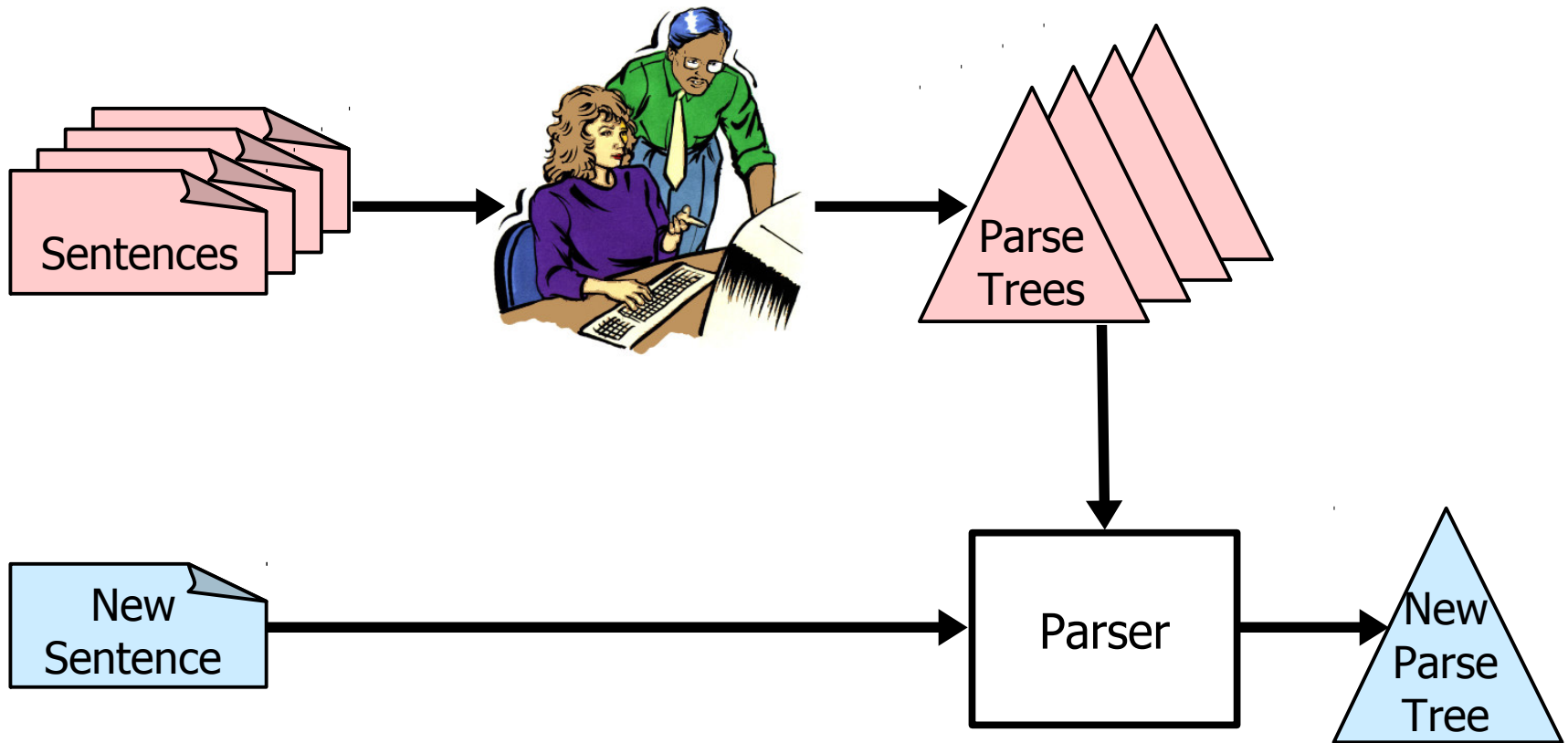
A CFG for English?

- Result: a **statistical parser** for English



A CFG for English?

- Result: a **statistical parser** for English



“Automatically Improved Category
Labels for SCFG-Based Statistical
Machine Translation”

How Can Computers Translate?

Remember:

Collect **training data** and use **statistics**

How Can Computers Translate?

85	冬瓜茶	Wintermelon Tea
86	礦泉水	Mineral Water
87	皇帽啤酒	Carlsberg Beer
87B	老虎啤酒	Tiger Beer
88	黑啤	Guinness Stout
88B	青島啤酒	TSINGTAO Beer
89	中國茶	Chinese Tea
90	紅酒	Red Wine
103	菊花茶	Chrysanthemum Tea

How Can Computers Translate?

85	冬瓜茶	Wintermelon Tea
86	礦泉水	Mineral Water
87	皇帽啤酒	Carlsberg Beer
87B	老虎啤酒	Tiger Beer
88	黑啤	Guinness Stout
88B	青島啤酒	TSINGTAO Beer
89	中國茶	Chinese Tea
90	紅酒	Red Wine
103	菊花茶	Chrysanthemum Tea

茶

tea

$$P(\text{茶} \mid \text{tea}) = 1$$

$$P(\text{tea} \mid \text{茶}) = 1$$

How Can Computers Translate?

85	冬瓜茶	Wintermelon Tea
86	礦泉水	Mineral Water
87	皇帽啤酒	Carlsberg Beer
87B	老虎啤酒	Tiger Beer
88	黑啤	Guinness Stout
88B	青島啤酒	TSINGTAO Beer
89	中國茶	Chinese Tea
90	紅酒	Red Wine
103	菊花茶	Chrysanthemum Tea

茶

tea

酒

beer

酒

wine

$$P(\text{酒} \mid \text{beer}) = 1$$

$$P(\text{beer} \mid \text{酒}) = 0.75$$

$$P(\text{酒} \mid \text{wine}) = 1$$

$$P(\text{wine} \mid \text{酒}) = 0.25$$

How Can Computers Translate?

85	冬瓜茶	Wintermelon Tea
86	礦泉水	Mineral Water
87	皇帽啤酒	Carlsberg Beer
87B	老虎啤酒	Tiger Beer
88	黑啤	Guinness Stout
88B	青島啤酒	TSINGTAO Beer
89	中國茶	Chinese Tea
90	紅酒	Red Wine
103	菊花茶	Chrysanthemum Tea

茶

tea

□酒

beer

酒

wine

$$P(\text{啤酒} \mid \text{beer}) = 1$$

$$P(\text{beer} \mid \text{啤酒}) = 1$$

$$P(\text{酒} \mid \text{wine}) = 1$$

$$P(\text{wine} \mid \text{酒}) = 1$$

How Can Computers Translate?

- This intuition goes far
- Basis of solving **word alignment** problems
 - Guess a model
 - Use model to explain data in most likely way
 - Use explanation to update model
 - Repeat (Expectation-Maximization algorithm)
- With big enough menus (**parallel corpora**), we can learn alignments well

How Can Computers Translate?

	Qu'est-ce qu'une société équitable ?
What is a fair society ?	

[illegible]

How Can Computers Translate?

	Qu'	est	-ce	qu'	une	société	équitable	?
What	■							
is		■						
a					■			
fair							■	
society						■		
?								■

société

society

How Can Computers Translate?

	Qu'	est	-ce	qu'	une	société	équitable	?
What	■							
is		■						
a					■			
fair							■	
society						■		
?								■

société

society

équitable

fair

une

a

How Can Computers Translate?

	Qu'	est	-ce	qu'	une	société	équitable	?
What	■							
is		■						
a					■			
fair							■	
society						■		
?								■

société

society

équitable

fair

une

a

société équitable

fair society

How Can Computers Translate?

	Qu'	est	-ce	qu'	une	société	équitable	?
What	■							
is		■						
a					■			
fair							■	
society						■		
?								■

société

society

équitable

fair

une

a

société équitable

fair society

Qu' est

What is

How Can Computers Translate?

	Qu'	est	-ce	qu'	une	société	équitable	?
What	■							
is		■						
a					■			
fair							■	
society						■		
?								■

société

society

équitable

fair

une

a

société équitable

fair society

Qu' est

What is

Qu' est -ce

What is

How Can Computers Translate?

	Qu'	est	-ce	qu'	une	société	équitable	?
What	■							
is		■						
a					■			
fair							■	
society						■		
?								■

société

society

équitable

fair

une

a

société équitable

fair society

Qu' est

What is

Qu' est -ce

What is

Qu' est -ce que

What is

How Can Computers Translate?

	Qu'	est	-ce	qu'	une	société	équitable	?
What	■							
is		■						
a					■			
fair							■	
society						■		
?								■

X

société	équitable	une
society	fair	a

société équitable	Qu' est
fair society	What is

Qu' est -ce	Qu' est -ce que
What is	What is

How Can Computers Translate?

Qu'	Qu' est -ce qu' une	société	équitable	une
What	What is a	society	fair	a
est	une société équitable	société équitable	Qu' est	
is	a fair society	fair society	What is	
est -ce	est -ce qu'	?	Qu' est -ce	Qu' est -ce qu'
is	is	?	What is	What is

Calculate statistics over each extracted phrase!

How Can Computers Translate?

Le vote final aura lieu demain à 12 heures .

How Can Computers Translate?

Le vote final aura lieu demain à 12 heures .

The	vote	last	have	place	tomorrow	in	12	hours	.
-----	------	------	------	-------	----------	----	----	-------	---

It	final vote	will take place	next day	at	noon	!
----	------------	-----------------	----------	----	------	---

The vote	final	will	spot		to	12 p.m.
	final will have				with 12	o'clock

How Can Computers Translate?

Le	vote	final	aura	lieu	demain	à	12	heures	.
The	vote	last	have	place	tomorrow	in	12	hours	.
It	final vote	will take place	next day	at	noon	!			
The vote	final	will	spot			to	12 p.m.		
	final will have					with 12	o'clock		



$P(\text{final will have} \mid \text{final aura})$

$P(\text{final aura} \mid \text{final will have})$

$P(3 \text{ words} \mid 2 \text{ words}) \dots$

How Can Computers Translate?



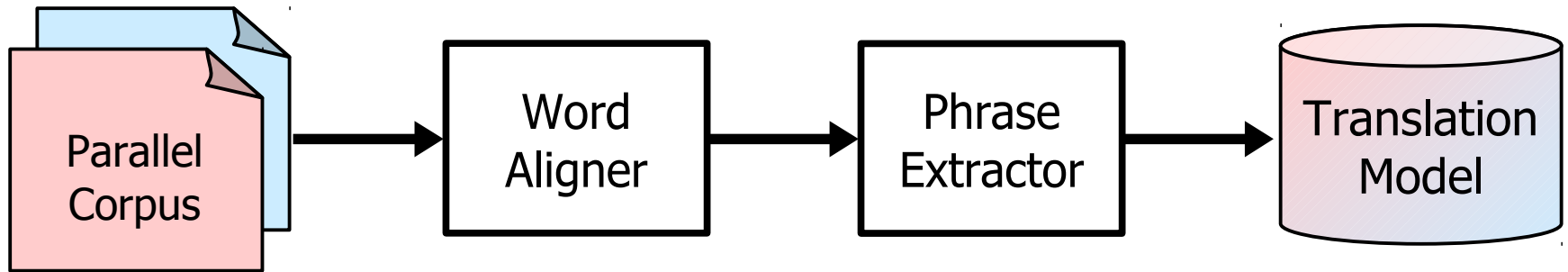
$P(11 \text{ words} \mid 10 \text{ words})$

$P(8 \text{ phrases} \mid 10 \text{ words})$

$P(\text{The final vote will take place tomorrow at 12 o'clock .})$

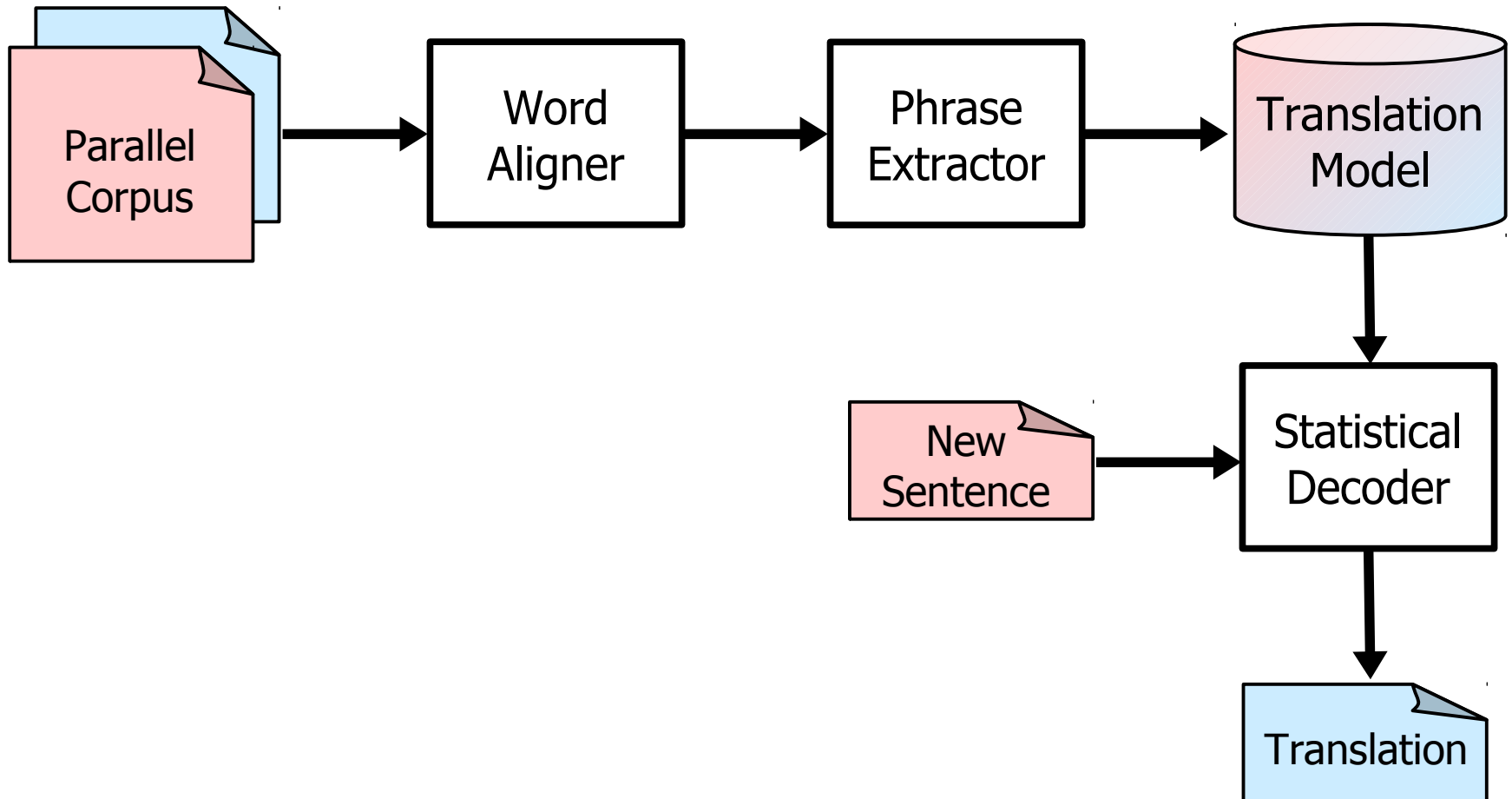
How Can Computers Translate?

- Result: **statistical machine translation**



How Can Computers Translate?

- Result: **statistical machine translation**



“Automatically Improved Category
Labels for SCFG-Based Statistical
Machine Translation”

SCFG-Based Translation

- Basic statistical machine translation uses no **linguistic** knowledge
- We can use a CFG to model the syntax of one language
- Can we use a **synchronous CFG** to model two languages at once?

SCFG-Based Translation

Le vote final du Parlement aura lieu demain .

The

vote

last

last vote

SCFG-Based Translation

Le vote final du Parlement aura lieu demain .

The

vote

last

last vote

Le vote provisoire du Parlement aura lieu demain .

The

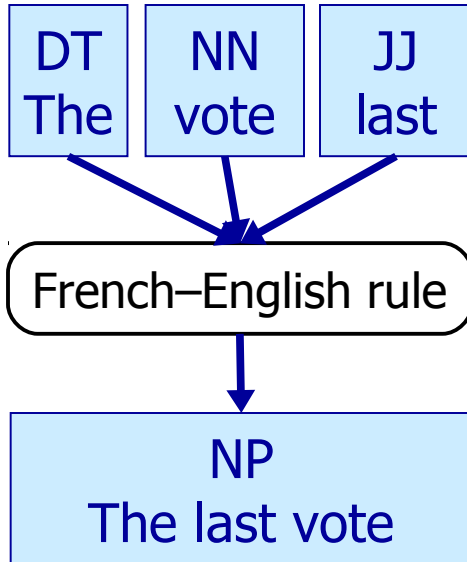
vote

provisional

???

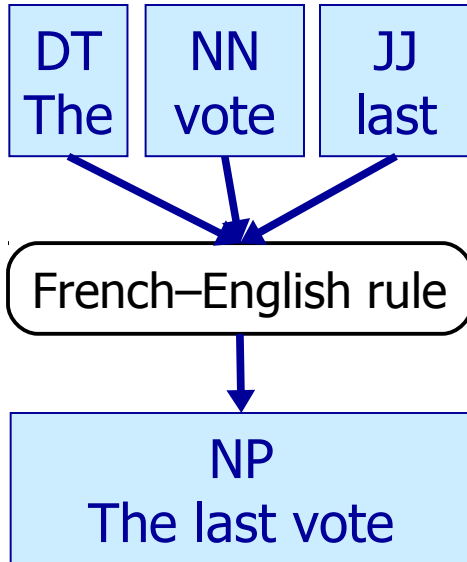
SCFG-Based Translation

Le vote final du Parlement aura lieu demain .



SCFG-Based Translation

Le vote final du Parlement aura lieu demain .

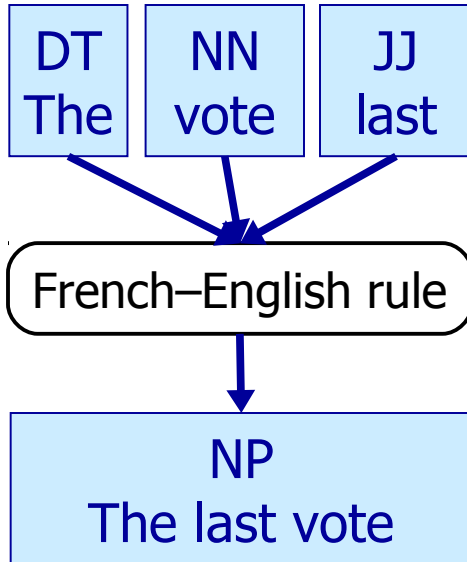


$NP \rightarrow D \ N \ A$

$NP \rightarrow DT \ JJ \ NN$

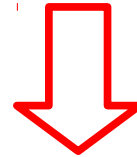
SCFG-Based Translation

Le vote final du Parlement aura lieu demain .



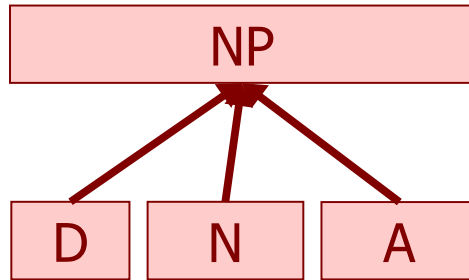
$NP \rightarrow D \ N \ A$

$NP \rightarrow DT \ JJ \ NN$

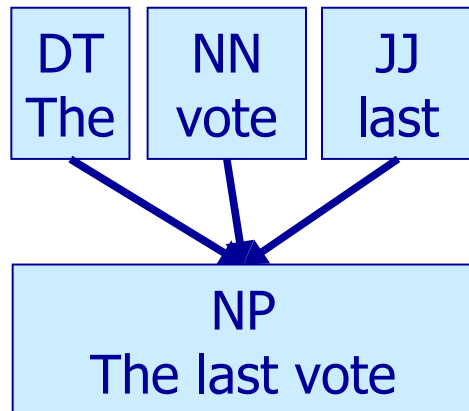


$NP::NP \rightarrow [D^1 \ N^2 \ A^3]::[DT^1 \ JJ^3 \ NN^2]$

SCFG-Based Translation

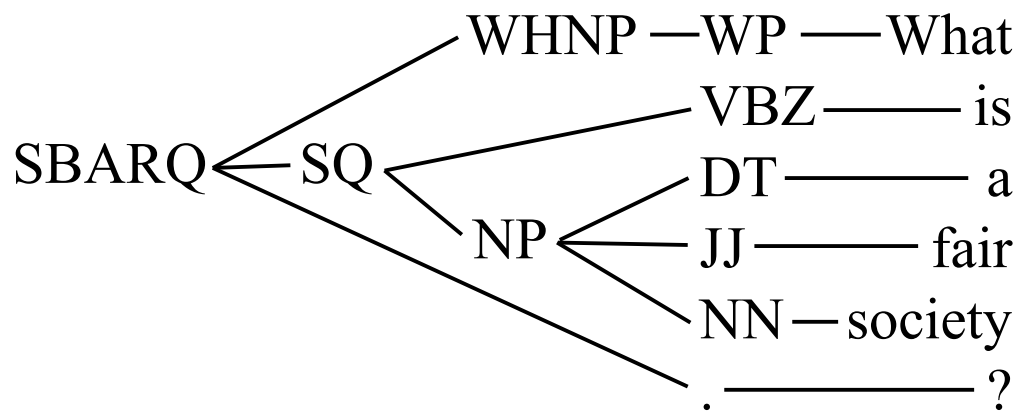
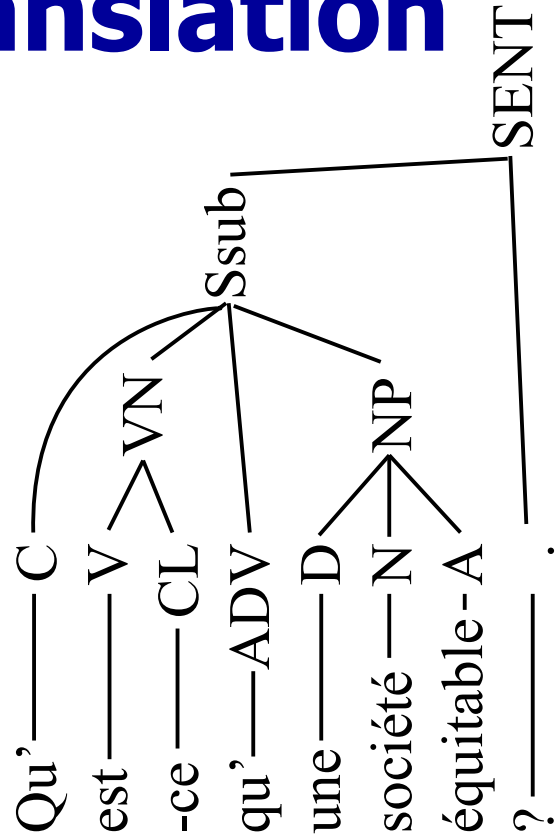


Le vote final du Parlement aura lieu demain .



$$\text{NP}::\text{NP} \rightarrow [\text{D}^1 \text{ N}^2 \text{ A}^3]::[\text{DT}^1 \text{ JJ}^3 \text{ NN}^2]$$

SCFG-Based Translation

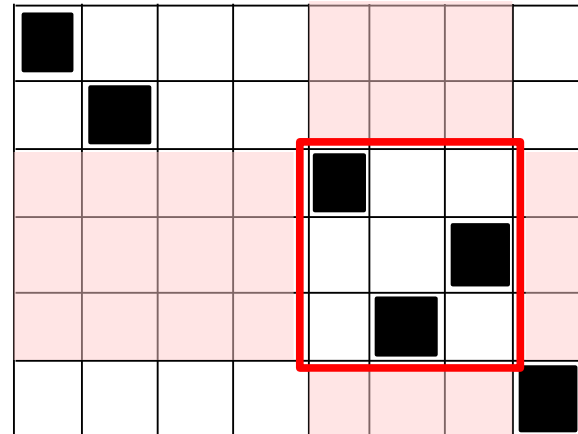
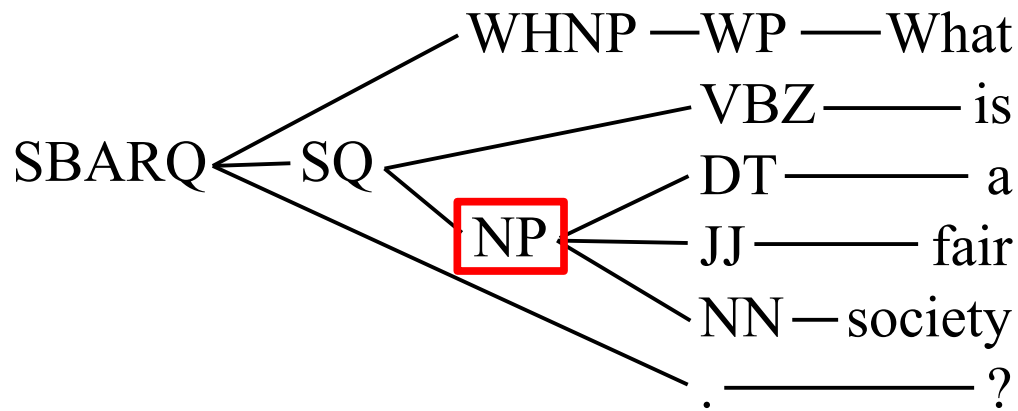
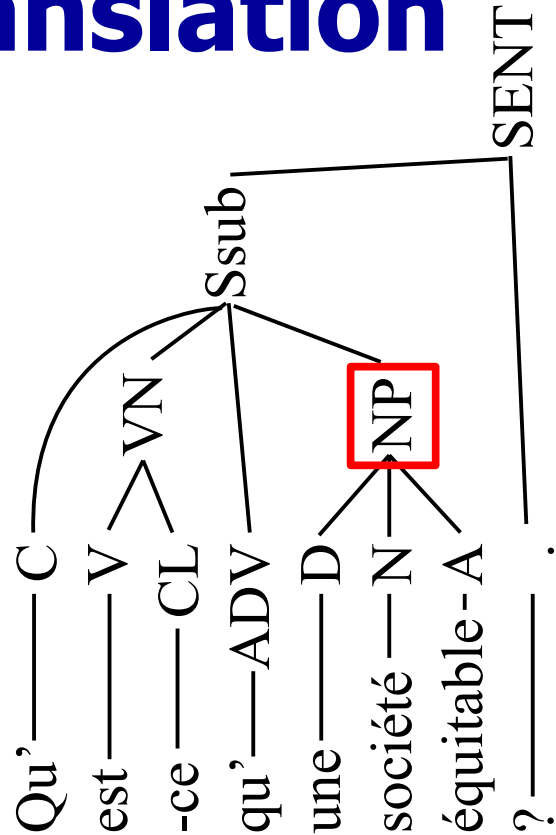


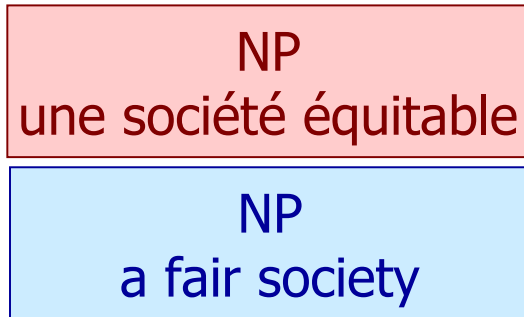
SCFG-Based Translation

NP
une société équitable

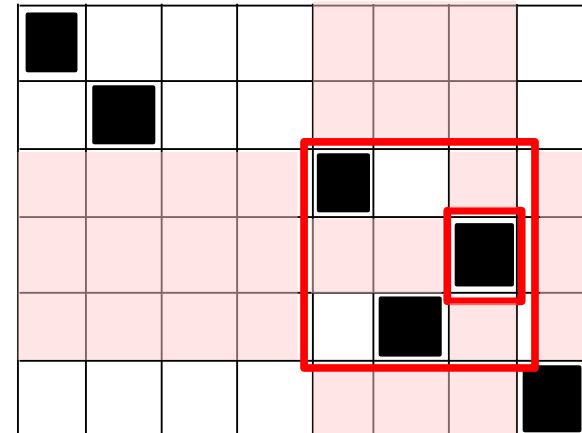
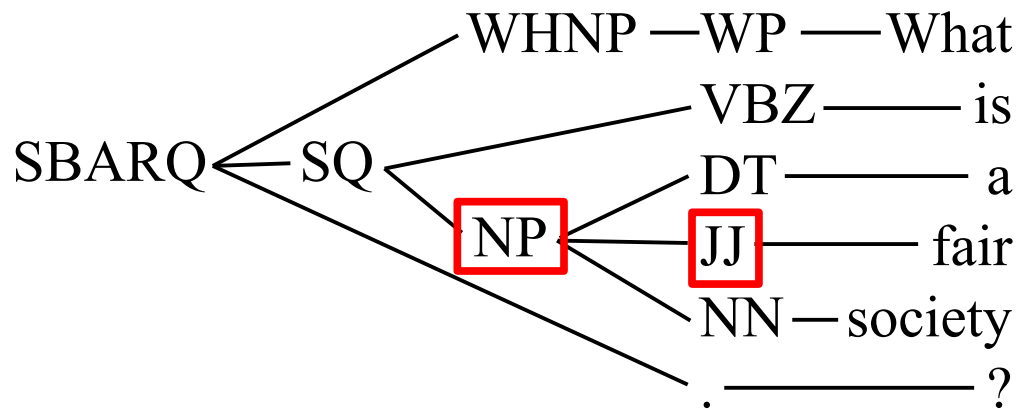
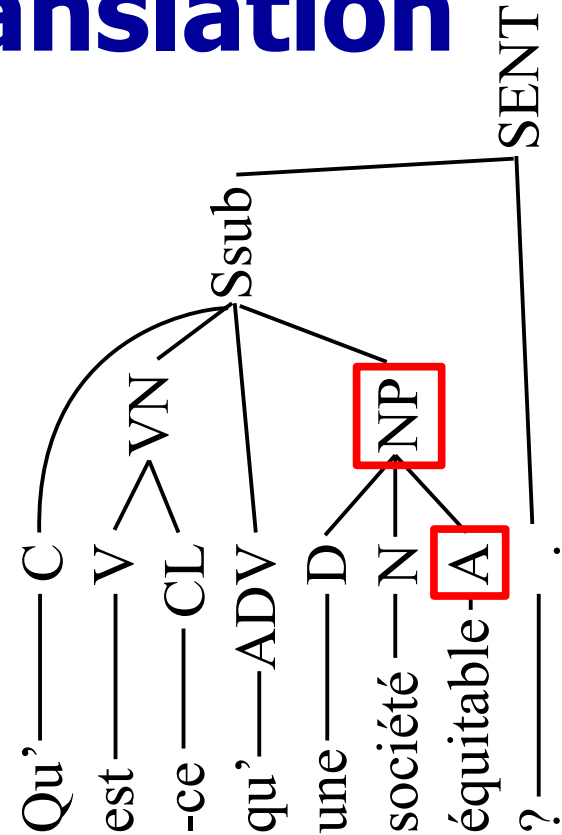
NP
a fair society



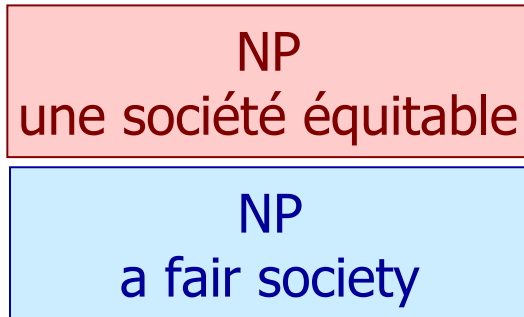
SCFG-Based Translation



$NP::NP \rightarrow [une\ société\ A^1]::[a\ JJ^1\ society]$

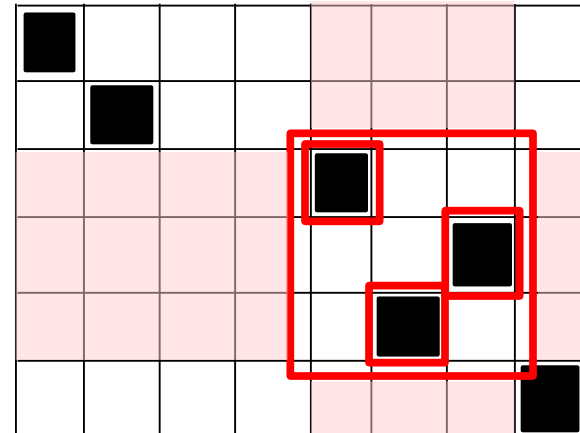
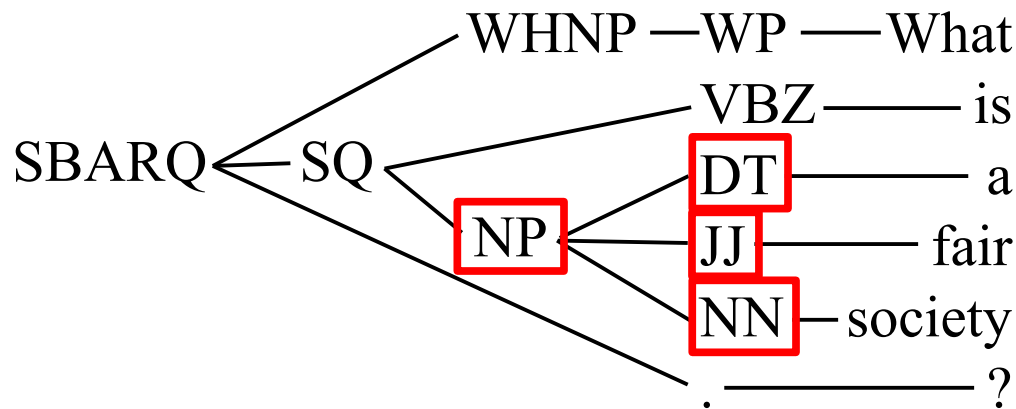
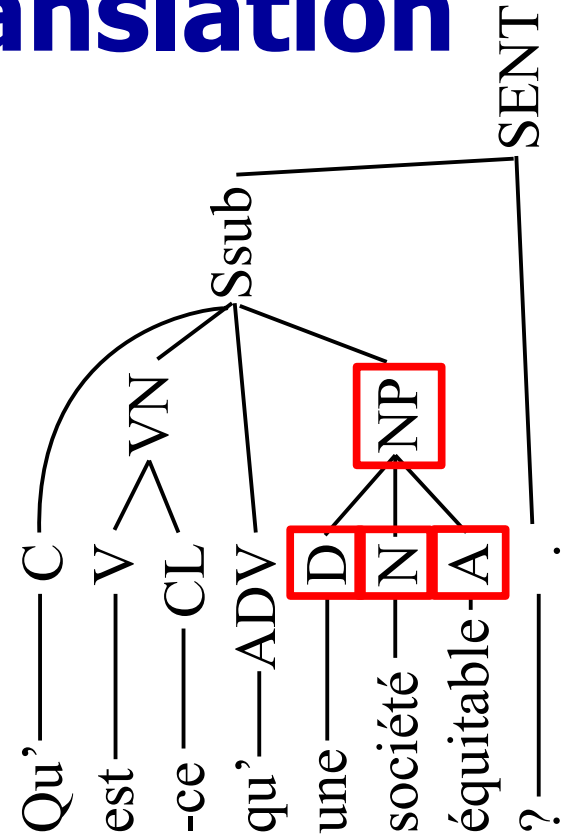


SCFG-Based Translation



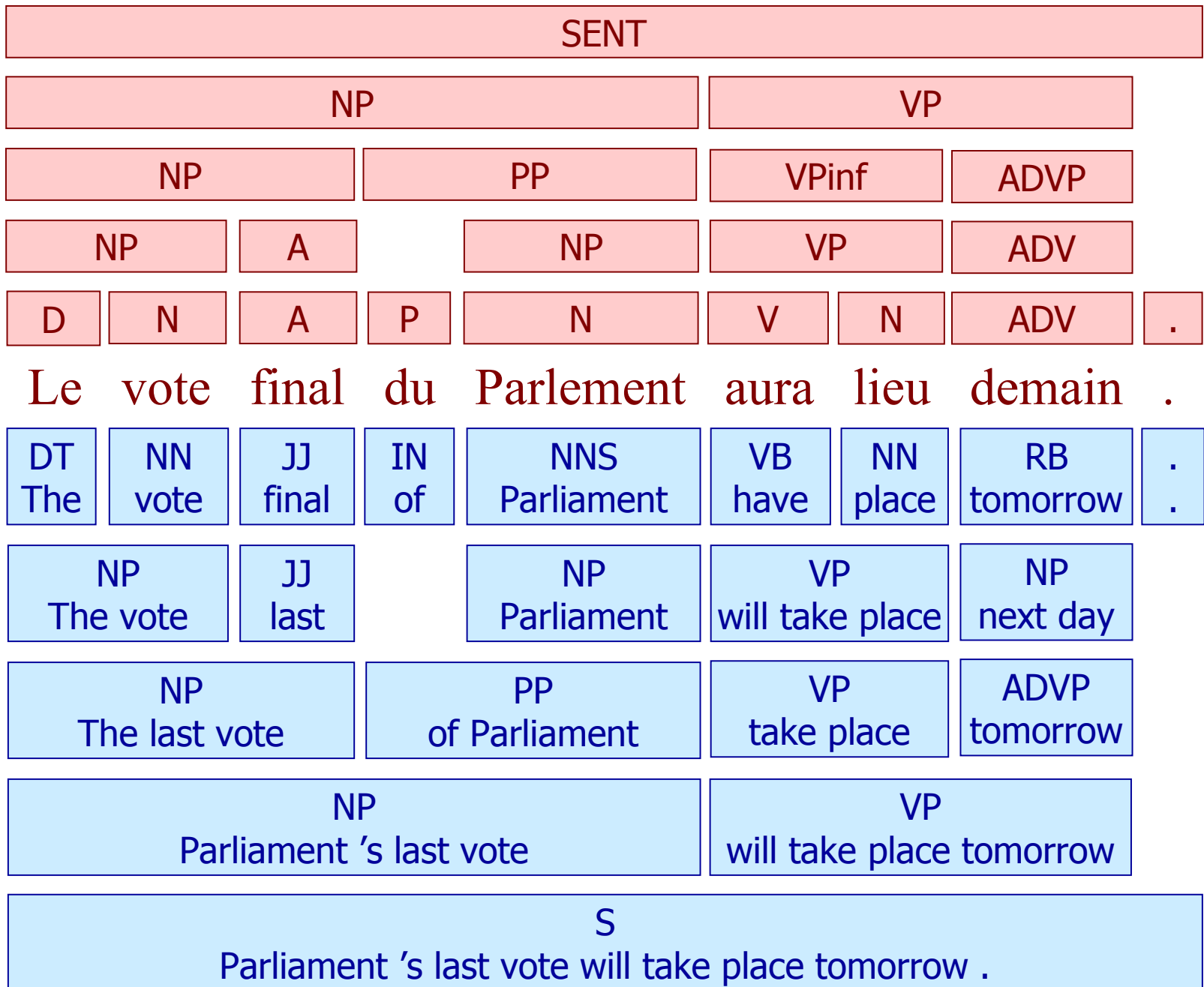
$NP::NP \rightarrow [une\ société\ A^1]::[a\ JJ^1\ society]$

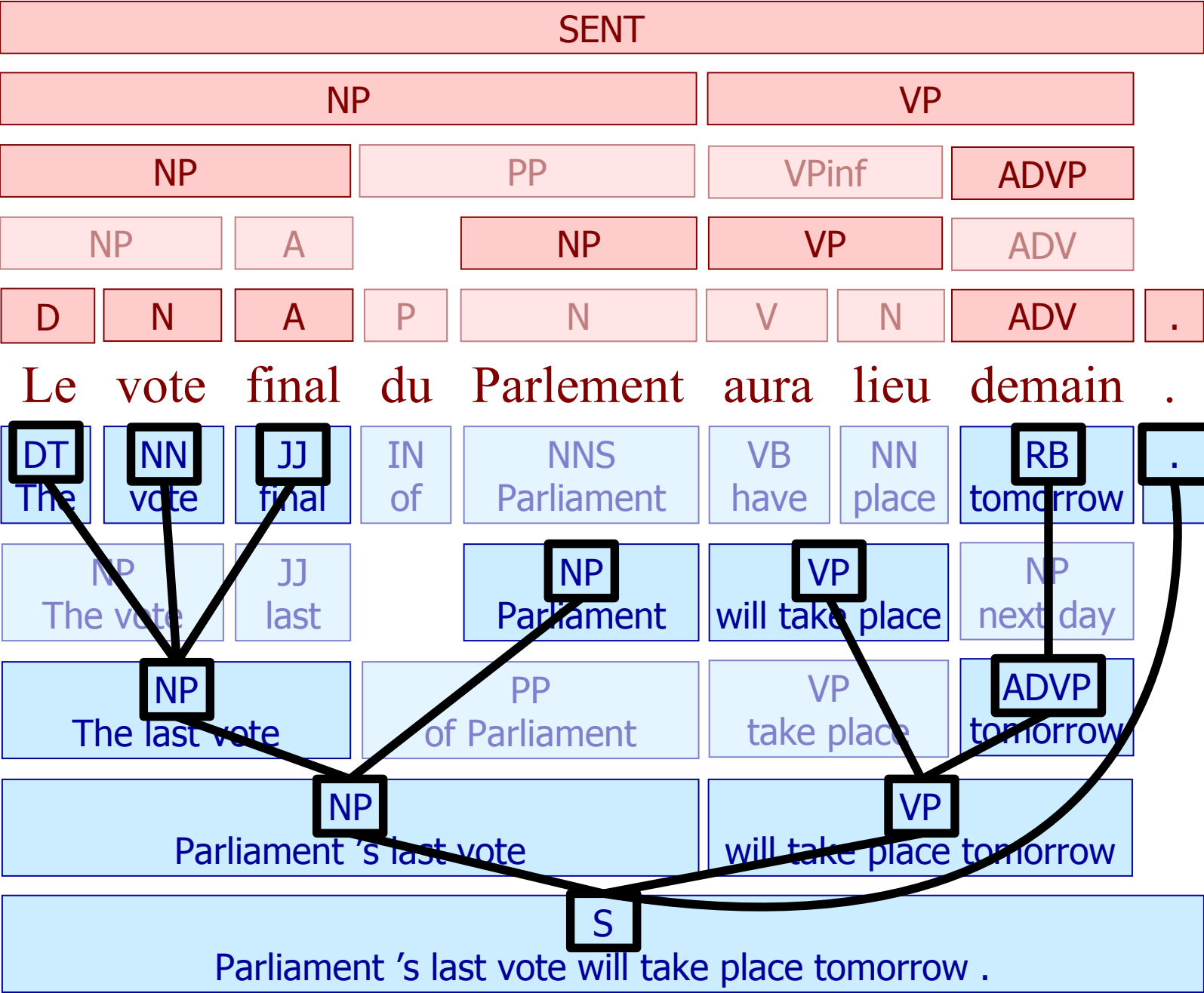
$NP::NP \rightarrow [D^1\ N^2\ A^3]::[DT^1\ JJ^3\ NN^2]$



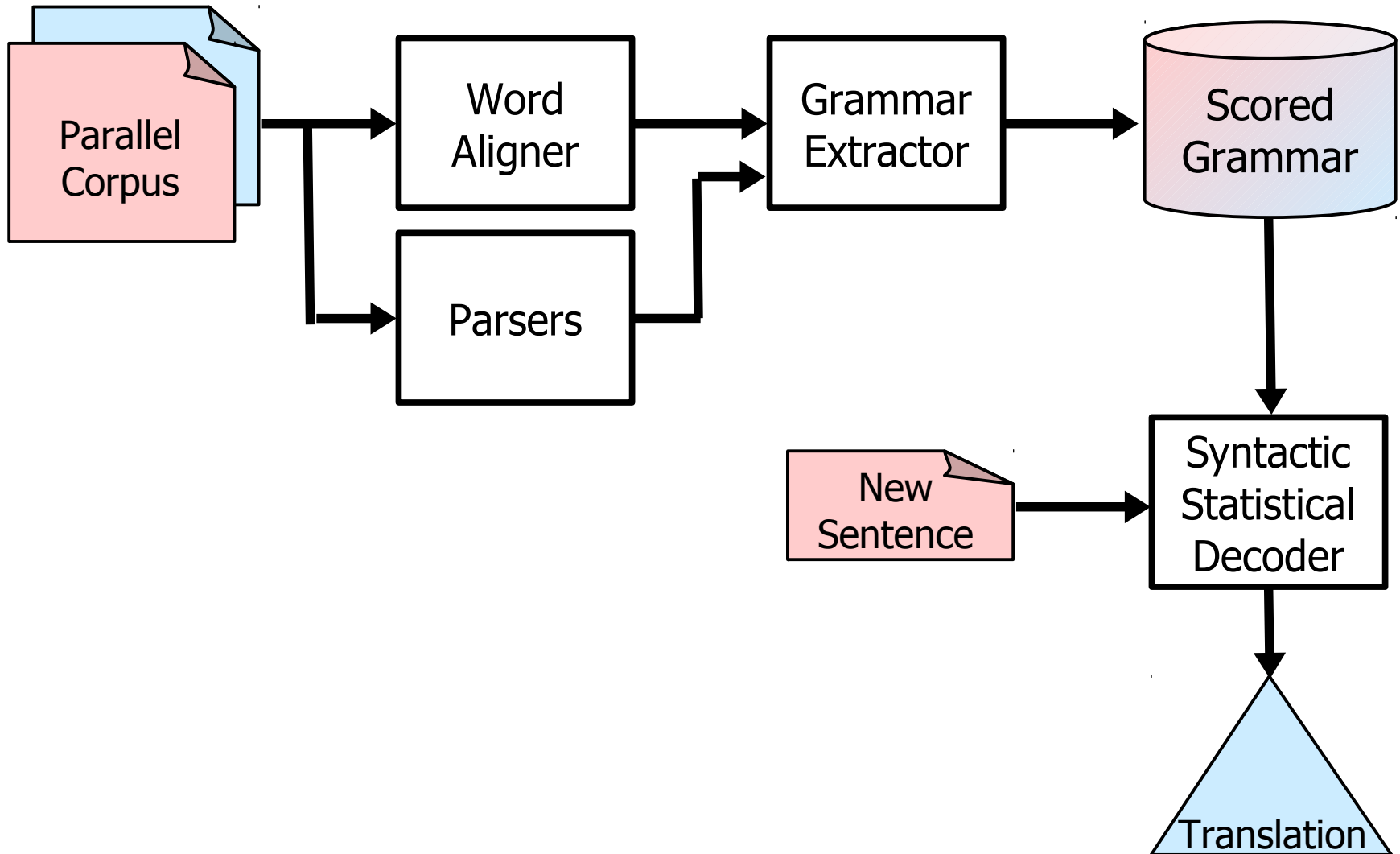
SCFG-Based Translation

Le vote final du Parlement aura lieu demain .





SCFG-Based Translation



**“Automatically Improved Category
Labels for SCFG-Based Statistical
Machine Translation”**

“Interesting Topics of Current Research in SCFG-Based Statistical Machine Translation”

Research Topics

- What is the best nonterminal set?
 - Where do labels come from?
 - How can we make them better?
 - Does the specific language pair matter?
- Which rules are good? Which are bad?
How many rules do I need?
- What happens if I don't have two parsers?
- How else can rules be learned?

Research Topics

- What statistics should I collect?
- Can grammars help word alignment?
- How can I efficiently store, search, and apply tons of rules?
- How can I tell how good a translation is?
- ...?

