

# Syntax-based Translation

## Part 2: Synchronous Grammars

March 20, 2014

# Goals

- Revisit why people thought syntax would not help machine translation
- Learn about **Synchronous Context Free Grammars**
- Introduce **notation**, and **basic algorithm**
- Understand how we **learn SCFGs from bitexts**
- Get a sense of the different flavors of SCFGs
  - **Hiero**
  - **SAMT**

# The Syntax Bet

- Longstanding debate about whether linguistic information can help statistical translation
- Two camps



# The Syntax Bet

- Longstanding debate about whether linguistic information can help statistical translation
- Two camps

Syntax will improve  
translation



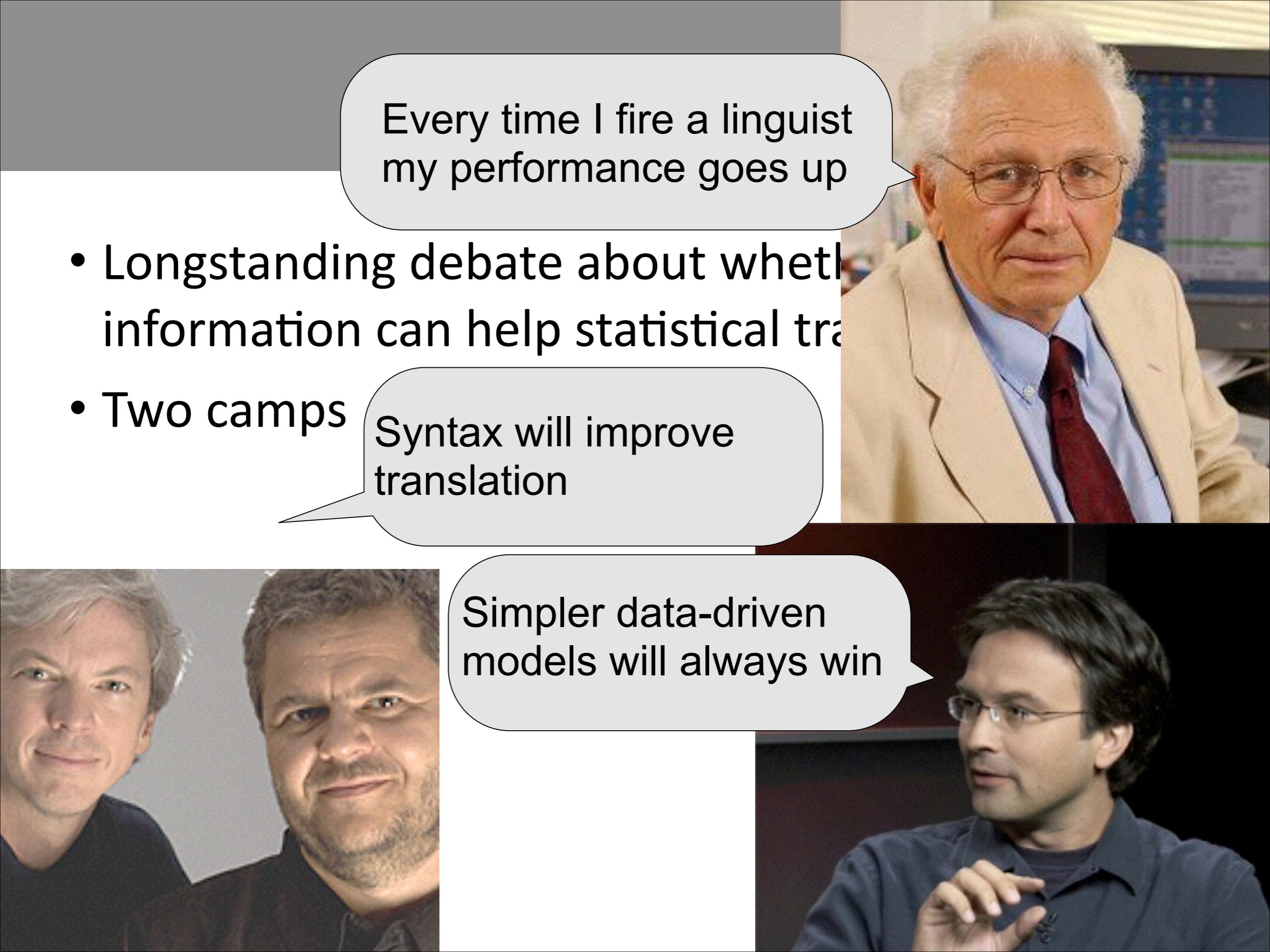
# The Syntax Bet

- Longstanding debate about whether linguistic information can help statistical translation
- Two camps



Syntax will improve translation

Simpler data-driven models will always win



Every time I fire a linguist  
my performance goes up

- Longstanding debate about whether information can help statistical translation
- Two camps

Syntax will improve translation

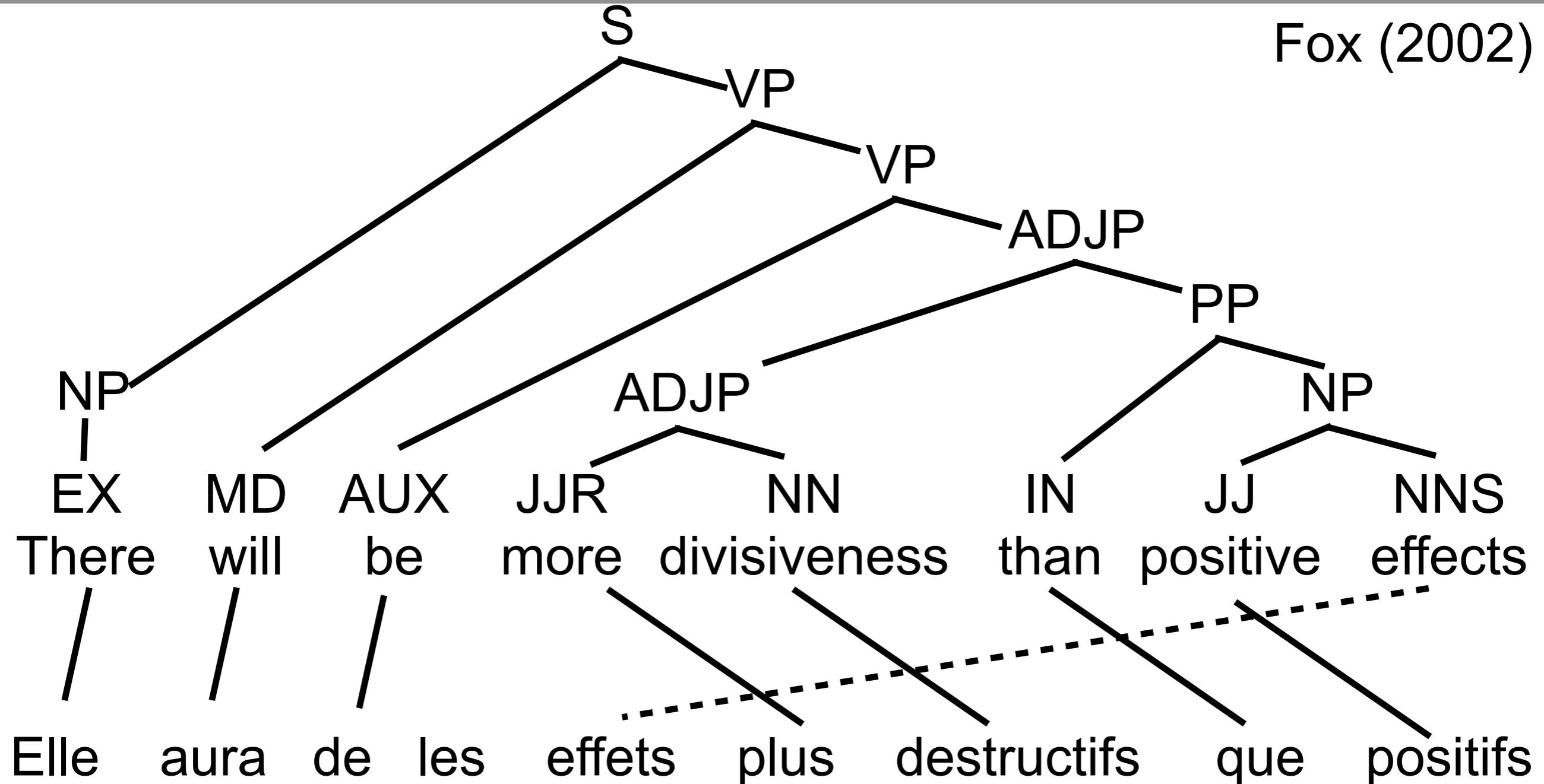
Simpler data-driven models will always win

# Syntax is bad for translation

- The IBM Models were the dominant approach to SMT from the ‘90s until mid 2000s
  - Eschewed linguistic information
- A number of studies cast doubt on whether linguistic info could help SMT
  - Fox (2002) showed that “phrasal cohesion” was less common than assumed across even related languages
  - Koehn et al (2003) empirically demonstrated that syntactically motivated phrases made PBMT worse

# Phrases aren't coherent in bitexts

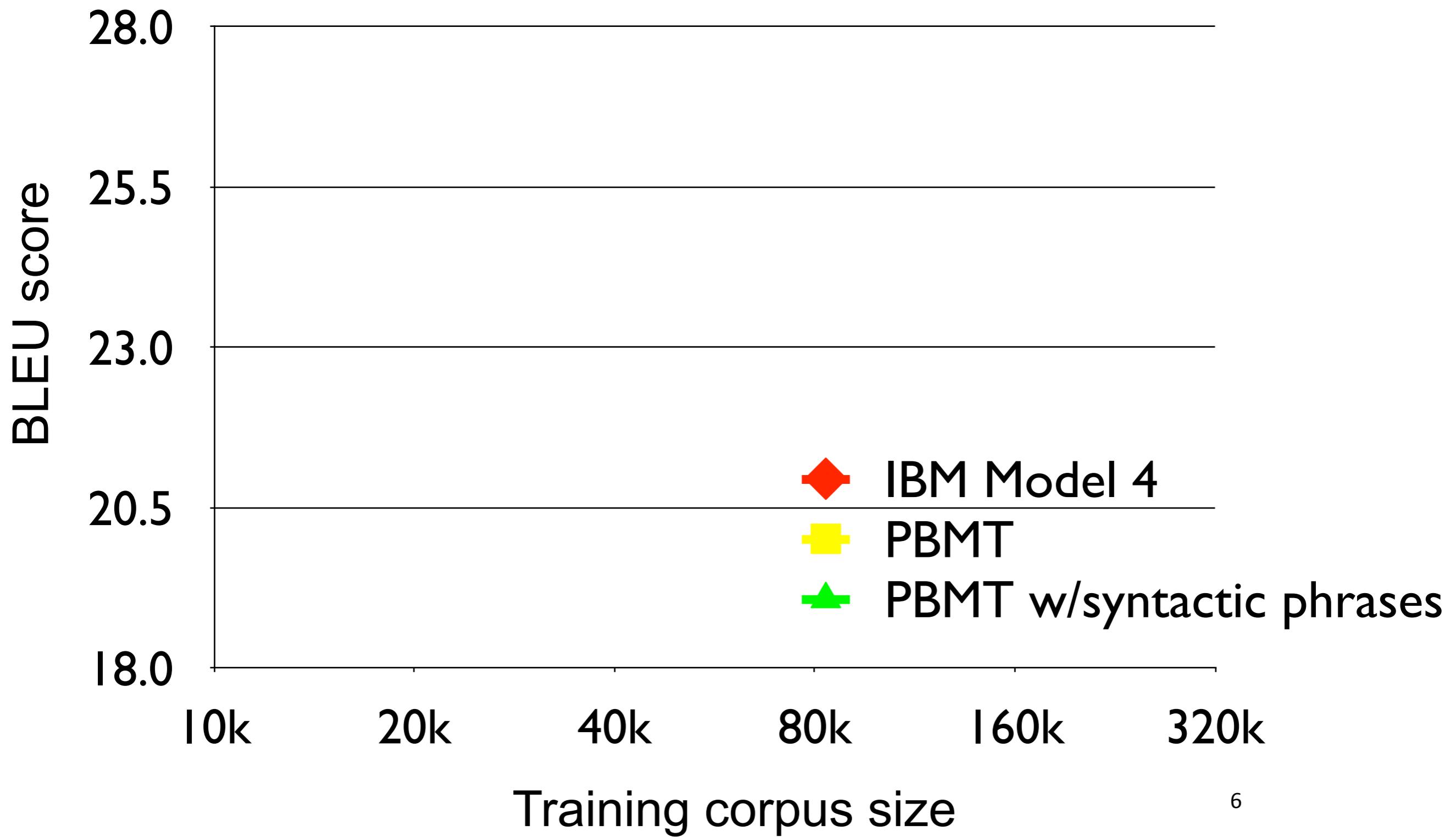
Fox (2002)



Gloss: *It will have effects more destructive than positive*

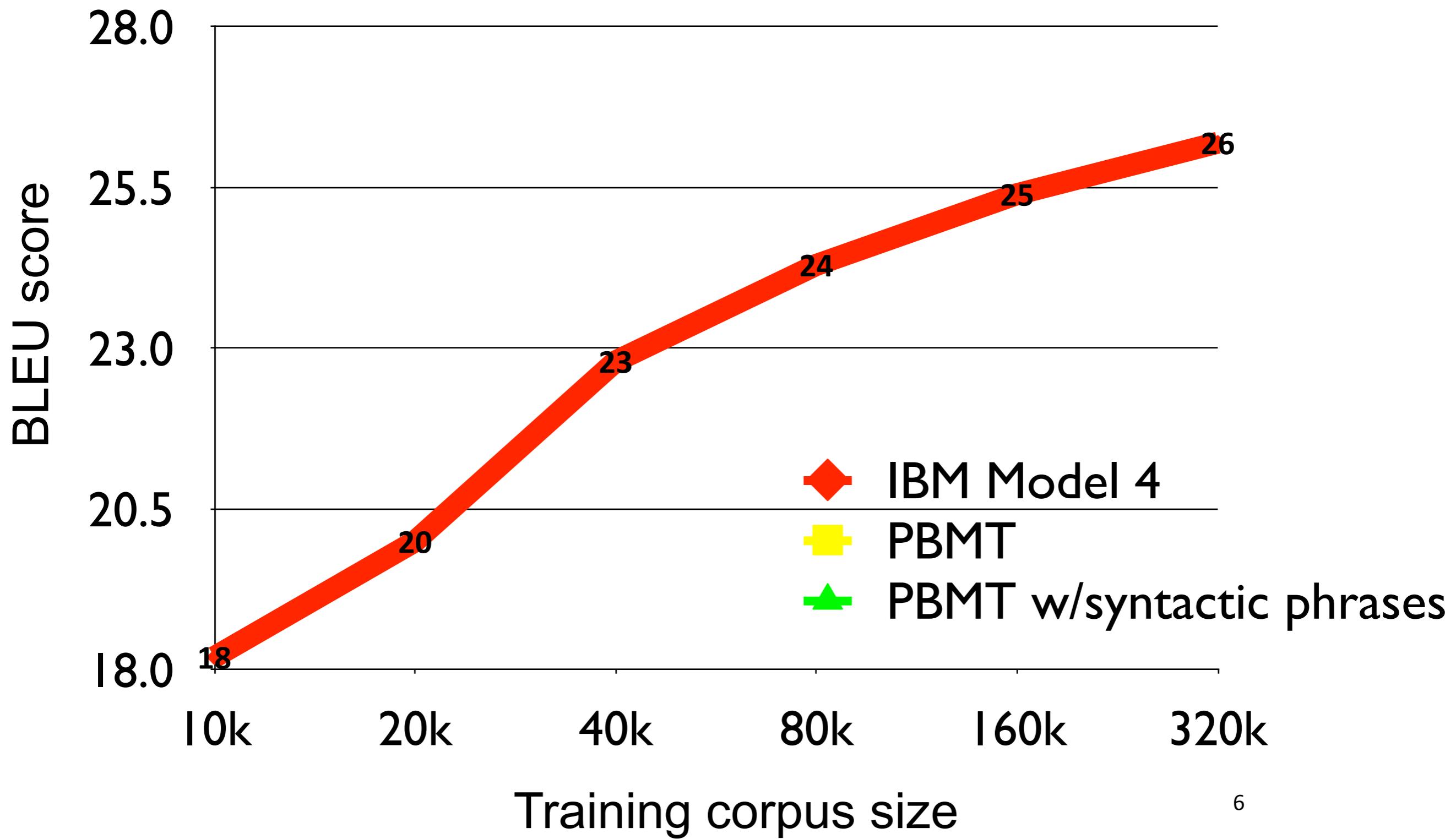
# Ouch! Syntax hurts!

Koehn et al (2003)



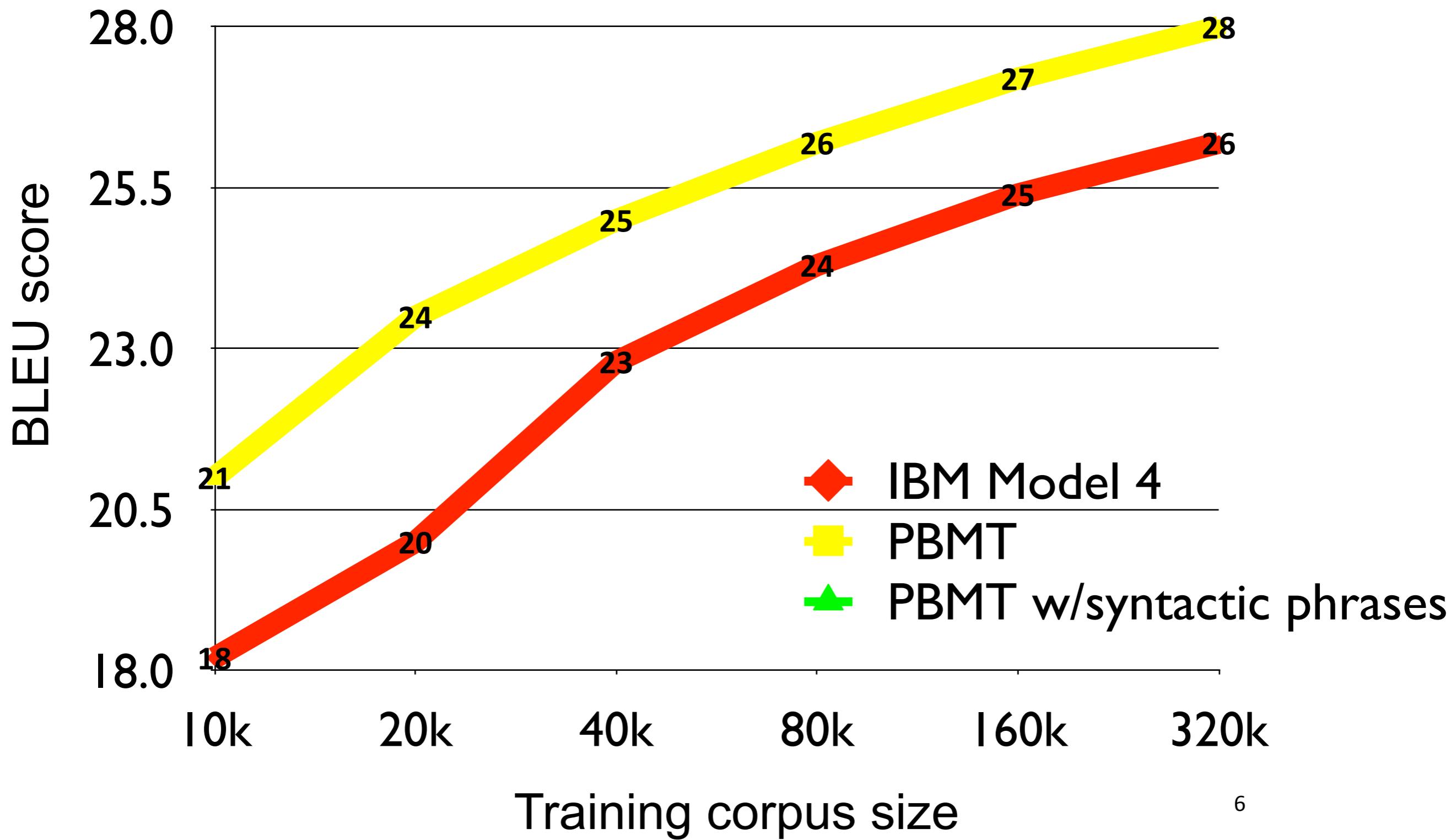
# Ouch! Syntax hurts!

Koehn et al (2003)



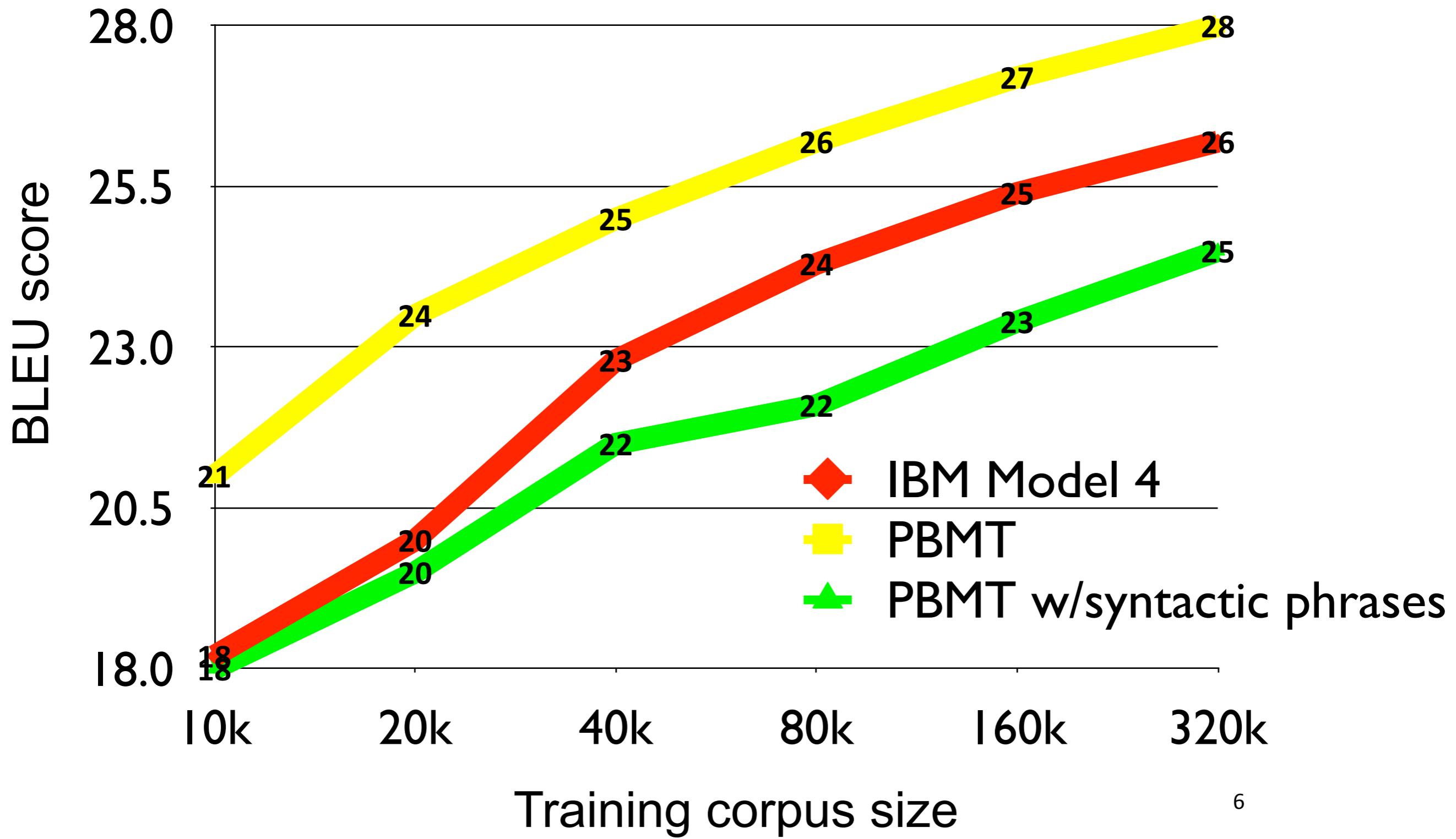
# Ouch! Syntax hurts!

Koehn et al (2003)



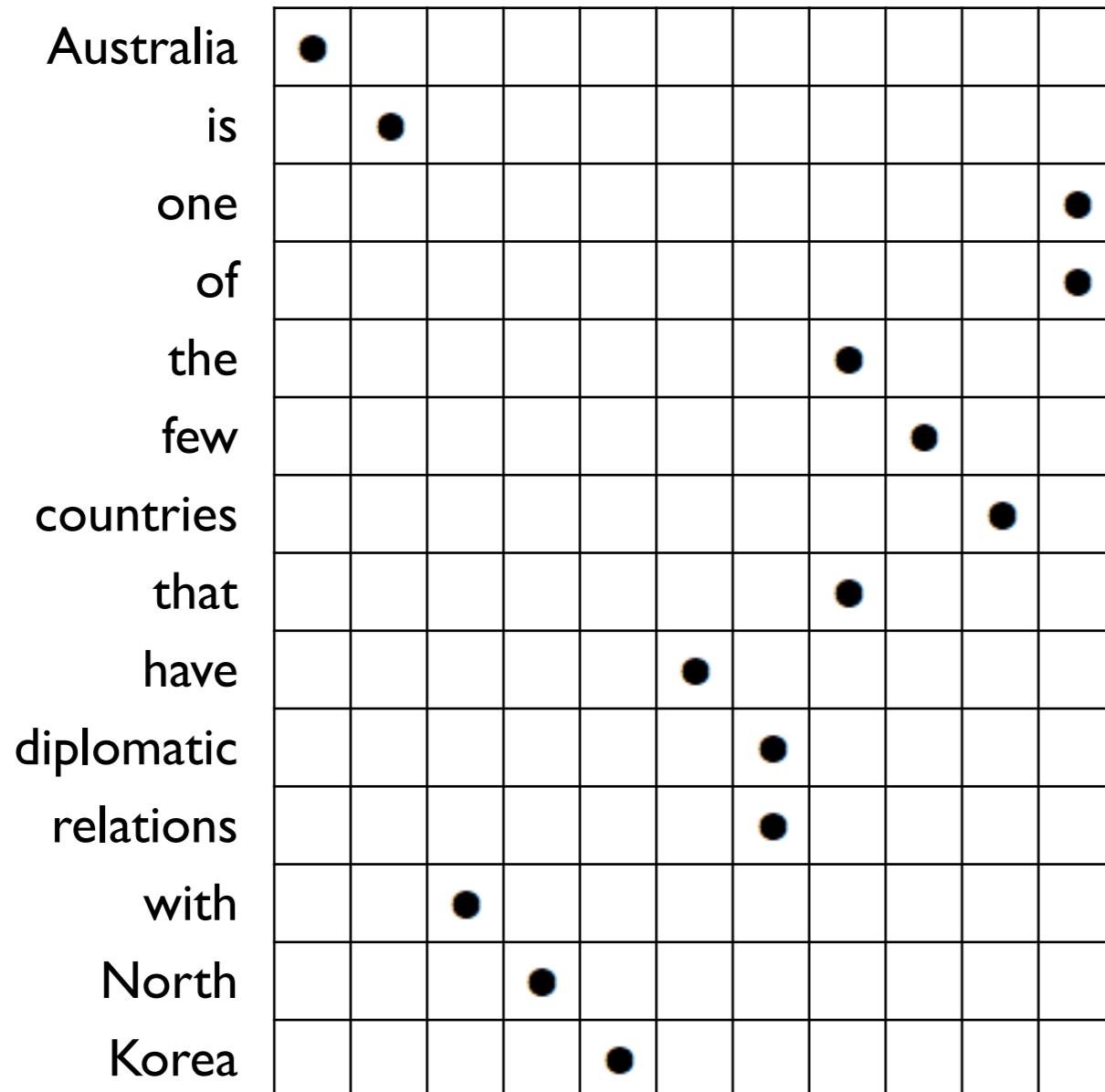
# Ouch! Syntax hurts!

Koehn et al (2003)



# Extracting phrase pairs

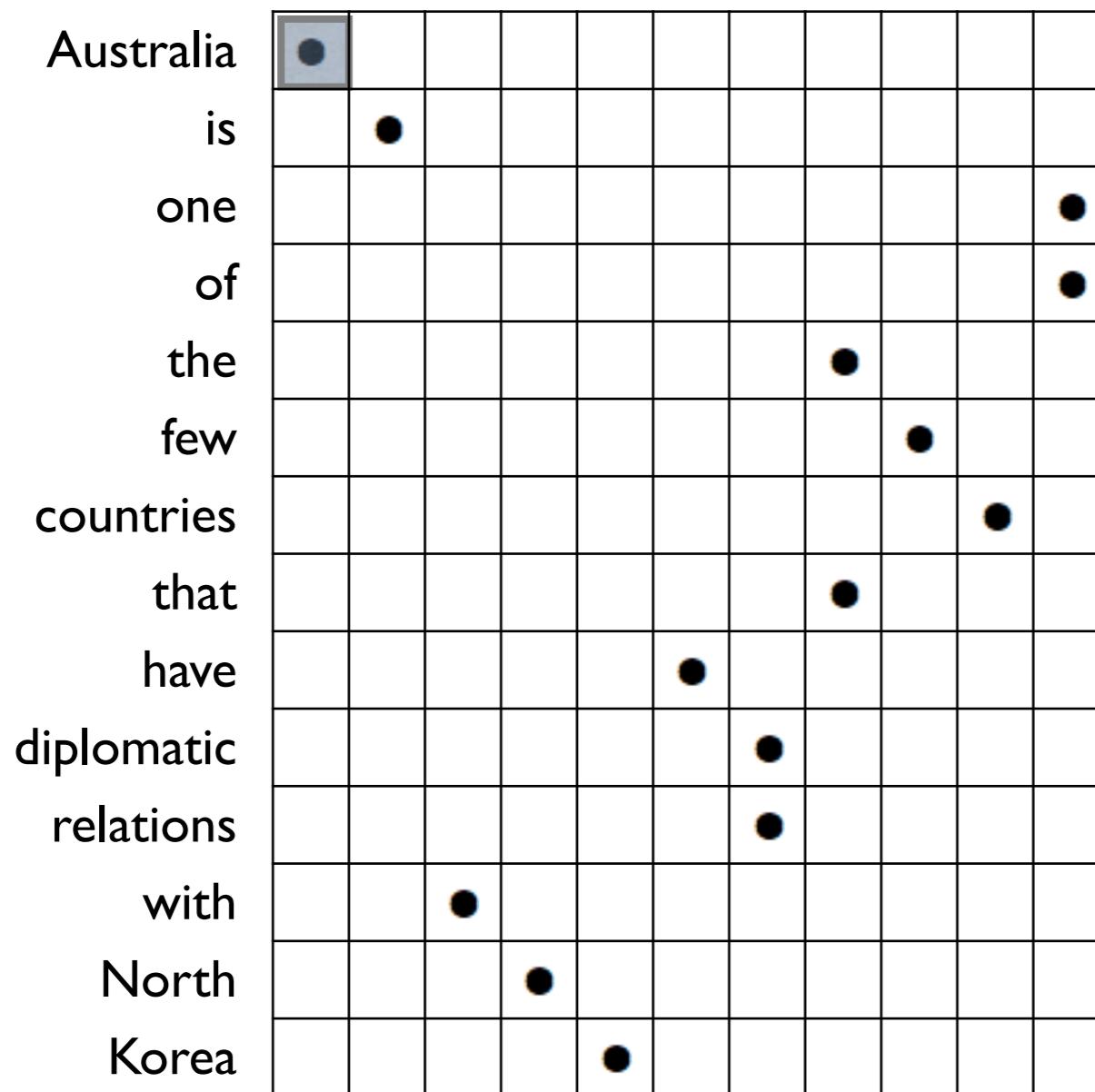
澳洲 是 与 北 韩 有 邦 交 的 少 数 国 家 之



# Extracting phrase pairs

# 澳洲, Australia

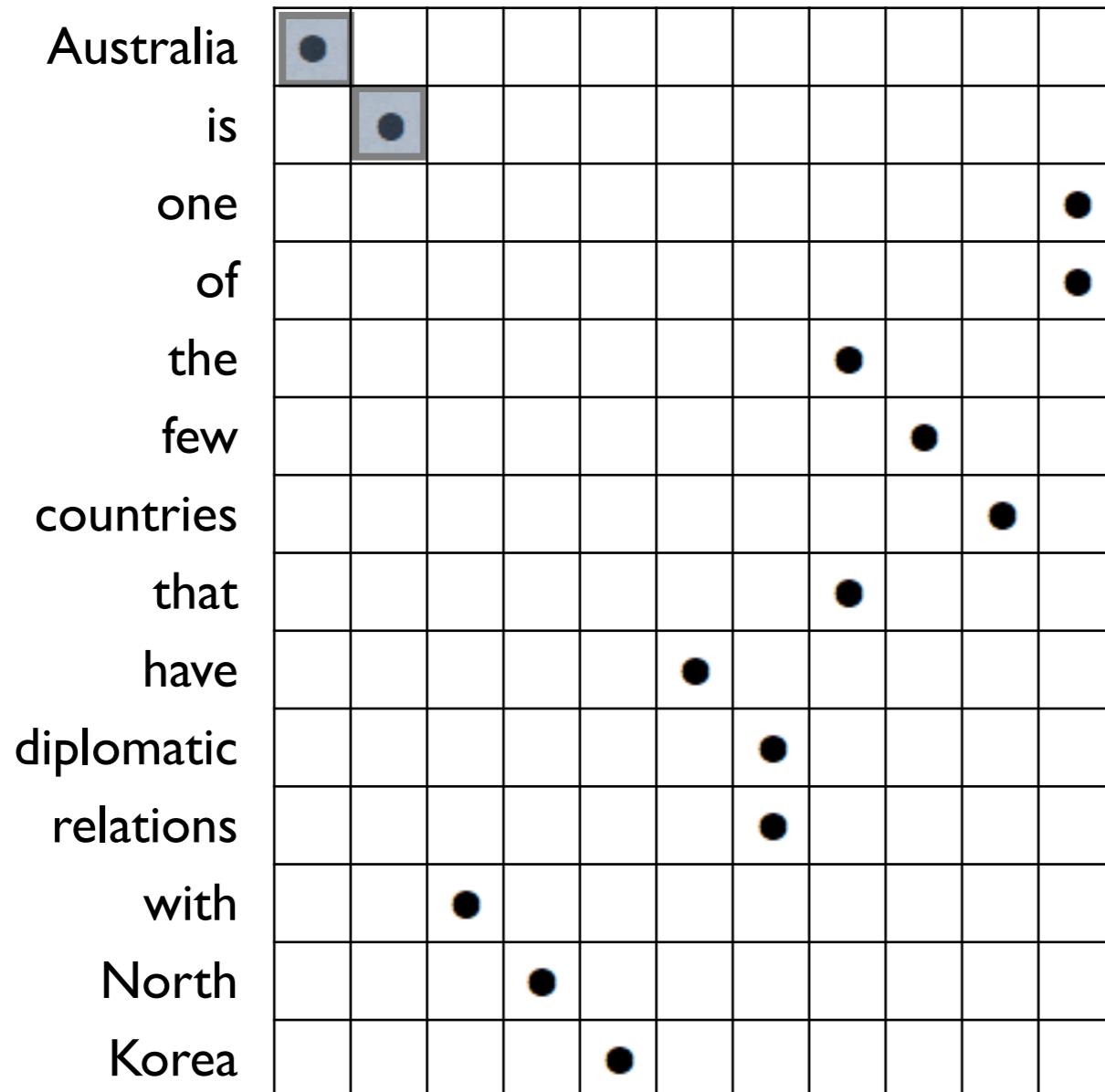
# 澳洲是与北韩有邦交的少数国家之一



# Extracting phrase pairs

澳洲 是 与 北 韩 有 邦 交 的 少 数 国 家 之  
洲

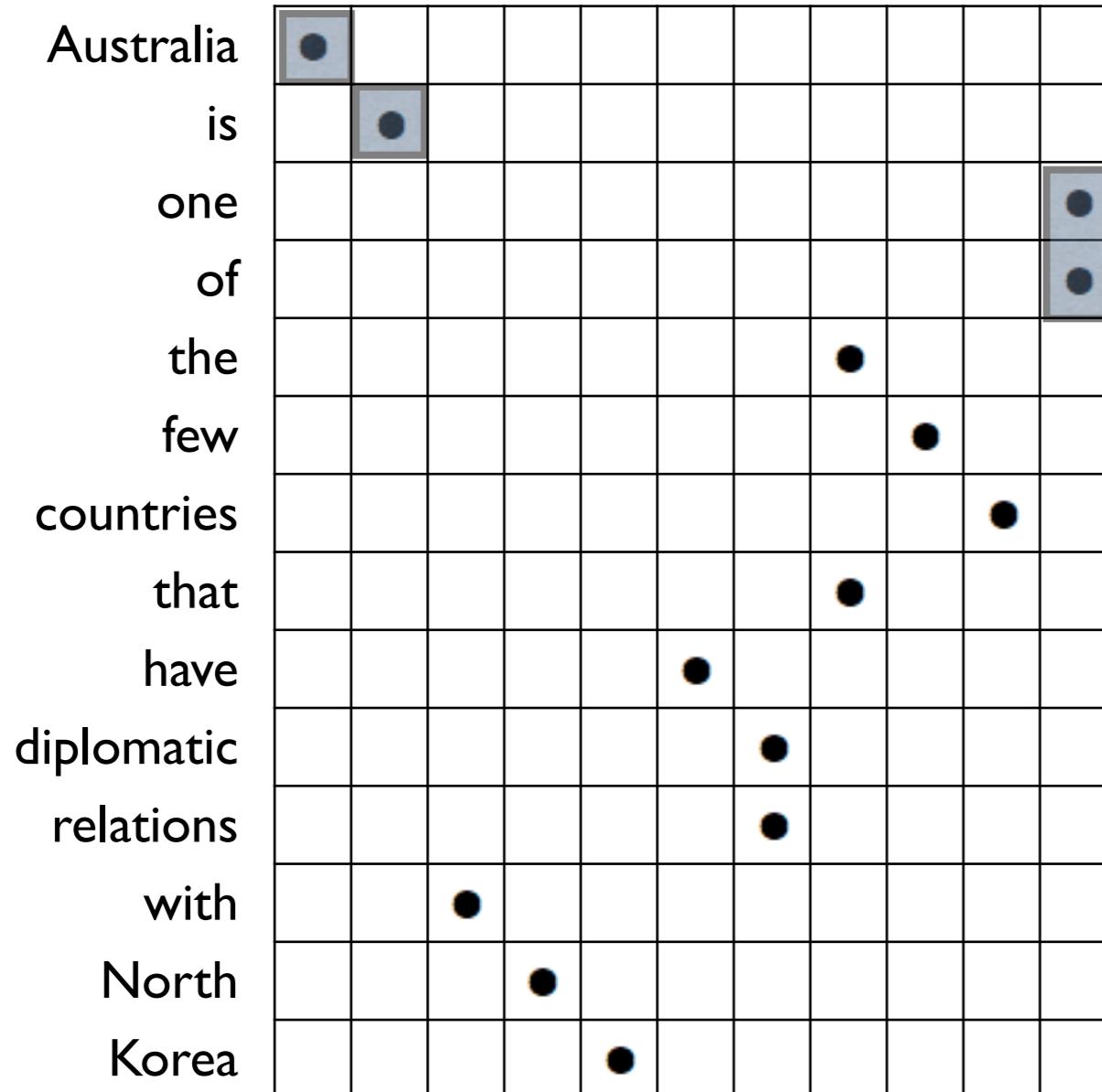
澳洲, Australia  
是, is



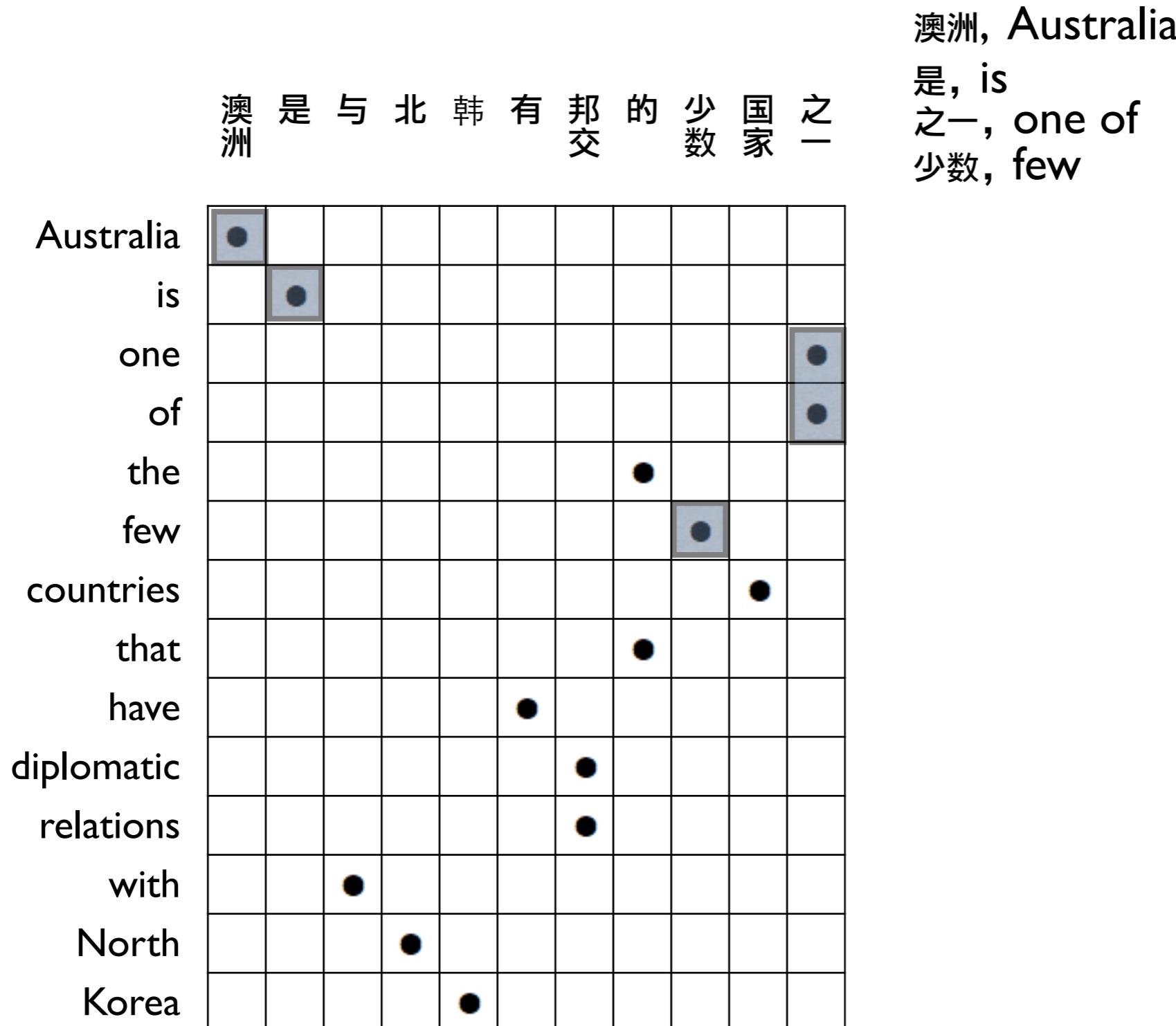
# Extracting phrase pairs

澳洲 是 与 北 韩 有 邦 交 的 少 数 国 家 之

澳洲, Australia  
是, is  
之一, one of

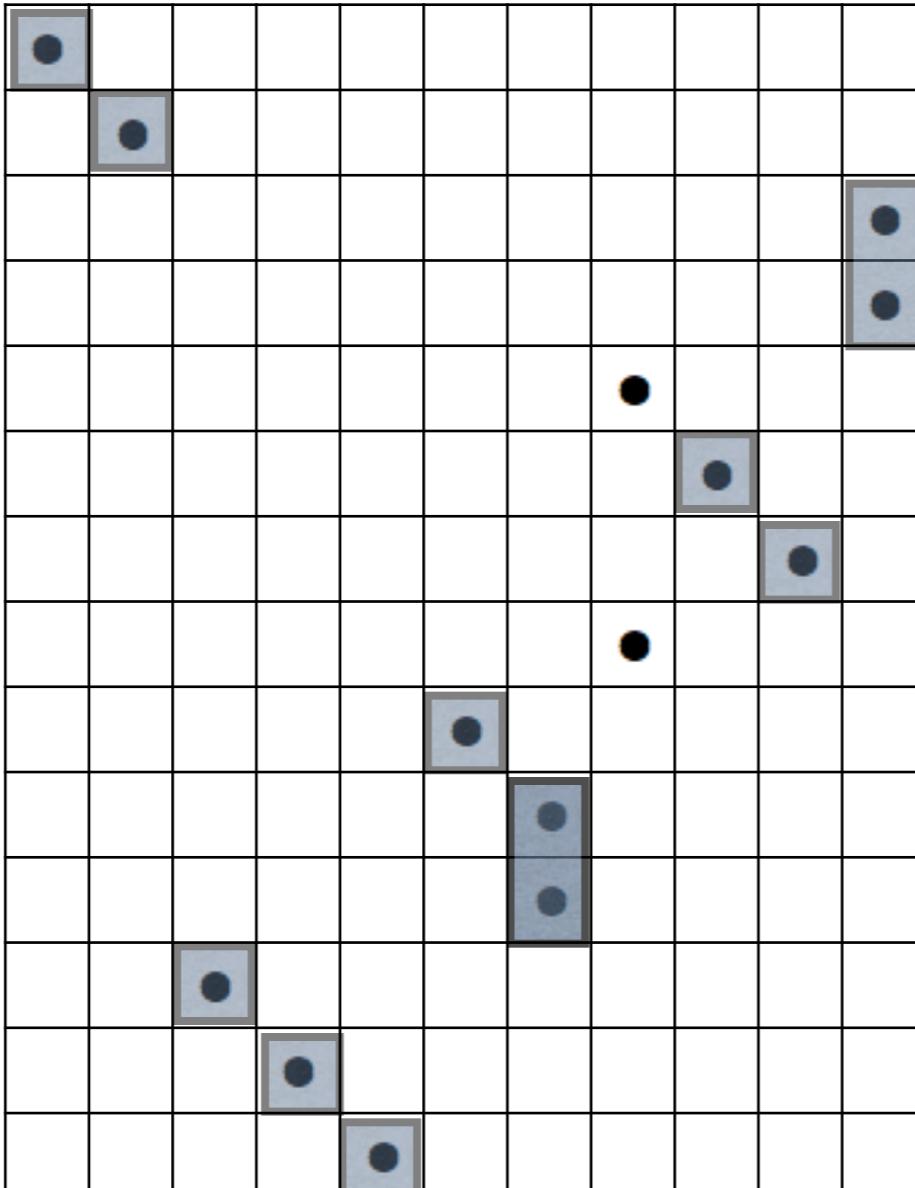


# Extracting phrase pairs



# Extracting phrase pairs

澳洲 是 与 北 韩 有 邦 交 的 少 数 国 家 之  
Australia is one of the few countries that have diplomatic relations with North Korea



澳洲, Australia

是, is

之一, one of

少数, few

国家, countries

有, have

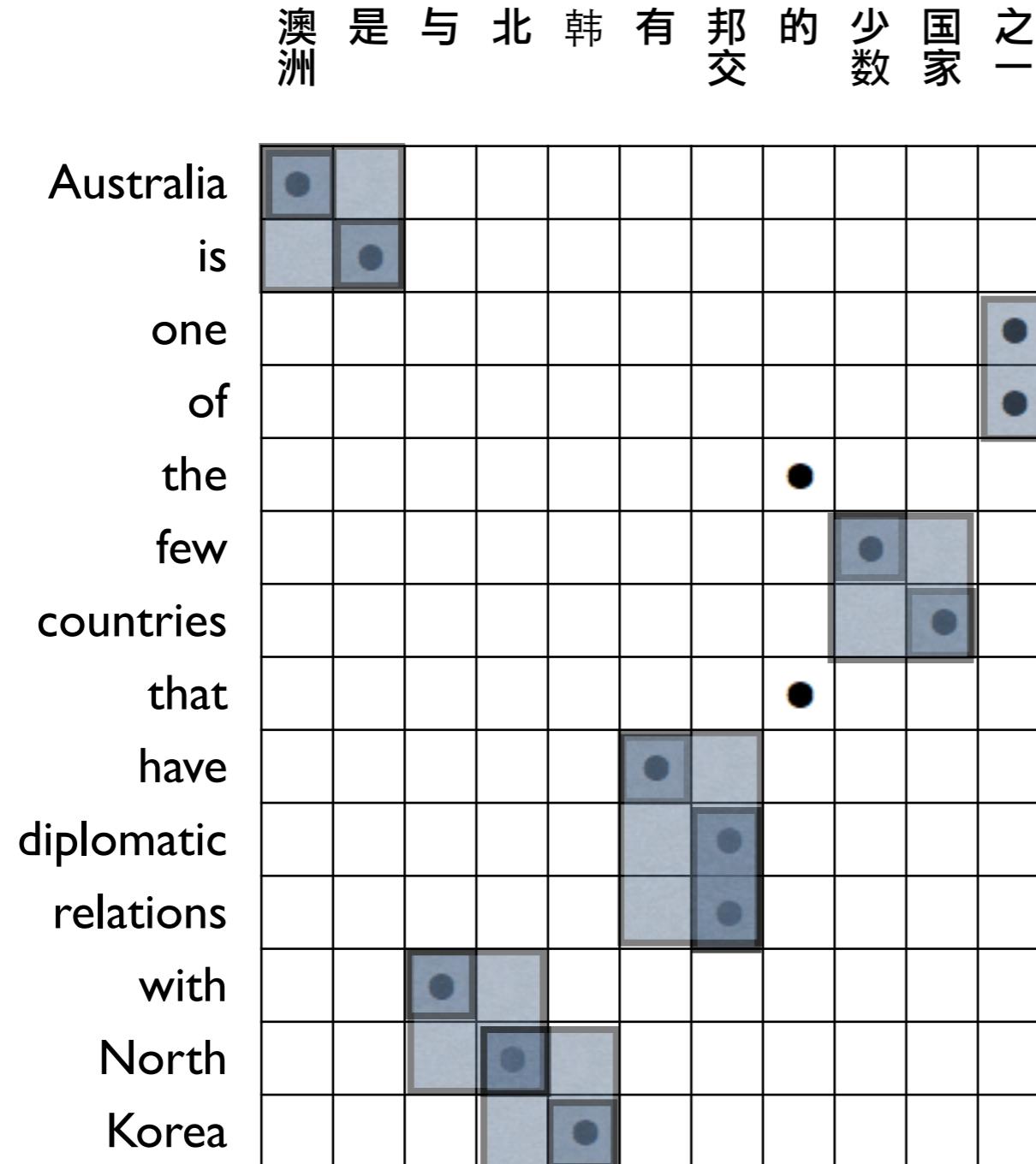
邦交, diplomatic relations

与, with

北, North

韩, Korea

# Extracting phrase pairs



澳洲, Australia

是, is

之一, one of

少数, few

国家, countries

有, have

邦交, diplomatic relations

与, with

北, North

韩, Korea

澳洲是, Australia is

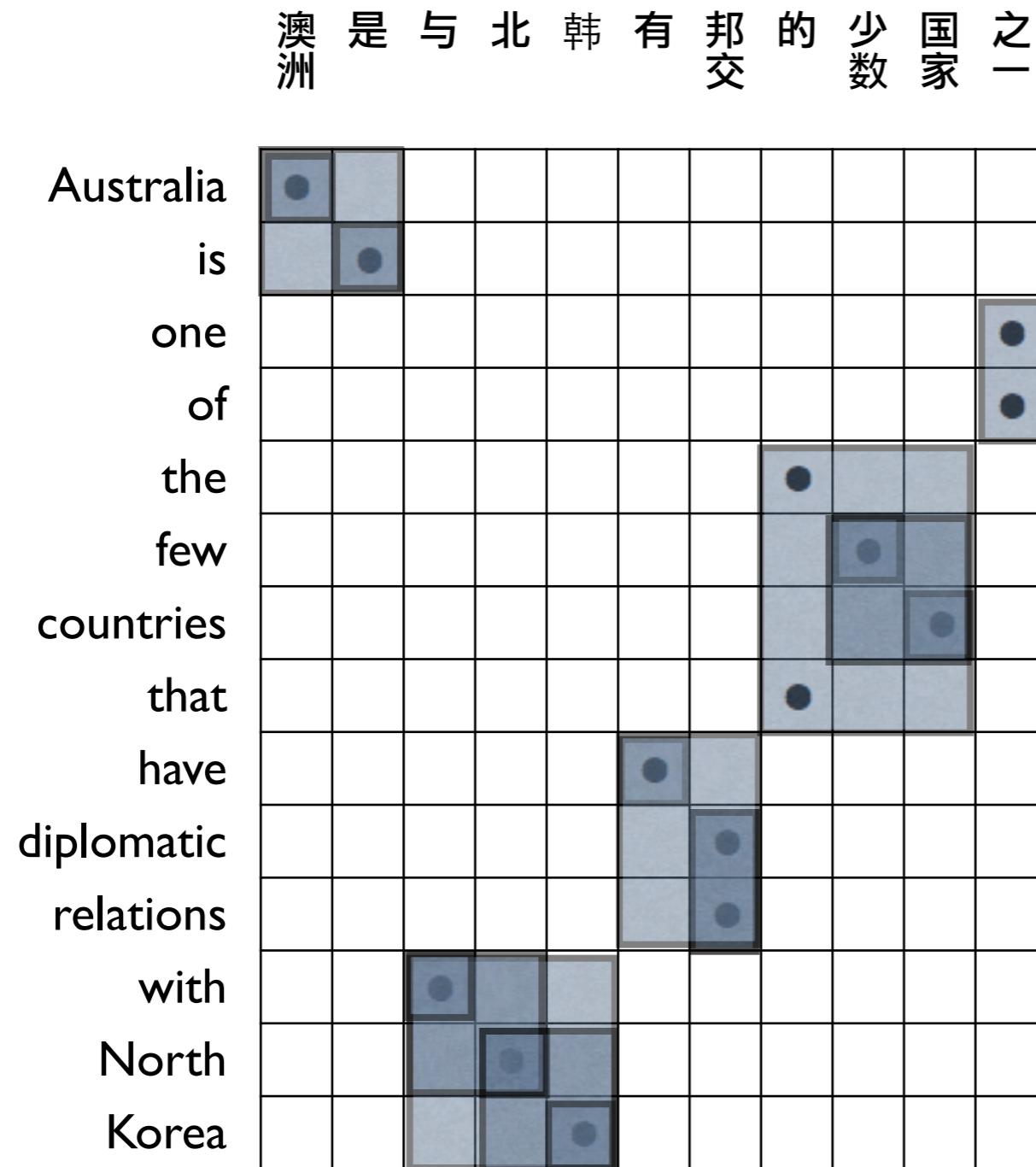
少数 国家, few countries

有邦交, have diplomatic relations

与北, with North

北韩, North Korea

# Extracting phrase pairs



澳洲, Australia

是, is

之一, one of

少数, few

国家, countries

有, have

邦交, diplomatic relations

与, with

北, North

韩, Korea

澳洲是, Australia is

少数 国家, few countries

有邦交, have diplomatic relations

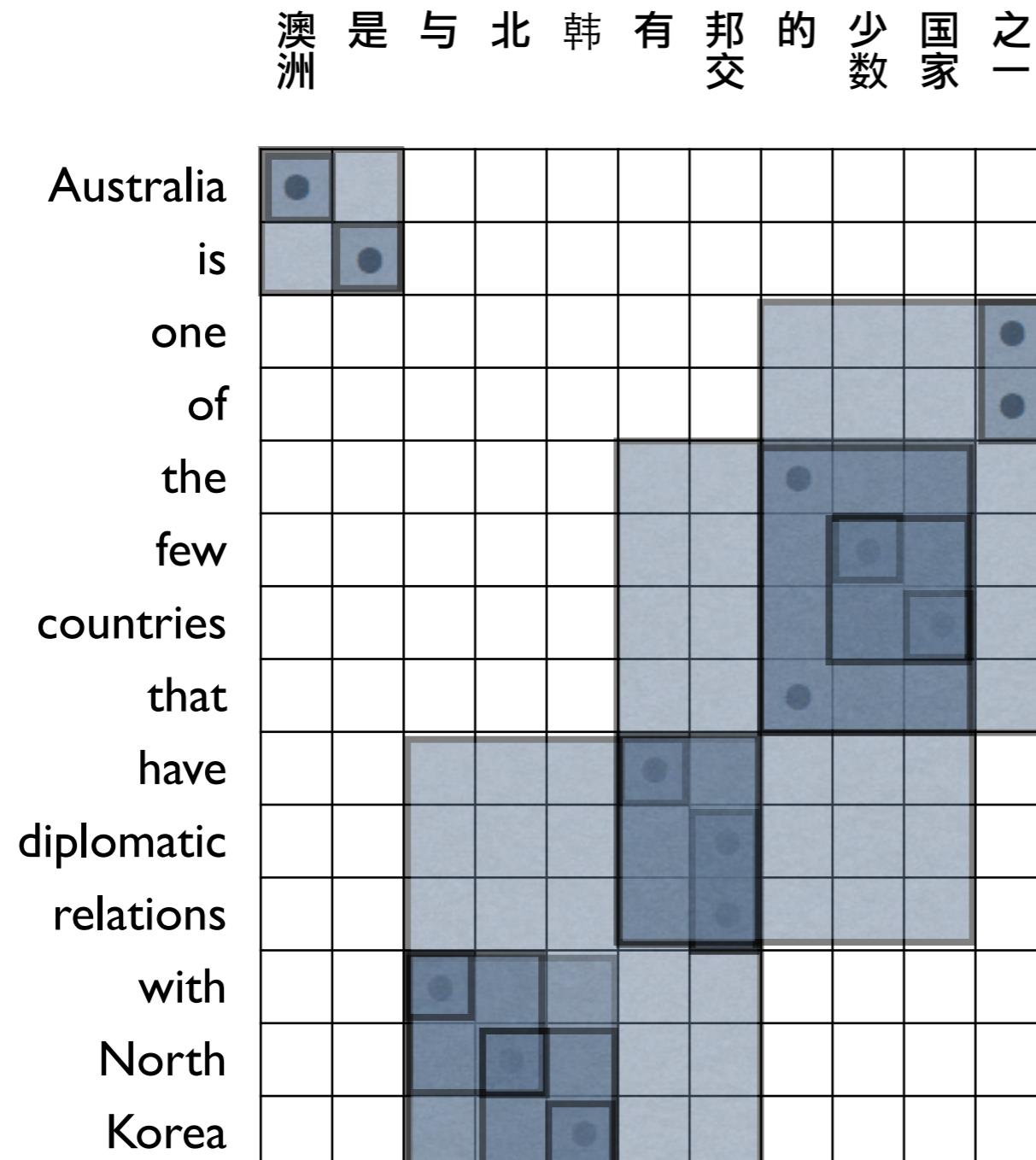
与北, with North

北韩, North Korea

的少数 国家, the few countries that

与北韩, with North Korea

# Extracting phrase pairs



澳洲, Australia

是, is

之一, one of

少数, few

国家, countries

有, have

邦交, diplomatic relations

与, with

北, North

韩, Korea

澳洲是, Australia is

少数 国家, few countries

有邦交, have diplomatic relations

与北, with North

北韩, North Korea

的少数 国家, the few countries that

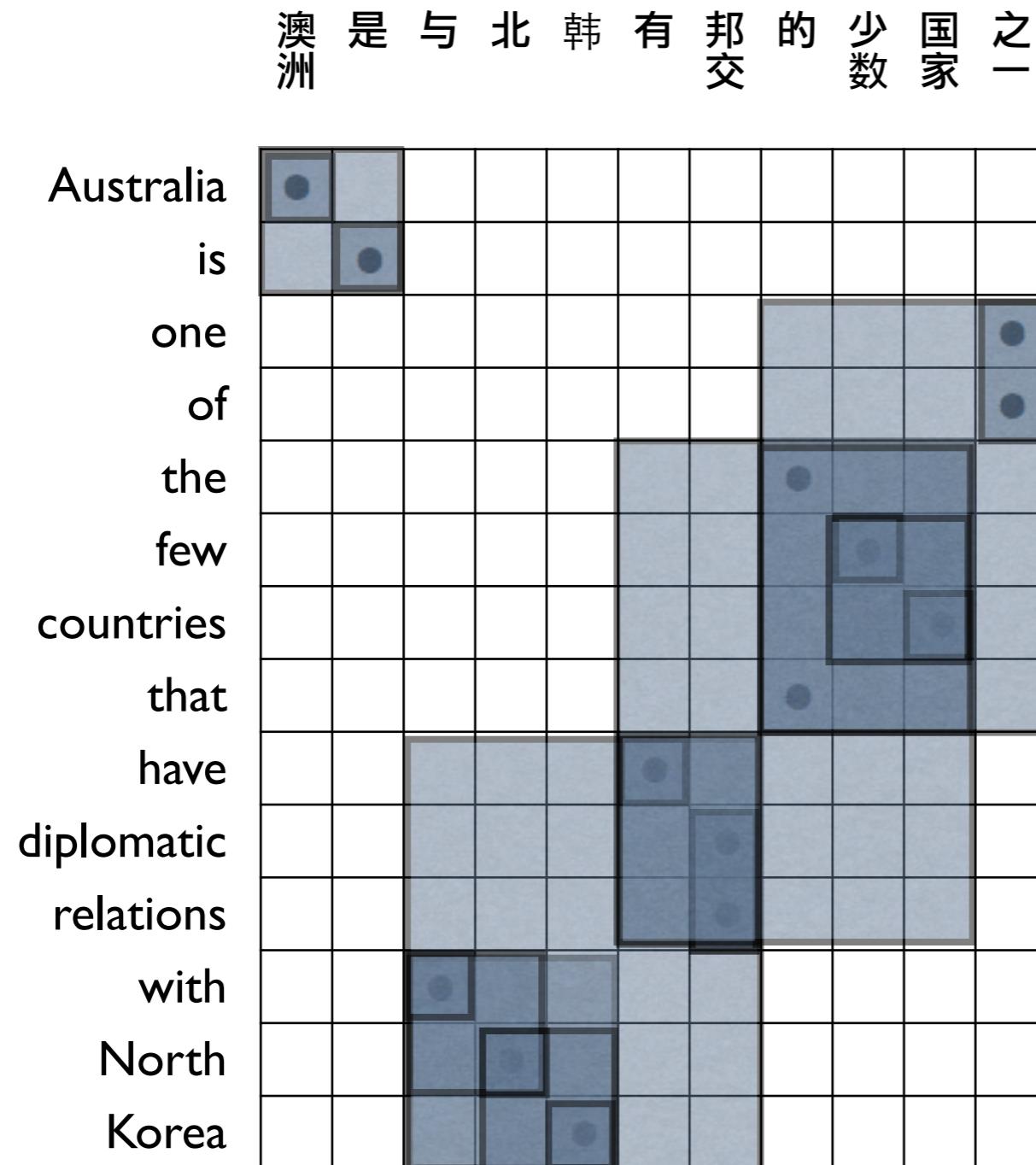
与北韩, with North Korea

之一的少数 国家, one of the the few  
countries that

与北韩 有邦交, have diplomatic  
relations with North Korea

有邦交 的少数 国家, the few countries  
that have diplomatic relations

# Extracting phrase pairs



澳洲, Australia  
是, is  
少数, few  
国家, countries  
有, have  
邦交, diplomatic relations  
与, with  
北, North  
韩, Korea  
  
少数 国家, few countries  
  
北韩, North Korea  
  
与北韩, with North Korea  
  
与北韩 有邦交, have diplomatic  
relations with North Korea

# Why does it hurt to limit to constituents?

- Massively **reduces the inventory** of phrases that can be used as translation units
- Eliminates **non-constituent phrases**, many of which are quite useful
  - *there are*
  - *note that*
  - *according to*

# So, what should we do?

- Drop **syntax** from statistical machine translation, since syntax is a bad fit for the data
- Abandon conventional English syntax and move towards **more robust grammars** that adapt to the parallel training corpus
- Maintain English syntax but **design different syntactic models**

# Synchronous Context Free Grammars

- A common way of representing syntax in NLP is through **context free grammars**
- **Synchronous** context free grammars generate pairs of corresponding strings
- Can be used to describe **translation** and **re-ordering** between languages
- SCFGs **translate sentences by parsing** them

# Example SCFG for Urdu

	Urdu	English
$S \rightarrow$	$NP\textcircled{1} VP\textcircled{2}$	$NP\textcircled{1} VP\textcircled{2}$
$VP \rightarrow$	$PP\textcircled{1} VP\textcircled{2}$	$VP\textcircled{2} PP\textcircled{1}$
$VP \rightarrow$	$V\textcircled{1} AUX\textcircled{2}$	$AUX\textcircled{2} V\textcircled{1}$
$PP \rightarrow$	$NP\textcircled{1} P\textcircled{2}$	$P\textcircled{2} NP\textcircled{1}$
$NP \rightarrow$	<i>hamd ansary</i>	<i>Hamid Ansari</i>
$NP \rightarrow$	<i>na}b sdr</i>	<i>Vice President</i>
$V \rightarrow$	<i>namzd</i>	<i>nominated</i>
$P \rightarrow$	<i>kylye</i>	<i>for</i>
$AUX \rightarrow$	<i>taa</i>	<i>was</i>

*hamd ansary*      *na}b sdr*      *kylye*      *namzd*      *taa*

NP1  
△

*hamd ansary*

*na}b sdr*

*kylye*

*namzd*

*taa*

NP1  
△

*Hamid Ansari*

**NP1**



*hamd ansary*

**NP2**



*na}b sdr*

*kylye*

*namzd*

*taa*

**NP1**

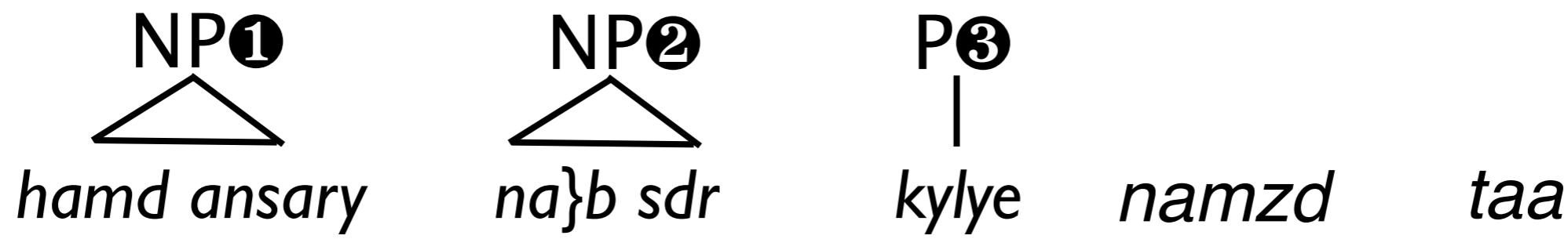


*Hamid Ansari*

**NP2**

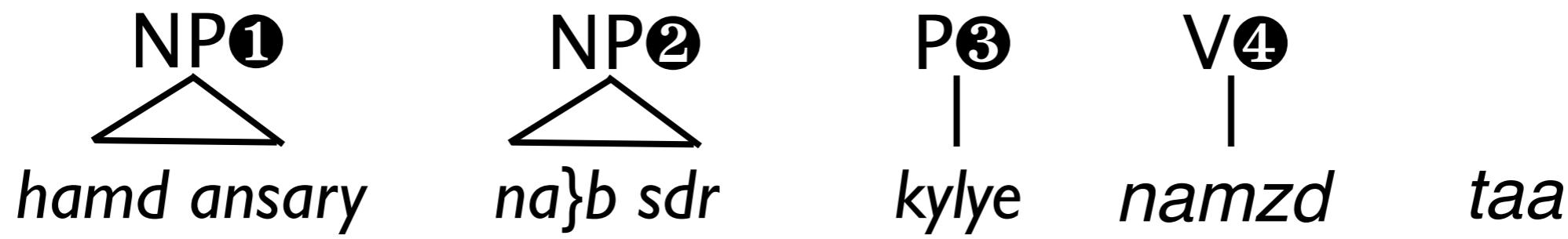


*Vice President*



NP1  
NP2  
P3

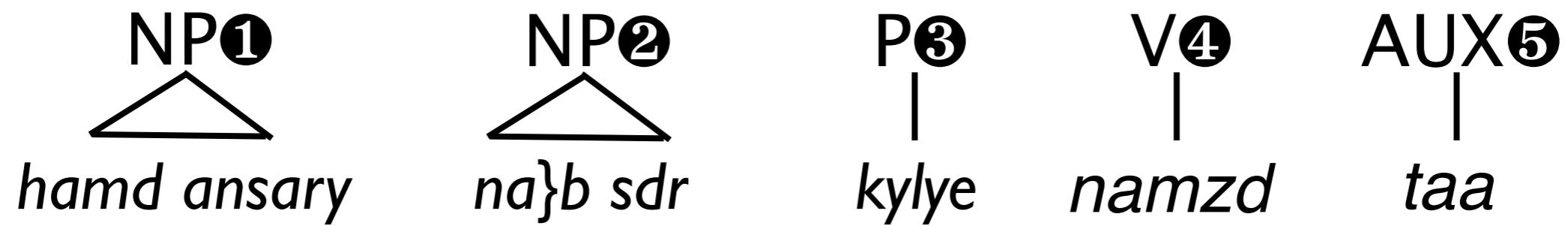
Hamid Ansari      Vice President      for



A morphological tree diagram showing the structure of the sentence 'Hamid Ansari was nominated for Vice President'. The diagram consists of five nodes arranged horizontally. From left to right: 1. A node labeled 'NP1' with a triangle underneath it, corresponding to the word 'Hamid'. 2. A node labeled 'NP2' with a triangle underneath it, corresponding to the word 'Ansari'. 3. A node labeled 'P3' with a vertical line underneath it, corresponding to the word 'was'. 4. A node labeled 'V4' with a vertical line underneath it, corresponding to the word 'nominated'. 5. A node labeled 'NP1' with a triangle underneath it, corresponding to the phrase 'for Vice President'.

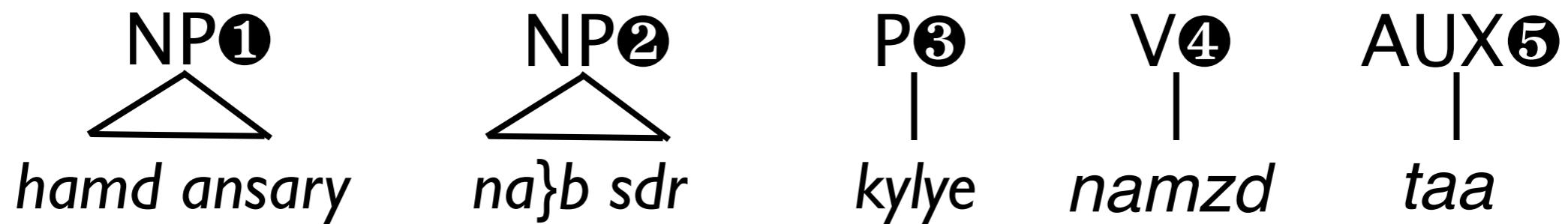
NP1  
NP2  
P3  
V4

Hamid Ansari   Vice President   for   nominated



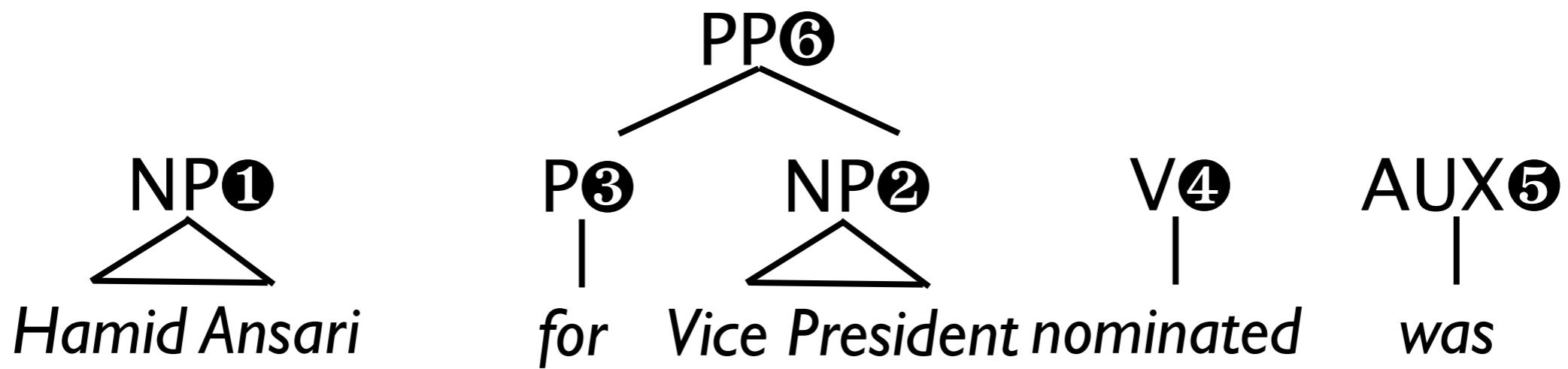
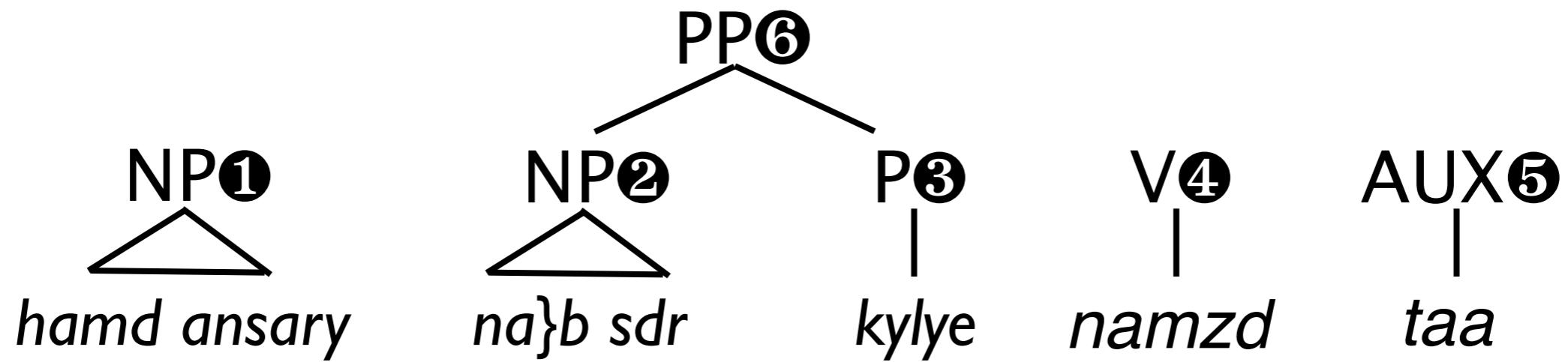
NP1  
NP2  
P3  
V4  
AUX5

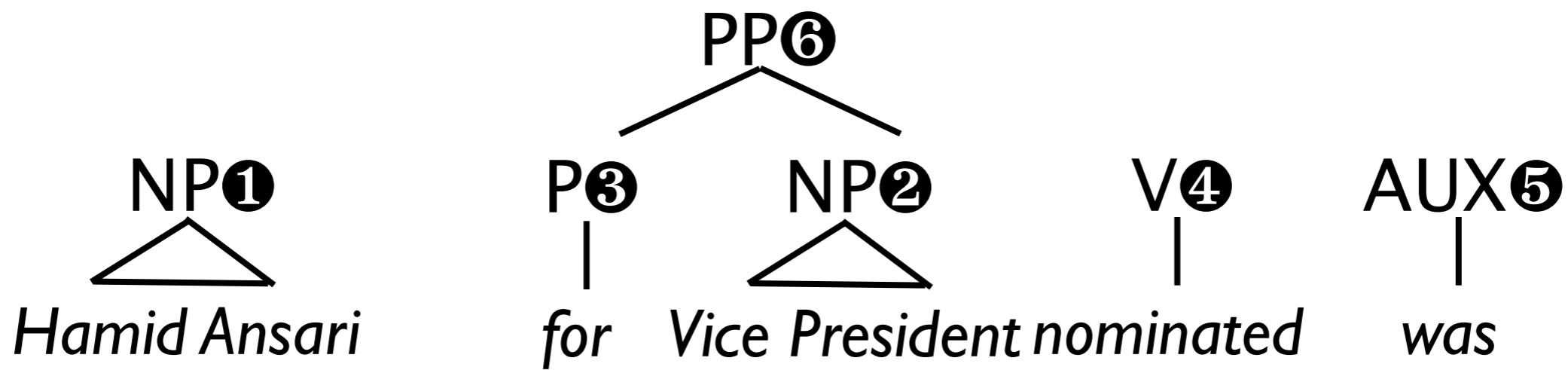
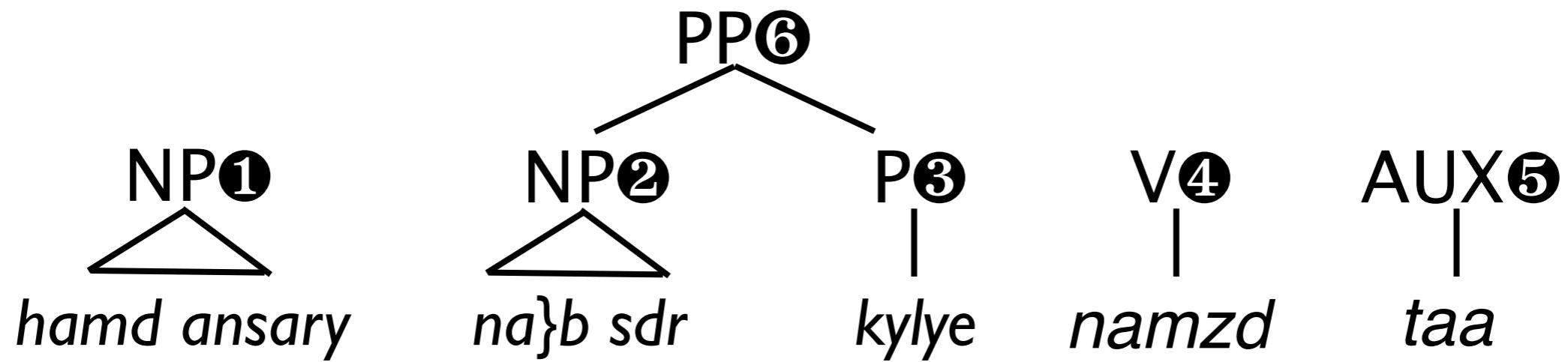
Hamid Ansari      Vice President  
                      for      nominated      was

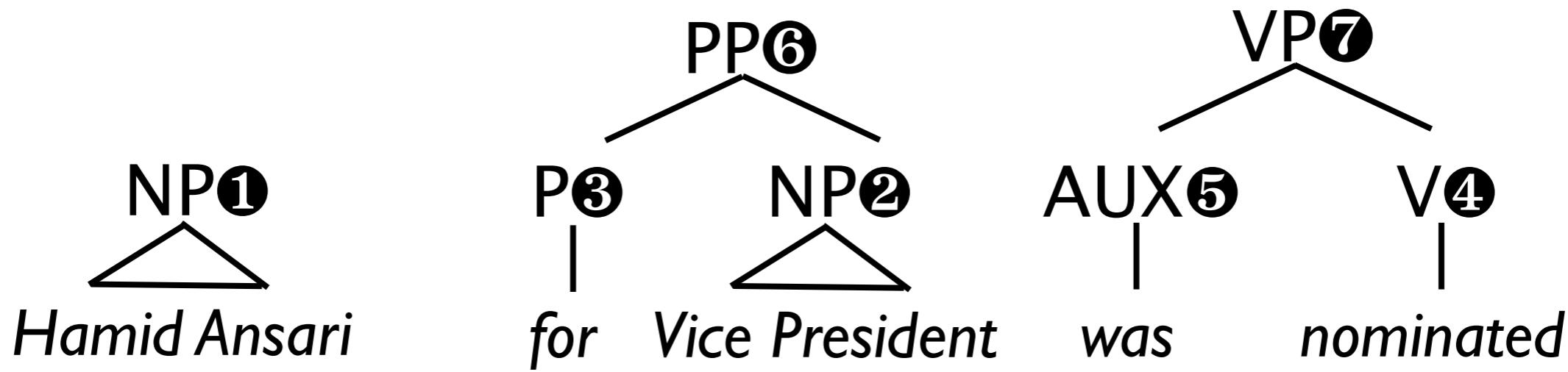
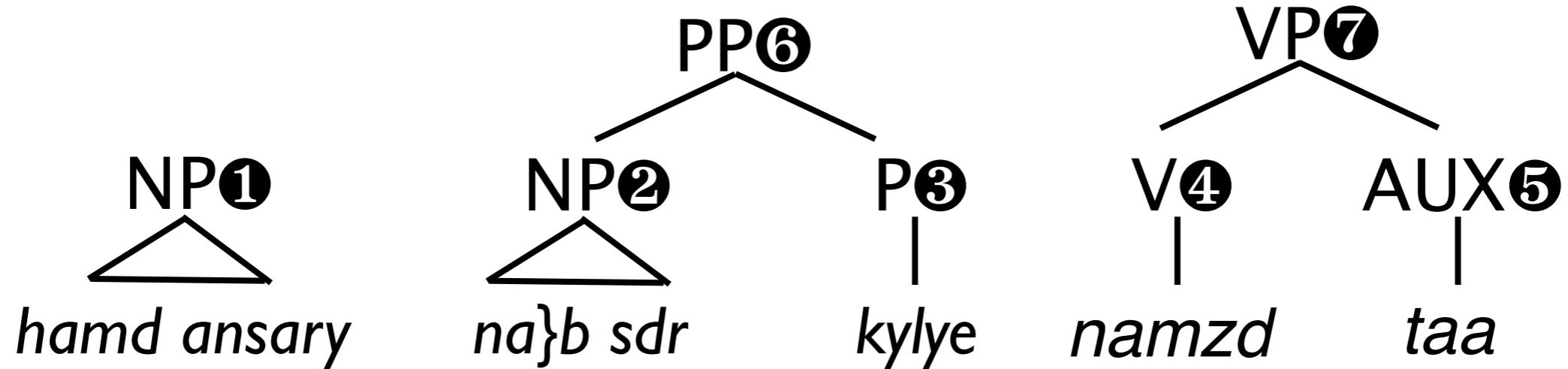


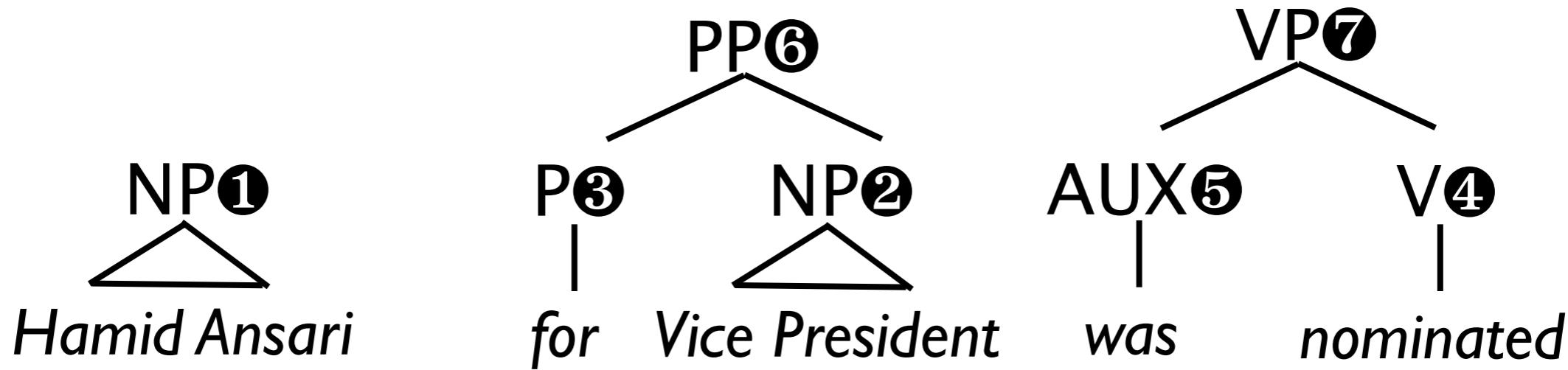
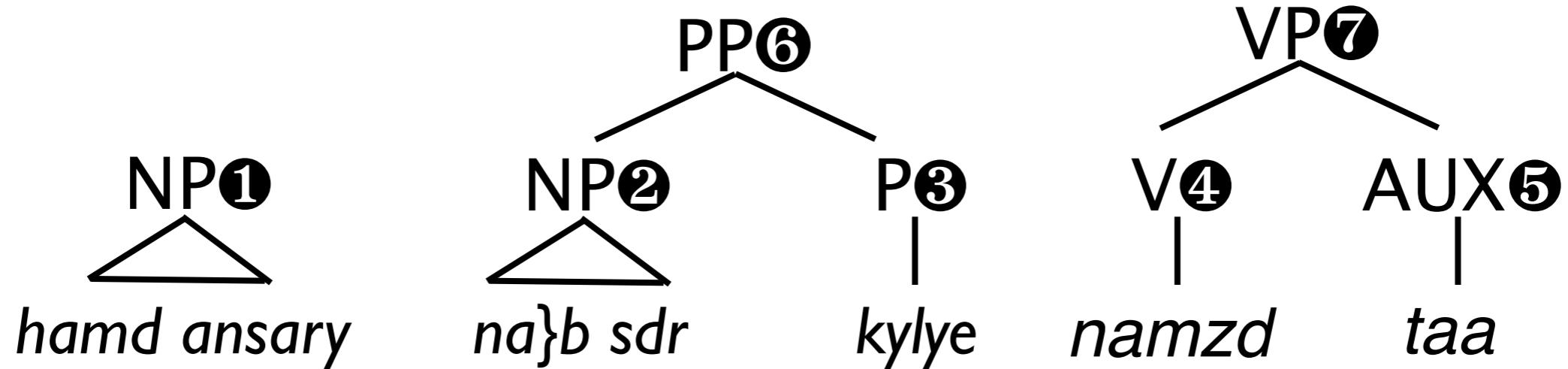
NP1  
NP2  
P3  
V4  
AUX5

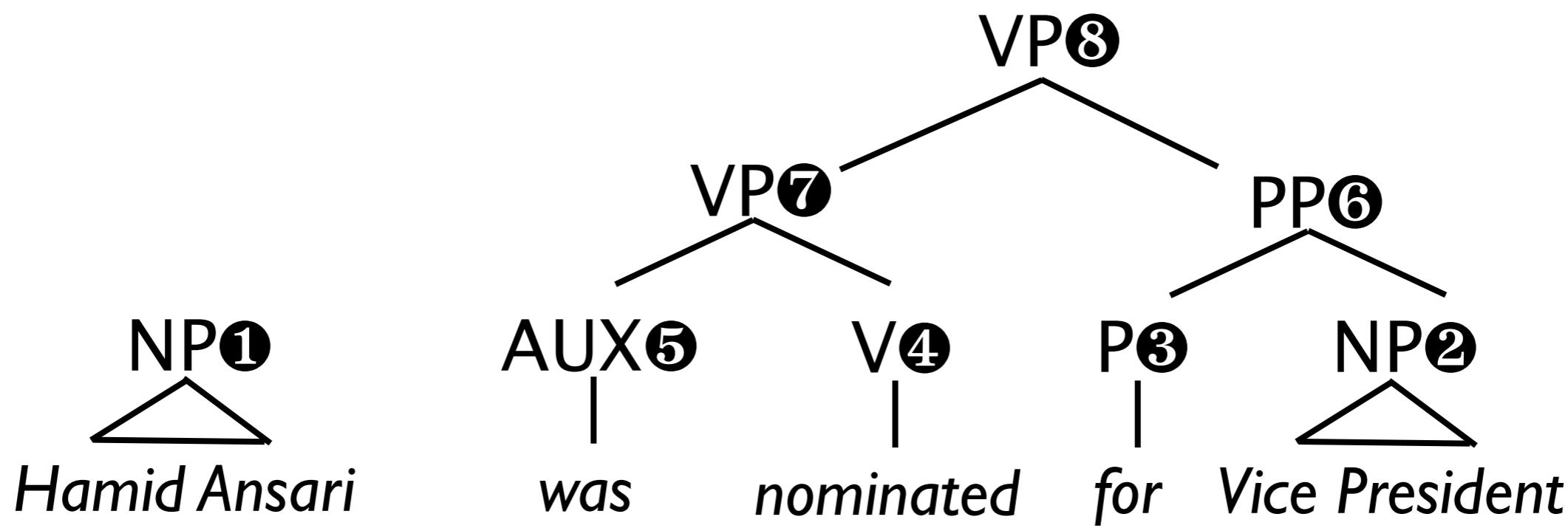
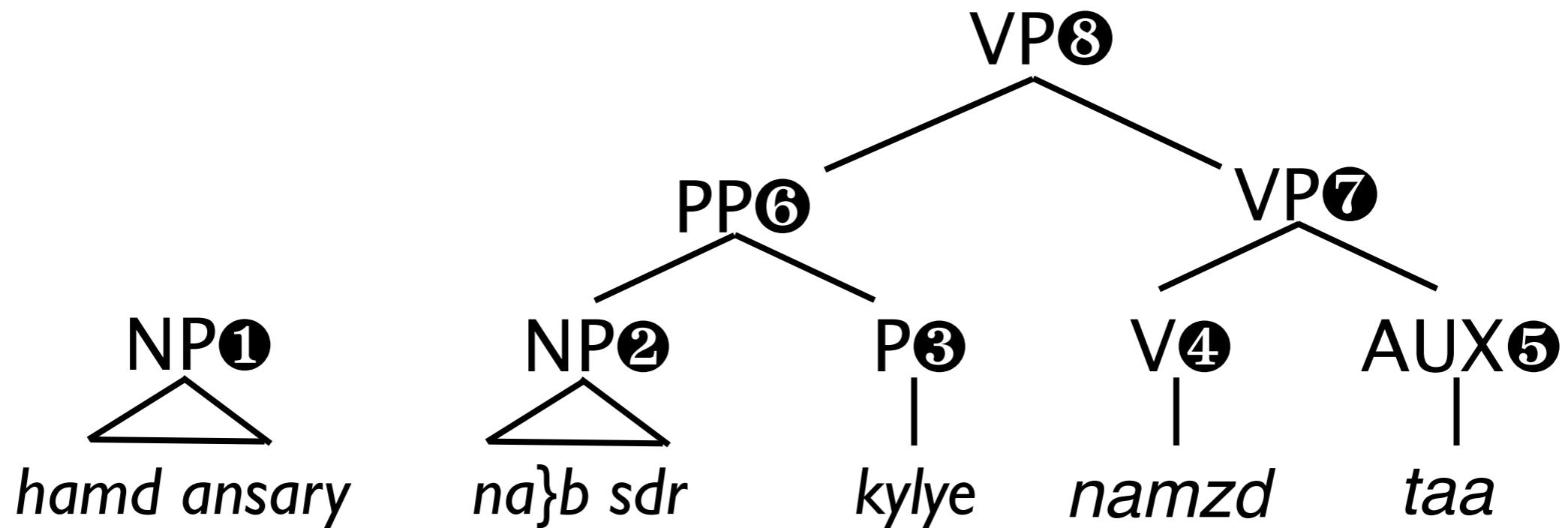
Hamid Ansari      Vice President  
                      for      nominated      was

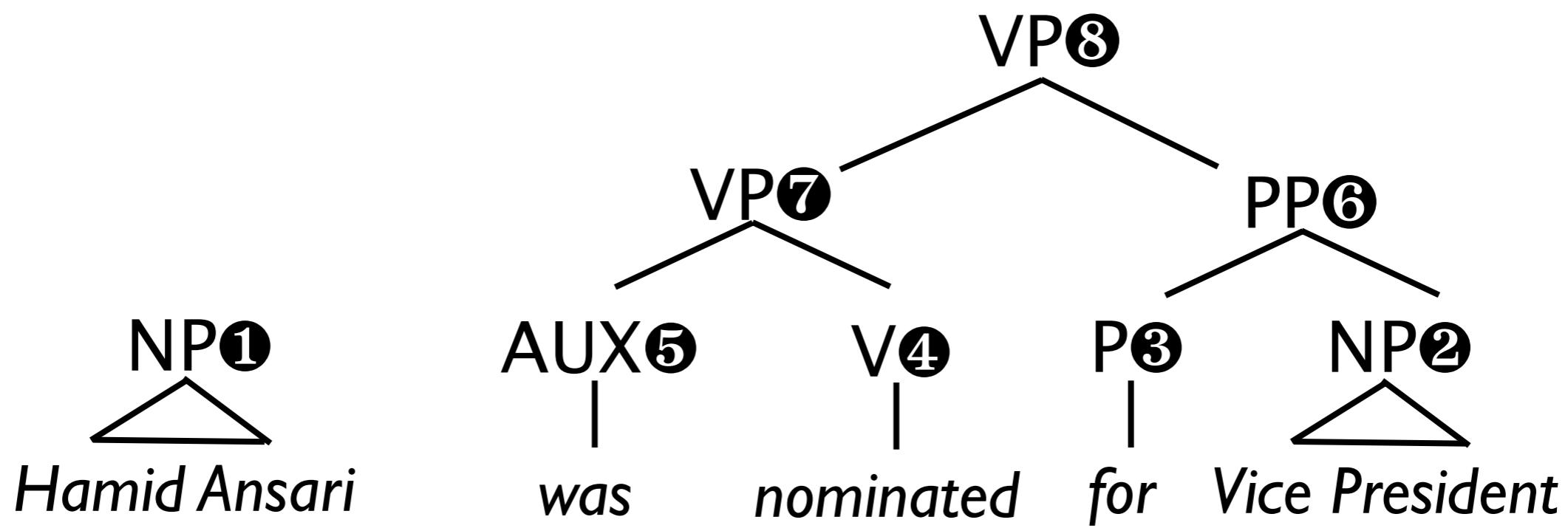
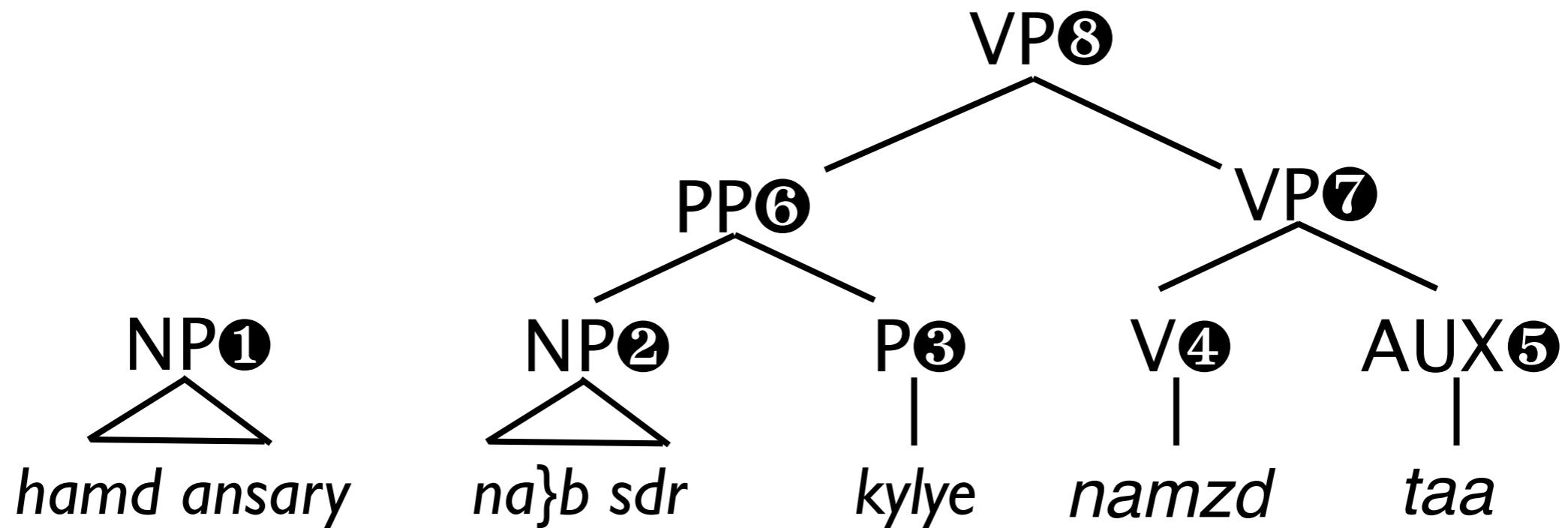


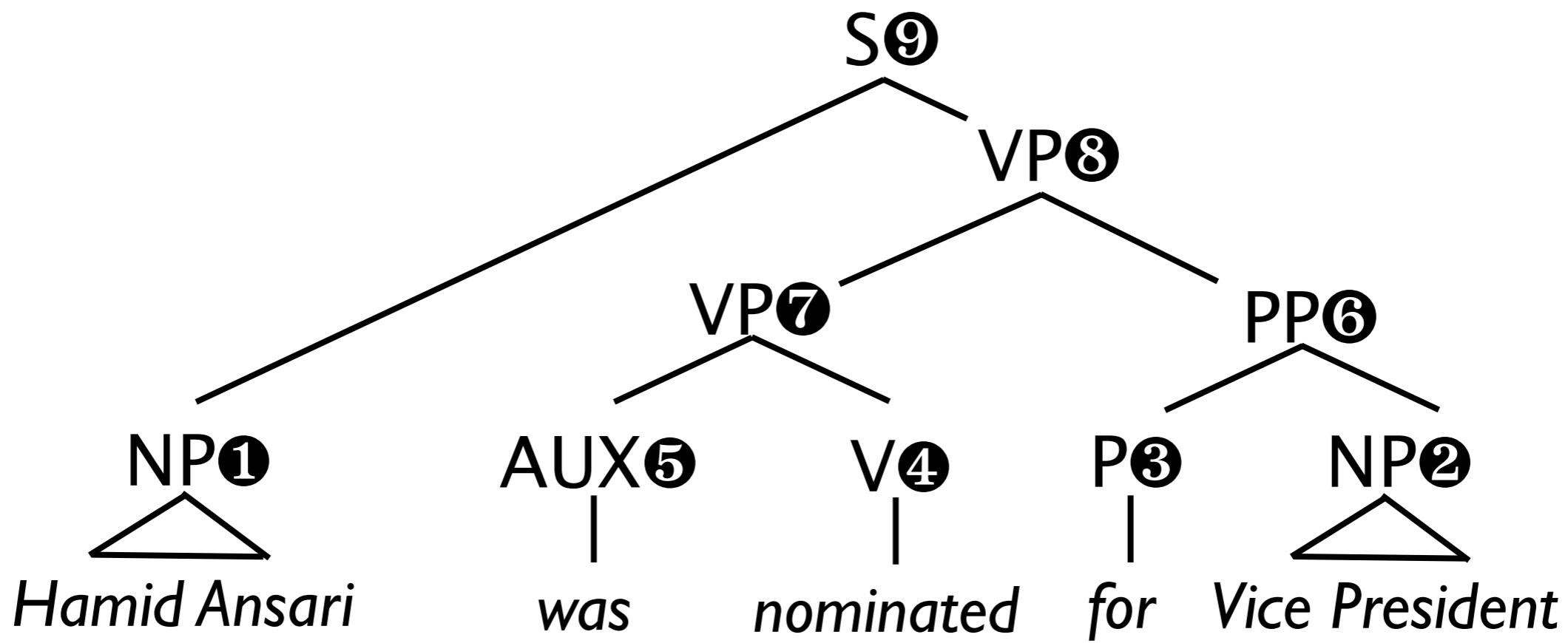
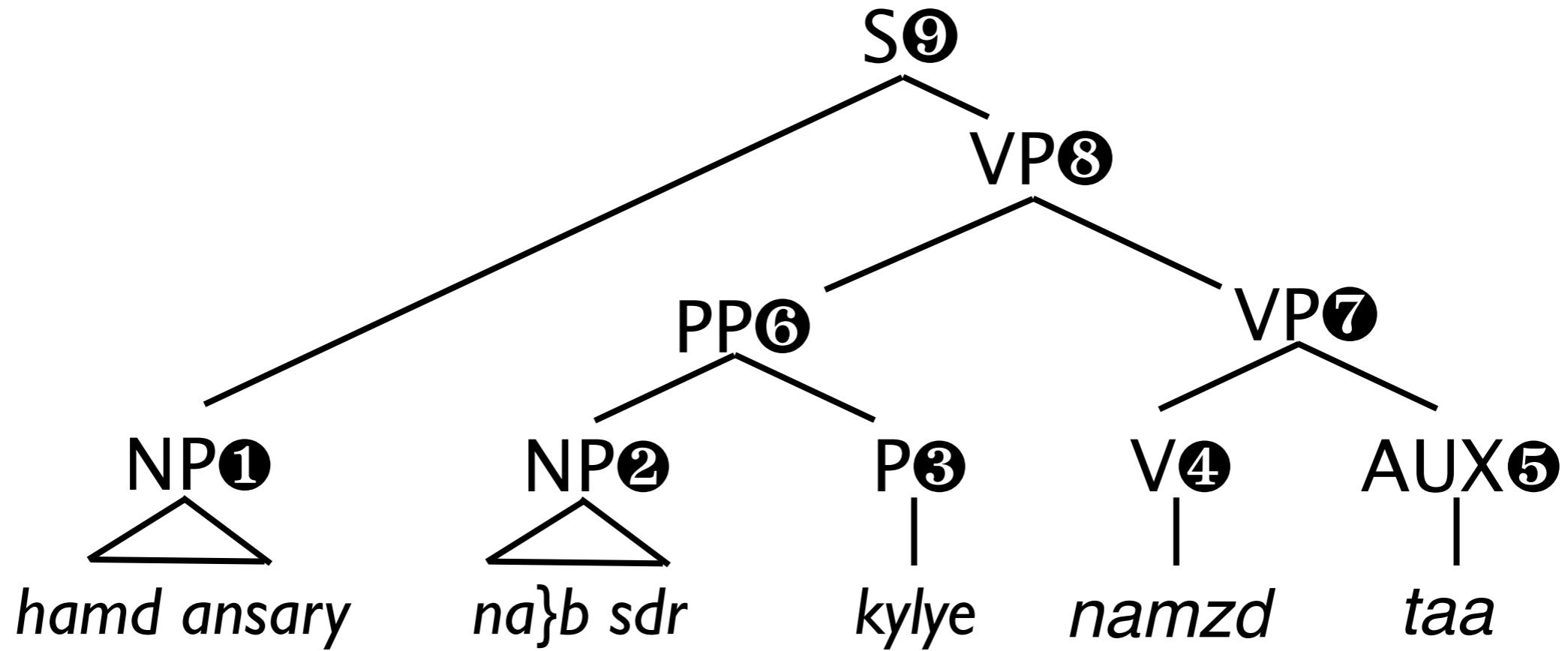










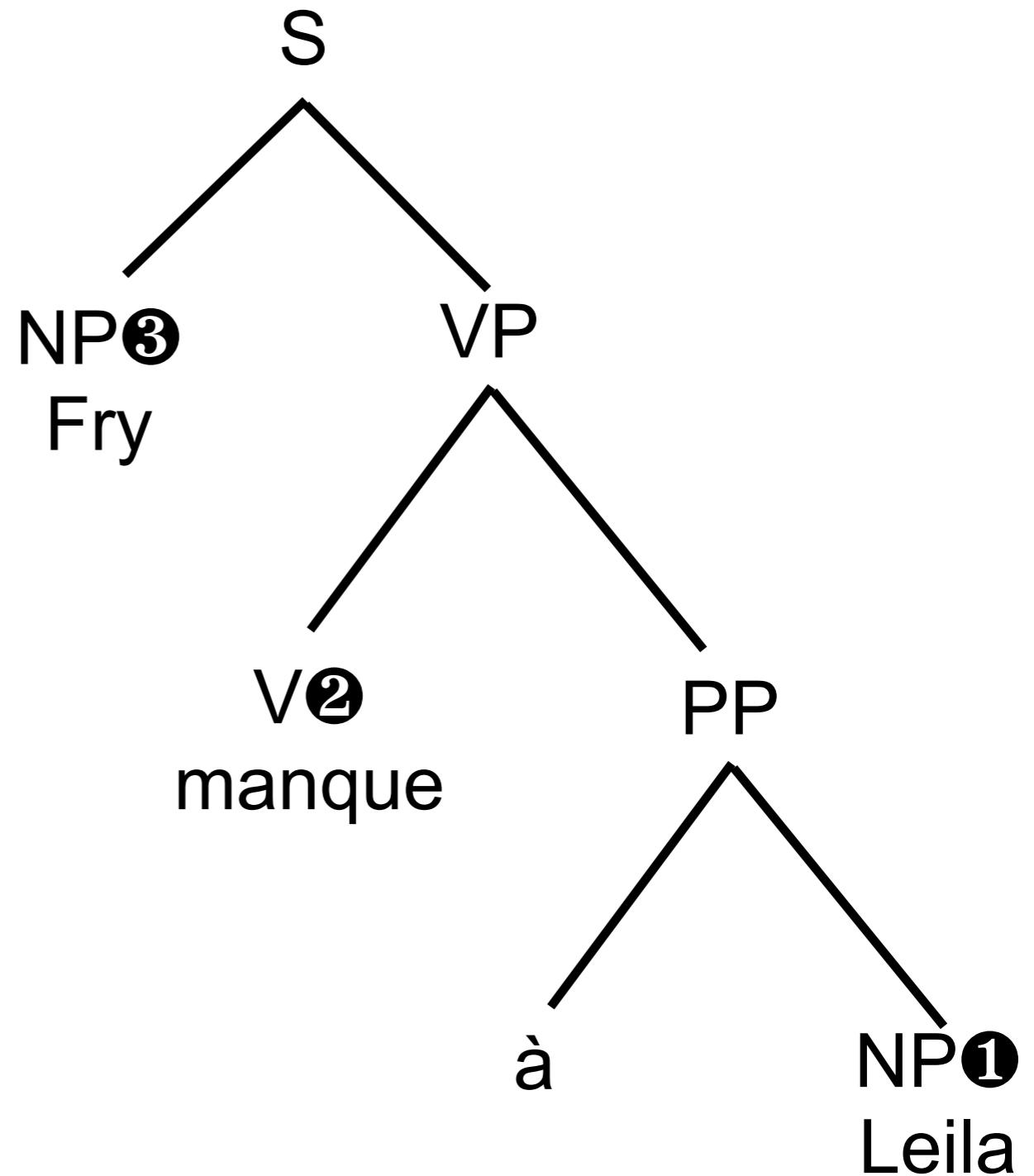
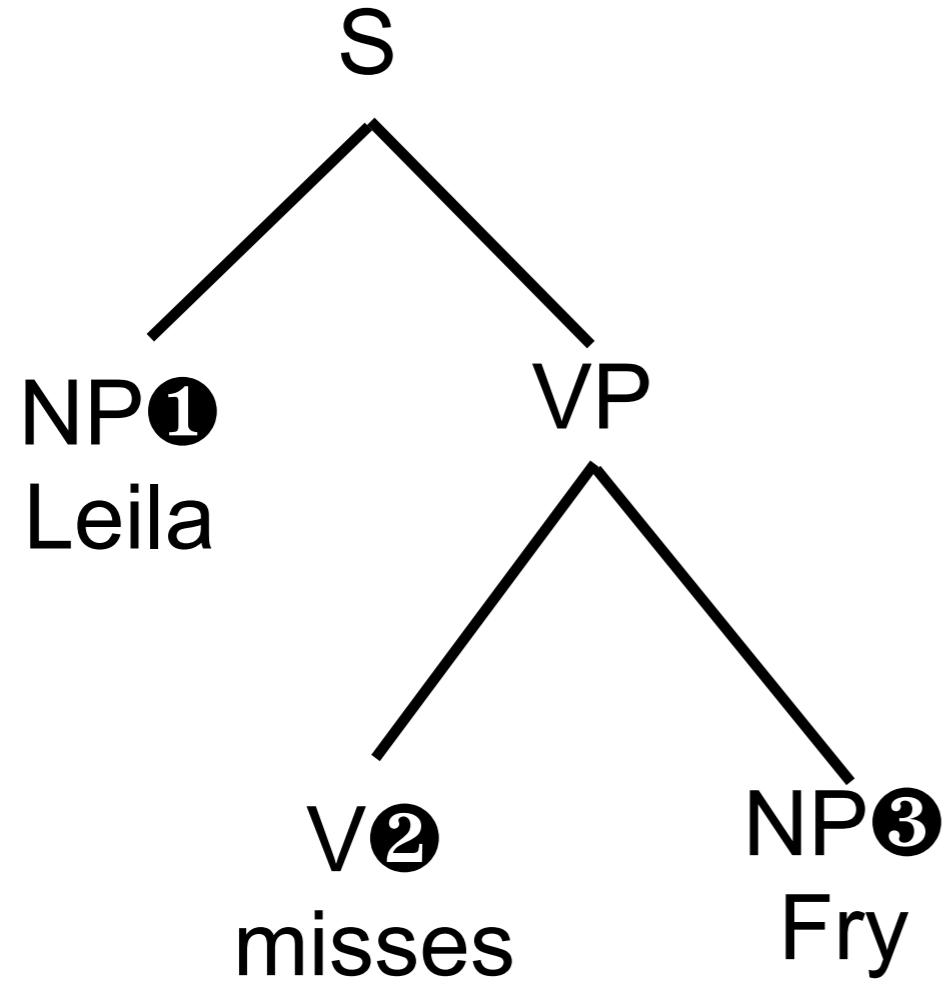


# Discussion: Do you like SCFG?

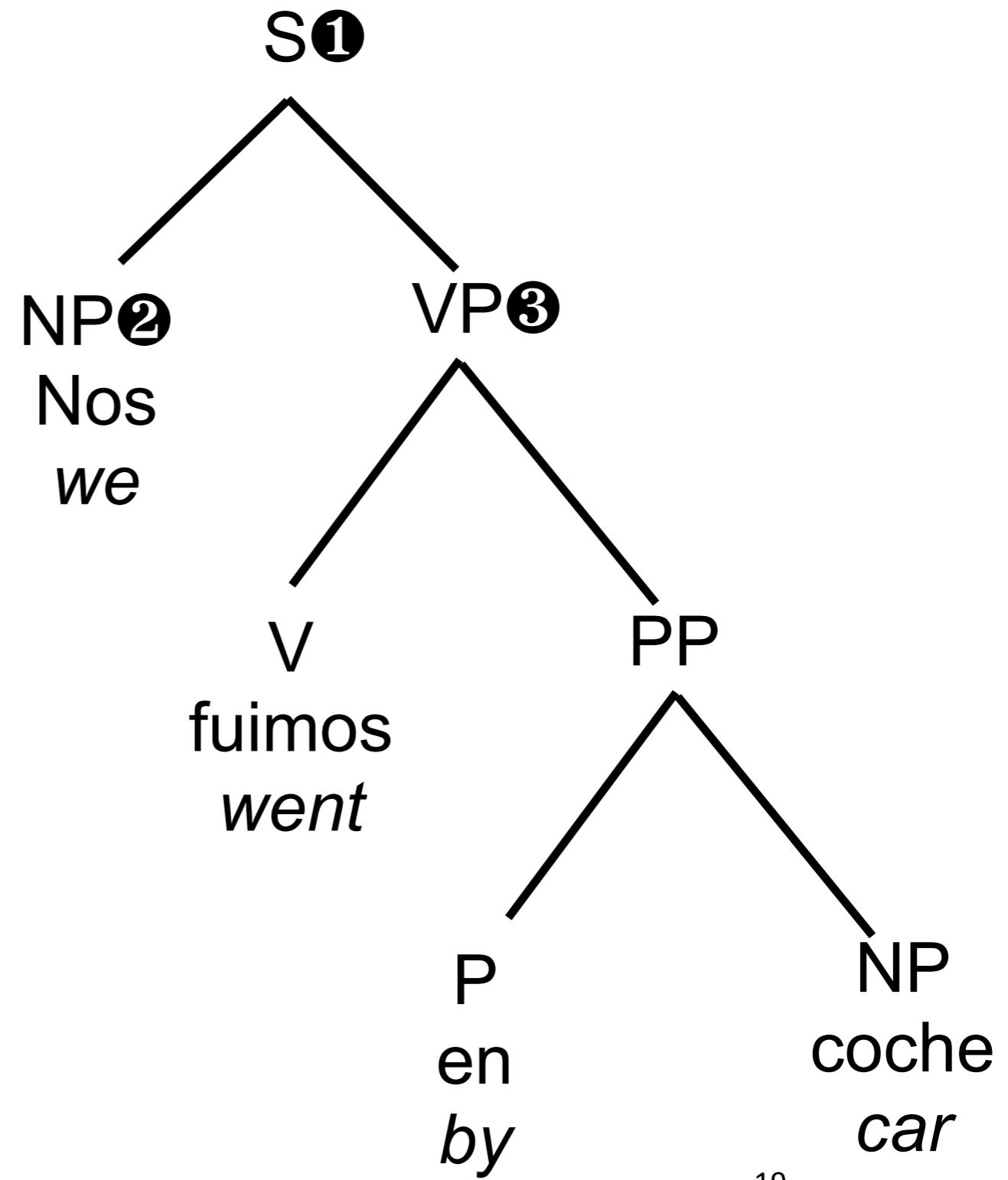
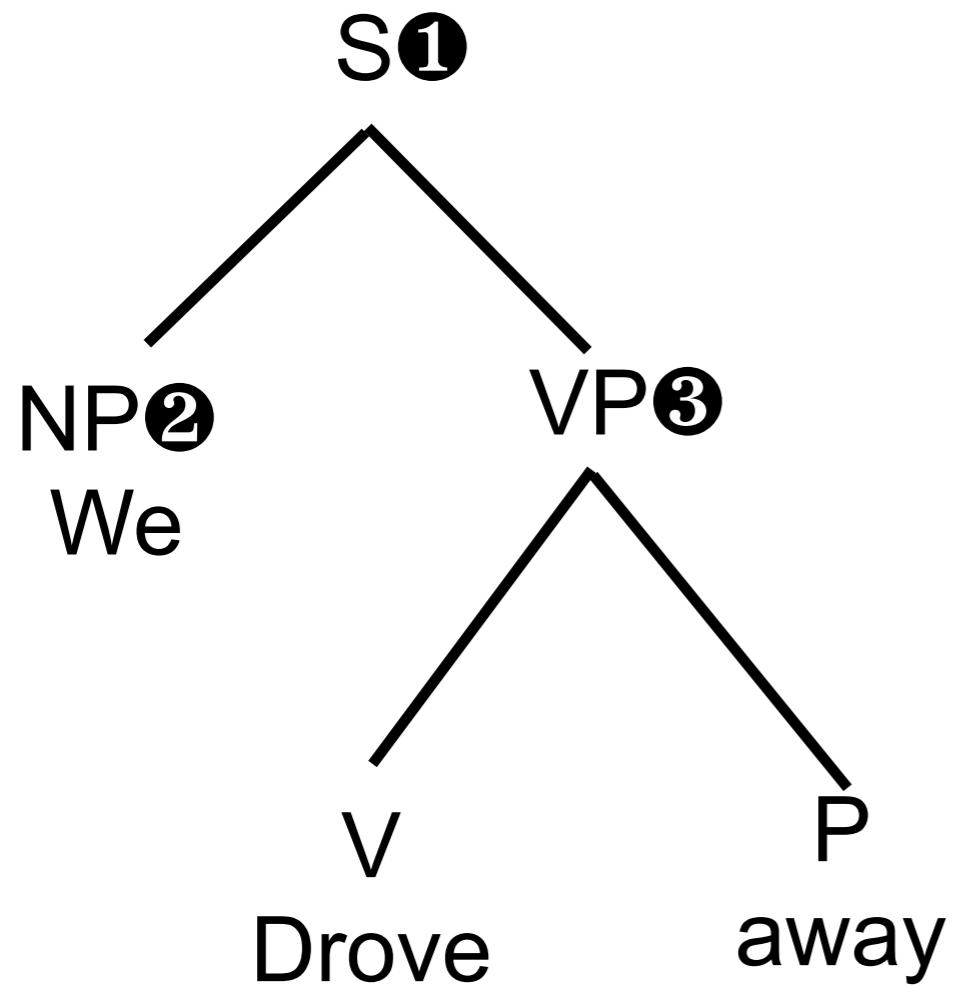
- In what ways are SCFGs better for describing reordering than what we saw before?
- Is this a good model of how languages relate?
- What do you think of the synchronous requirement?

(Discuss with your neighbor)

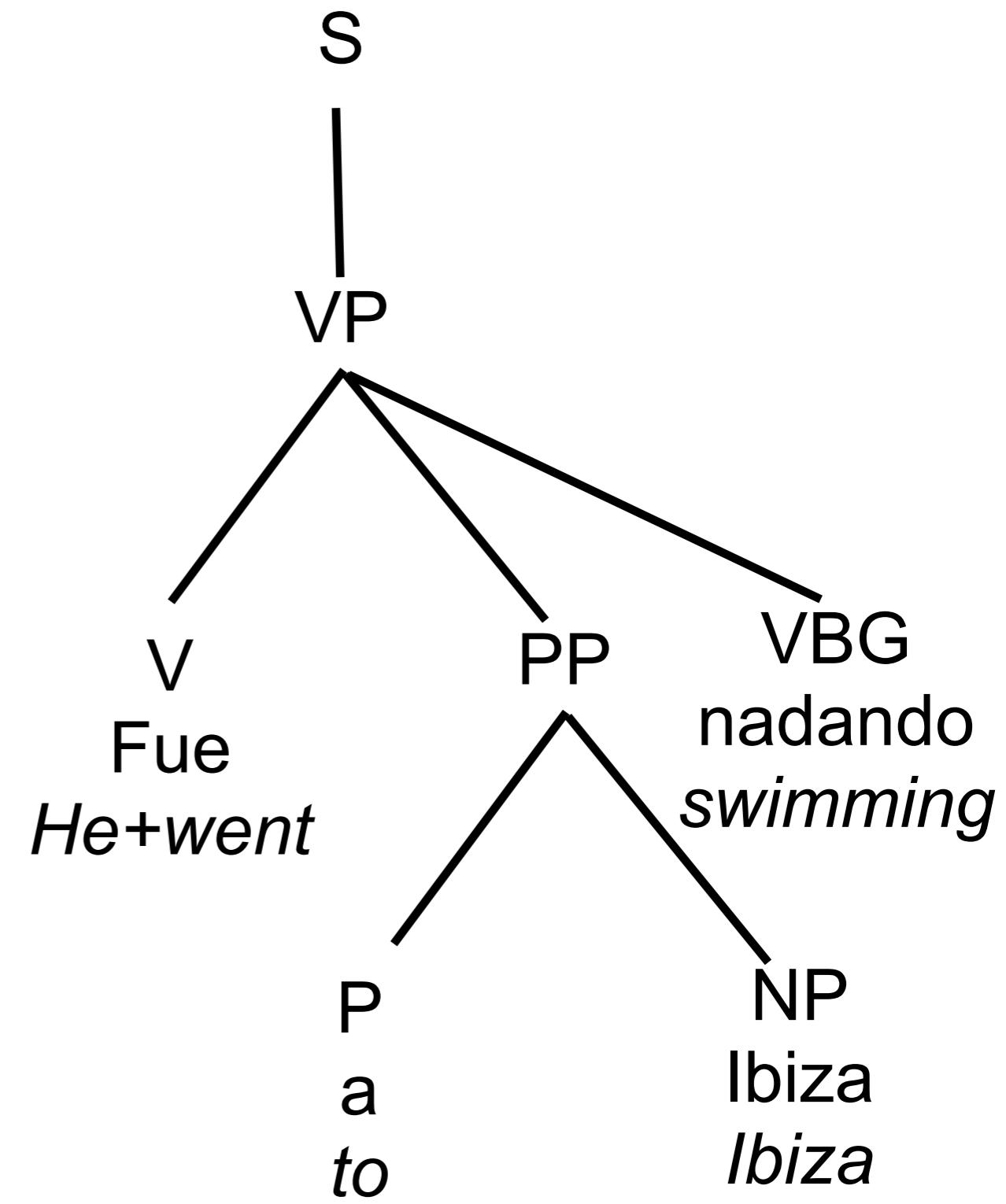
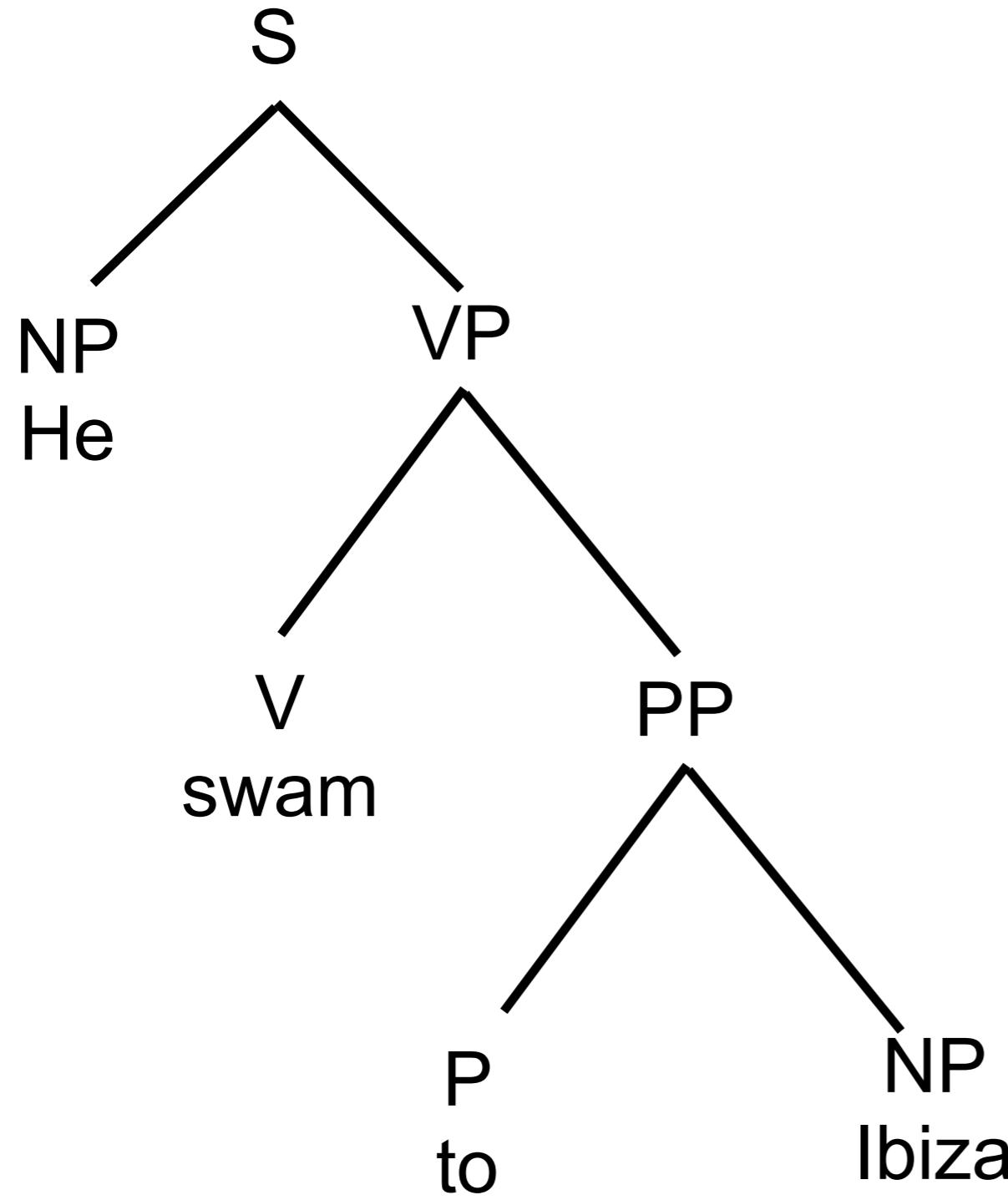
# Sometimes languages are mismatched



# Spanish motion verb



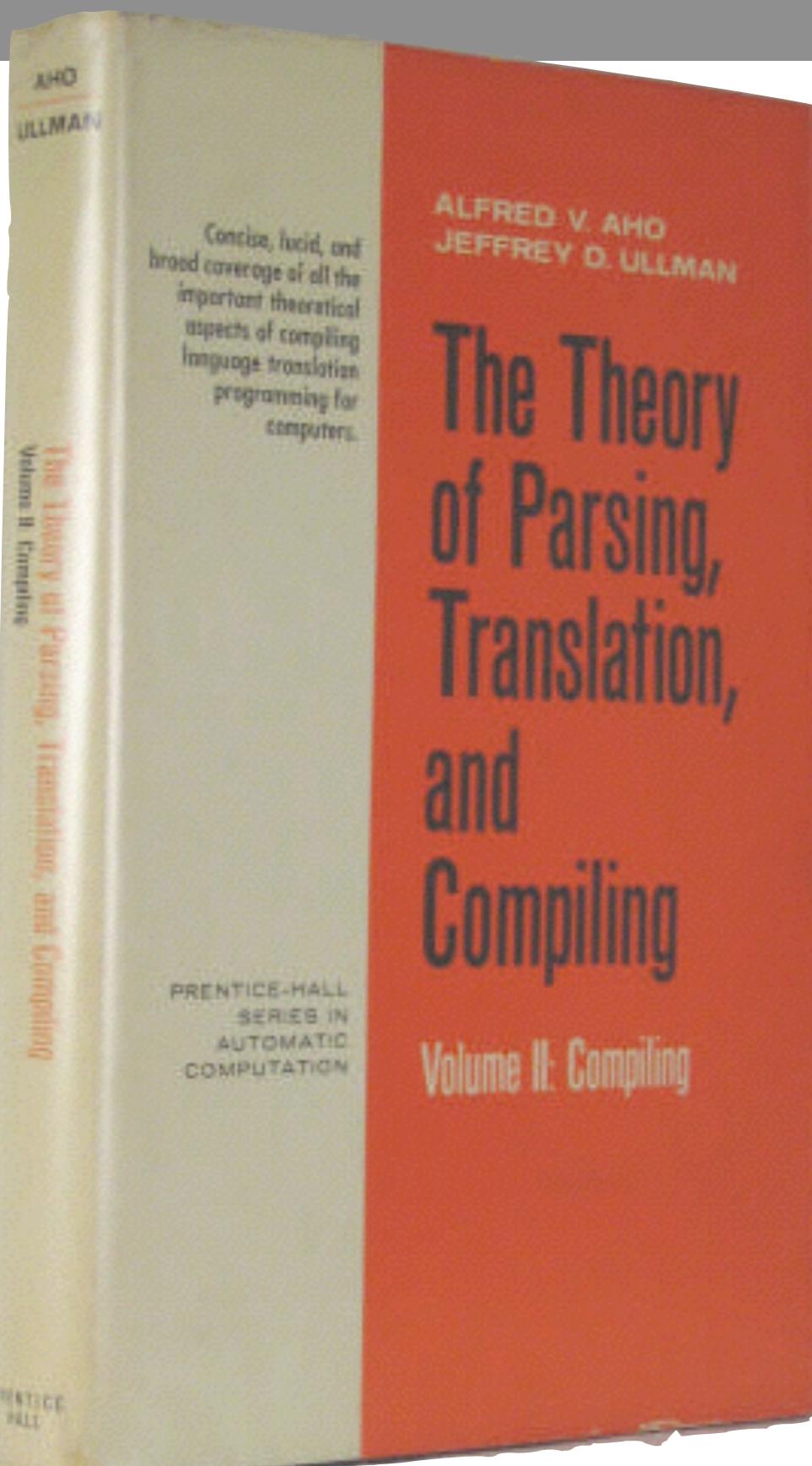
# Spanish motion verb, pro-drop



# We are going to use them anyway

- SCFGs are **mismatched** with some linguistic phenomena
- But they have nice **formal properties** and **well-defined algorithms**

# Formal definition of SCFGs



- Aho and Ullman worked all of this out in the '60s and '70s
- Compiler theory

# Formal definition of SCFGs

- A synchronous context free grammar is formally defined by a tuple

$$G = \langle N, T_S, T_T, R, S \rangle$$

- Where

# Formal definition of SCFGs

An synchronous context free grammar is defined by a tuple

$$G = \langle N, T_S, T_T, R, S \rangle$$

- Where
  - $N$  is a shared set of non-terminal symbols

# Formal definition of SCFGs

*hamd ansary, na}b sdr,  
namzd, kylye, taa*

A synchronous context free grammar is defined by a tuple

S, NP, VP, PP,  
P, V, AUX

$$G = \langle N, T_S, T_T, R, S \rangle$$

- Where
  - $N$  is a shared set of non-terminal symbols
  - $T_S$  is the set of source language terminals

# Formal definition of SCFGs

*hamd ansary, na}b sdr,  
namzd, kylye, taa*

*for, Hamid Ansari, nominated,  
Vice President, was*

A synchronous context free grammar is defined by a tuple

S, NP, VP, PP,  
P, V, AUX

$$G = \langle N, T_S, T_T, R, S \rangle$$

- Where
  - $N$  is a shared set of non-terminal symbols
  - $T_S$  is the set of source language terminals
  - $T_T$  is the set of target language terminals

# Formal definition of SCFGs

*hamd ansary, na}b sdr,  
namzd, kylye, taa*

*for, Hamid Ansari, nominated,  
Vice President, was*

A synchronous context free grammar is  
defined by a tuple

$$G = \langle N, T_S, T_T, R, S \rangle$$

- Where
  - $N$  is a shared set of non-terminal symbols
  - $T_S$  is the set of source language terminals
  - $T_T$  is the set of target language terminals
  - $R$  is a set of production rules

# Formal definition of SCFGs

*hamd ansary, na}b sdr,  
namzd, kylye, taa*

*for, Hamid Ansari, nominated,  
Vice President, was*

A synchronous context free grammar is defined by a tuple

S, NP, VP, PP,  
P, V, AUX

$$G = \langle N, T_S, T_T, R, S \rangle$$

S

- Where
  - $N$  is a shared set of non-terminal symbols
  - $T_S$  is the set of source language terminals
  - $T_T$  is the set of target language terminals
  - $R$  is a set of production rules
  - $S \in N$ , designated as the goal state

# Formal definition of SCFGs

- Each production rule has the form

$$X \rightarrow \langle \alpha, \beta, \sim, w \rangle$$

- Where
  - $X \in N$
  - $\alpha \in (N \cup T_S)^*$
  - $\beta \in (N \cup T_T)^*$
  - $\sim$  is a one-to-one correspondence between the non terminals in  $\alpha$  and  $\beta$
  - $w$  is a weight assigned to the rule

# Algorithms for SCFGs

- Translation with SCFGs is done via parsing
- How do we write an algorithm for parsing?
- One way to do it is as a deductive proof system

# The CKY Parsing Algorithm

Axioms	$\frac{}{A \rightarrow \alpha}$	for all $(A \rightarrow \alpha) \in R$
Inference rules	$\frac{A}{[A, i, i+1]}$ $\frac{[B, i, j] \ [C, j, k] \ A \rightarrow BC}{[A, i, k]}$	
Goal	$[S, 0, n]$	

Axioms		Inference rule used	Goal
$S \rightarrow$	$NP \ VP$		
$VP \rightarrow$	$PP \ VP$		
$VP \rightarrow$	$V \ AUX$		$[S, 0, 5]$
$PP \rightarrow$	$NP \ P$		
$NP \rightarrow$	<i>hamd ansary</i>		
$NP \rightarrow$	<i>na}b sdr</i>		
$V \rightarrow$	<i>namzd</i>		
$P \rightarrow$	<i>kylye</i>		
$AUX \rightarrow$	<i>taa</i>		

Axioms		Inference rule used	Goal
$S \rightarrow$	$NP \ VP$		
$VP \rightarrow$	$PP \ VP$		
$VP \rightarrow$	$V \ AUX$		$[S, 0, 5]$
$PP \rightarrow$	$NP \ P$		
$NP \rightarrow$	<i>hamd ansary</i>		
$NP \rightarrow$	<i>na}b sdr</i>		
$V \rightarrow$	<i>namzd</i>		
$P \rightarrow$	<i>kylye</i>		
$AUX \rightarrow$	<i>taa</i>		

$0 \ hamd \ ansary \quad 1 \ na}b \ sdr \quad 2 \ kylye \quad 3 \ namzd \quad 4 \ taa \quad 5$

<b>Axioms</b>	S → NP VP	<b>Inference rule used</b>	<b>Goal</b>
VP →	PP VP		
VP →	V AUX	<u>NP → hamd ansary<sub>1</sub></u>	[S, 0, 5]
PP →	NP P		
NP →	<i>hamd ansary</i>	[NP, 0, 1]	
NP →	<i>na}b sdr</i>		
V →	<i>namzd</i>		
P →	<i>kylye</i>		
AUX →	<i>taa</i>		

0 *hamd ansary*    1 *na}b sdr*    2 *kylye*    3 *namzd*    4 *taa*    5

Axioms		Inference rule used	Goal
$S \rightarrow NP VP$			
$VP \rightarrow PP VP$			
$VP \rightarrow V AUX$		<u><math>NP \rightarrow hamd ansary_1</math></u>	$[S, 0, 5]$
$PP \rightarrow NP P$			
$NP \rightarrow hamd ansary$			
$NP \rightarrow na}b sdr$		$[NP, 0, 1]$	
$V \rightarrow namzd$			
$P \rightarrow kylye$			
$AUX \rightarrow taa$			

0 *hamd ansary* 1 *na}b sdr* 2 *kylye* 3 *namzd* 4 *taa* 5

[NP, 0, 1]

Axioms		Inference rule used	Goal
$S \rightarrow$	$NP \ VP$		
$VP \rightarrow$	$PP \ VP$		
$VP \rightarrow$	$V \ AUX$		$[S, 0, 5]$
$PP \rightarrow$	$NP \ P$		
$NP \rightarrow$	<i>hamd ansary</i>		
$NP \rightarrow$	<i>na}b sdr</i>		
$V \rightarrow$	<i>namzd</i>		
$P \rightarrow$	<i>kylye</i>		
$AUX \rightarrow$	<i>taa</i>		

0    *hamd ansary*    1    *na}b sdr*    2    *kylye*    3    *namzd*    4    *taa*    5

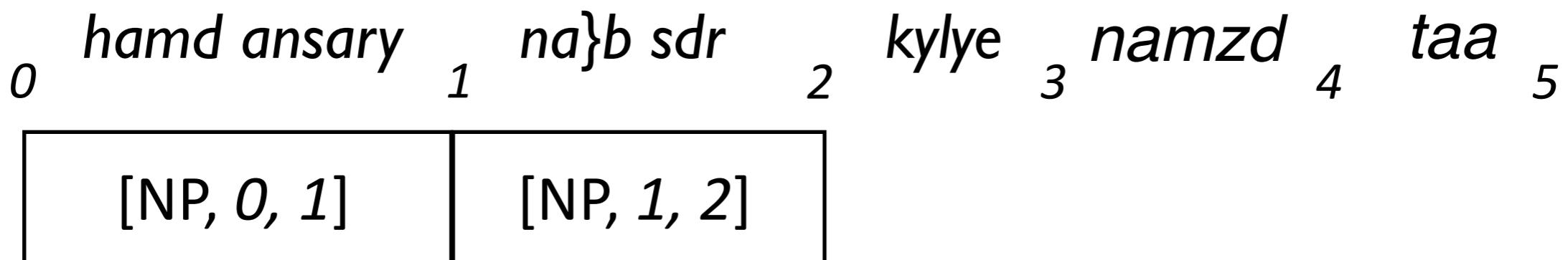
[NP, 0, 1]

Axioms		Inference rule used	Goal
$S \rightarrow$	$NP \ VP$		
$VP \rightarrow$	$PP \ VP$		
$VP \rightarrow$	$V \ AUX$	$NP \rightarrow na\}b \ sdr_2$	$[S, 0, 5]$
$PP \rightarrow$	$NP \ P$		
$NP \rightarrow$	<i>hamd ansary</i>		
$NP \rightarrow$	<i>na\}b sdr</i>		
$V \rightarrow$	<i>namzd</i>		
$P \rightarrow$	<i>kylye</i>		
$AUX \rightarrow$	<i>taa</i>		

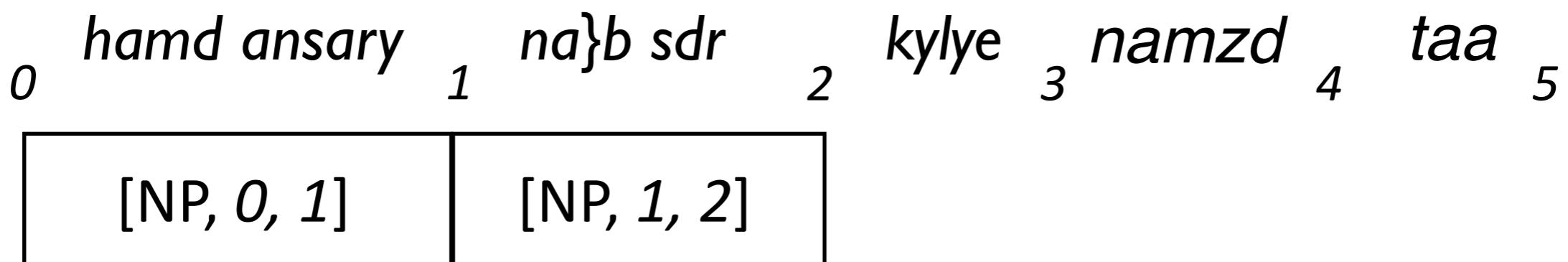
0 *hamd ansary* 1 *na\}b sdr* 2 *kylye* 3 *namzd* 4 *taa* 5

[NP, 0, 1]

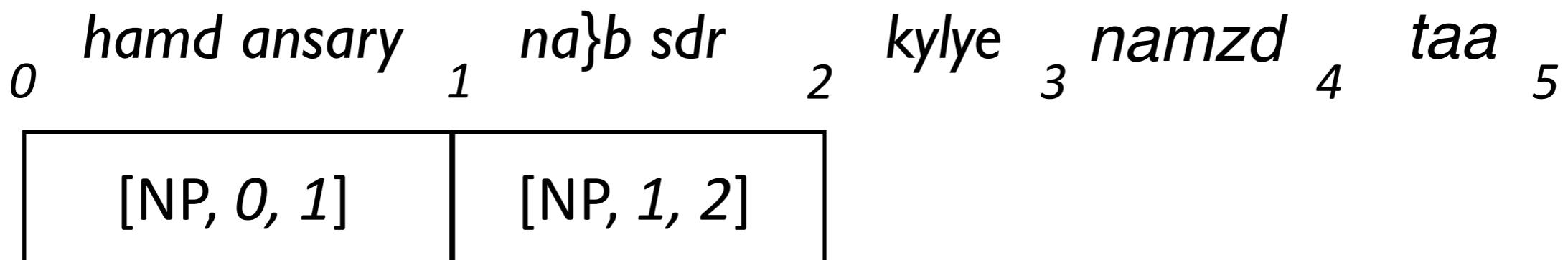
Axioms	S → NP VP	Inference rule used	Goal
VP →	PP VP		
VP →	V AUX	<u>NP → na}b sdr<sub>2</sub></u>	[S, 0, 5]
PP →	NP P		
NP →	<i>hamd ansary</i>	[NP, 1, 1]	
NP →	<i>na}b sdr</i>		
V →	<i>namzd</i>		
P →	<i>kylye</i>		
AUX →	<i>taa</i>		



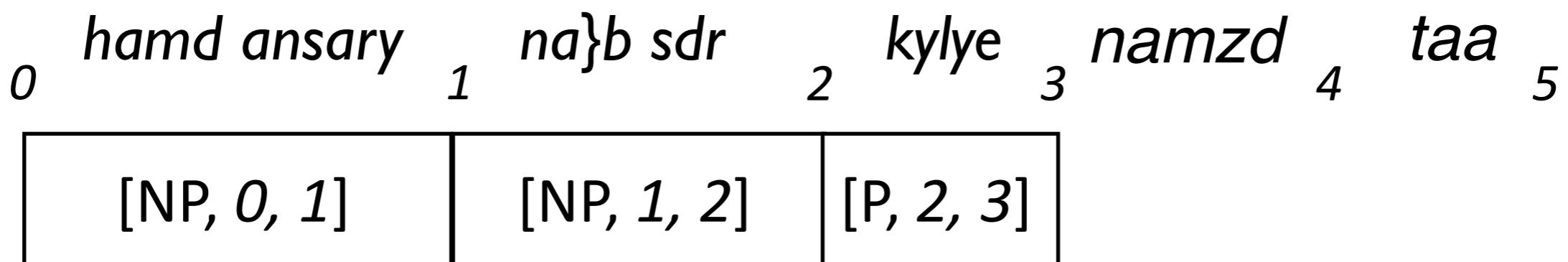
Axioms		Inference rule used	Goal
$S \rightarrow$	$NP \ VP$		
$VP \rightarrow$	$PP \ VP$		
$VP \rightarrow$	$V \ AUX$		$[S, 0, 5]$
$PP \rightarrow$	$NP \ P$		
$NP \rightarrow$	<i>hamd ansary</i>		
$NP \rightarrow$	<i>na}b sdr</i>		
$V \rightarrow$	<i>namzd</i>		
$P \rightarrow$	<i>kylye</i>		
$AUX \rightarrow$	<i>taa</i>		



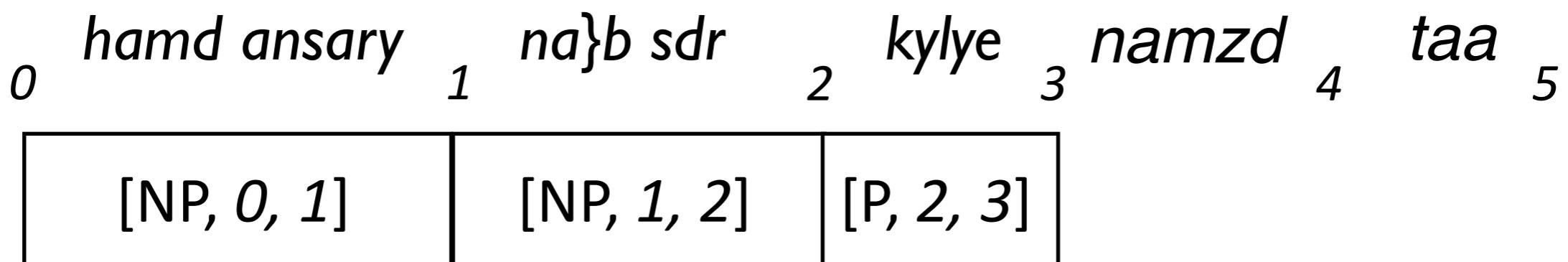
Axioms	S → NP VP	Inference rule used	Goal
VP →	PP VP		
VP →	V AUX	<u>P → kylye</u> <sub>3</sub>	[S, 0, 5]
PP →	NP P		
NP →	<i>hamd ansary</i>	[P, 2, 3]	
NP →	<i>na}b sdr</i>		
V →	<i>namzd</i>		
P →	<i>kylye</i>		
AUX →	<i>taa</i>		



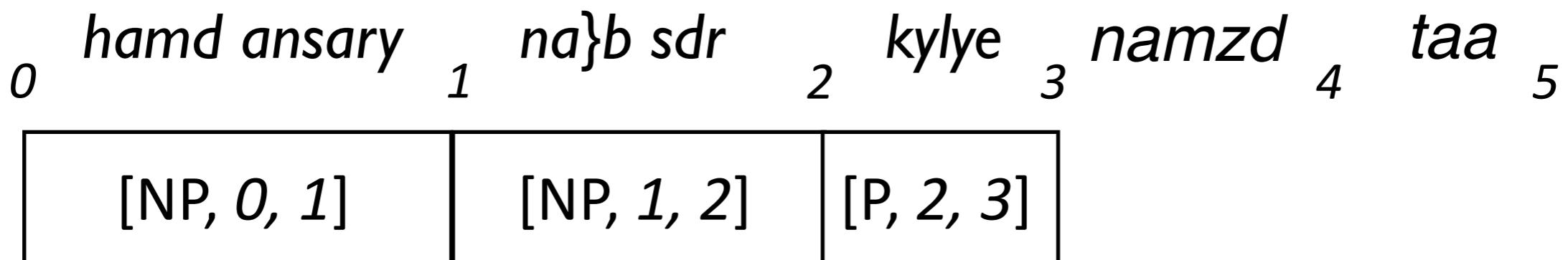
Axioms	S → NP VP	Inference rule used	Goal
VP →	PP VP		
VP →	V AUX	<u>P → kylye</u> <sub>3</sub>	[S, 0, 5]
PP →	NP P		
NP →	<i>hamd ansary</i>	[P, 2, 3]	
NP →	<i>na}b sdr</i>		
V →	<i>namzd</i>		
P →	<i>kylye</i>		
AUX →	<i>taa</i>		



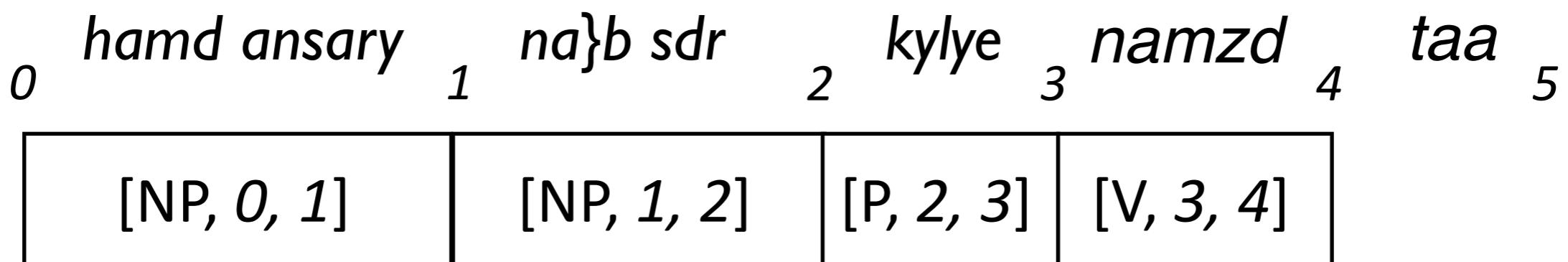
Axioms		Inference rule used	Goal
$S \rightarrow$	$NP \ VP$		
$VP \rightarrow$	$PP \ VP$		
$VP \rightarrow$	$V \ AUX$		$[S, 0, 5]$
$PP \rightarrow$	$NP \ P$		
$NP \rightarrow$	<i>hamd ansary</i>		
$NP \rightarrow$	<i>na}b sdr</i>		
$V \rightarrow$	<i>namzd</i>		
$P \rightarrow$	<i>kylye</i>		
$AUX \rightarrow$	<i>taa</i>		



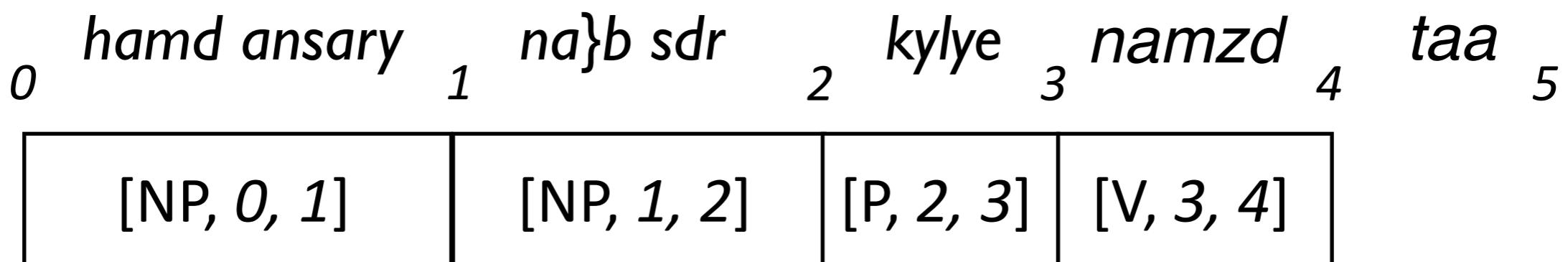
Axioms	S → NP VP	Inference rule used	Goal
VP →	PP VP		
VP →	V AUX	<u>V → namzd<sub>4</sub></u>	[S, 0, 5]
PP →	NP P		
NP →	<i>hamd ansary</i>	[V, 3, 4]	
NP →	<i>na}b sdr</i>		
V →	<i>namzd</i>		
P →	<i>kylye</i>		
AUX →	<i>taa</i>		



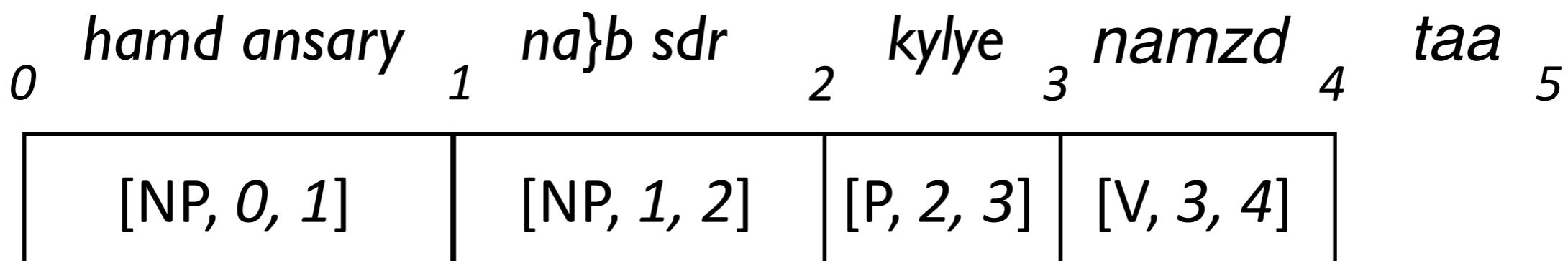
Axioms	S → NP VP	Inference rule used	Goal
VP →	PP VP		
VP →	V AUX	<u>V → namzd<sub>4</sub></u>	[S, 0, 5]
PP →	NP P		
NP →	<i>hamd ansary</i>	[V, 3, 4]	
NP →	<i>na}b sdr</i>		
V →	<i>namzd</i>		
P →	<i>kylye</i>		
AUX →	<i>taa</i>		



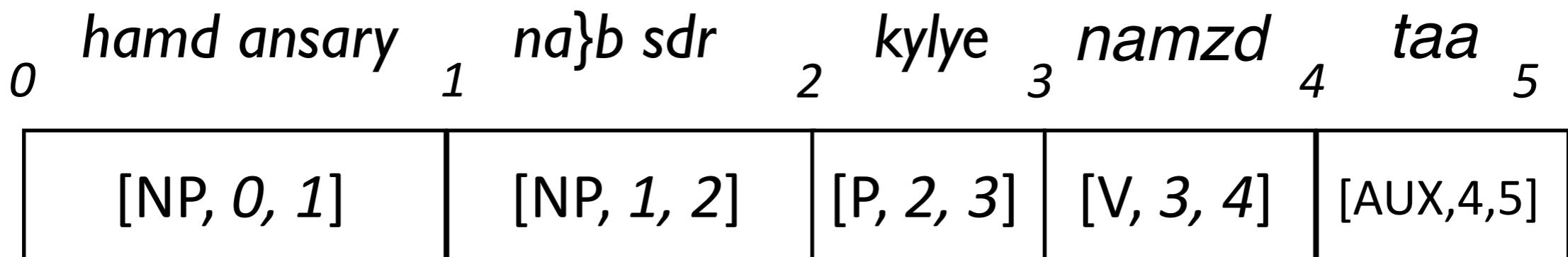
Axioms		Inference rule used	Goal
$S \rightarrow$	$NP \ VP$		
$VP \rightarrow$	$PP \ VP$		
$VP \rightarrow$	$V \ AUX$		$[S, 0, 5]$
$PP \rightarrow$	$NP \ P$		
$NP \rightarrow$	<i>hamd ansary</i>		
$NP \rightarrow$	<i>na}b sdr</i>		
$V \rightarrow$	<i>namzd</i>		
$P \rightarrow$	<i>kylye</i>		
$AUX \rightarrow$	<i>taa</i>		



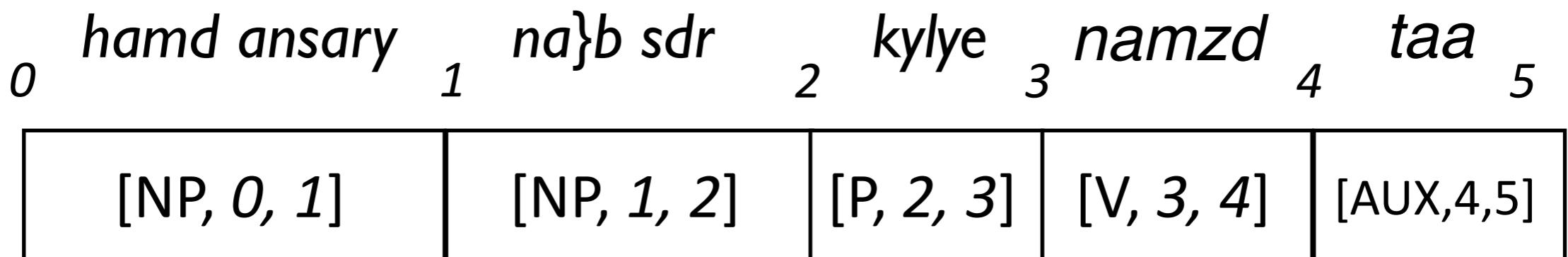
Axioms	S → NP VP	Inference rule used	Goal
VP →	PP VP		
VP →	V AUX	<u>AUX → taa<sub>5</sub></u>	[S, 0, 5]
PP →	NP P		
NP →	<i>hamd ansary</i>	[AUX, 4, 5]	
NP →	<i>na}b sdr</i>		
V →	<i>namzd</i>		
P →	<i>kylye</i>		
AUX →	<i>taa</i>		



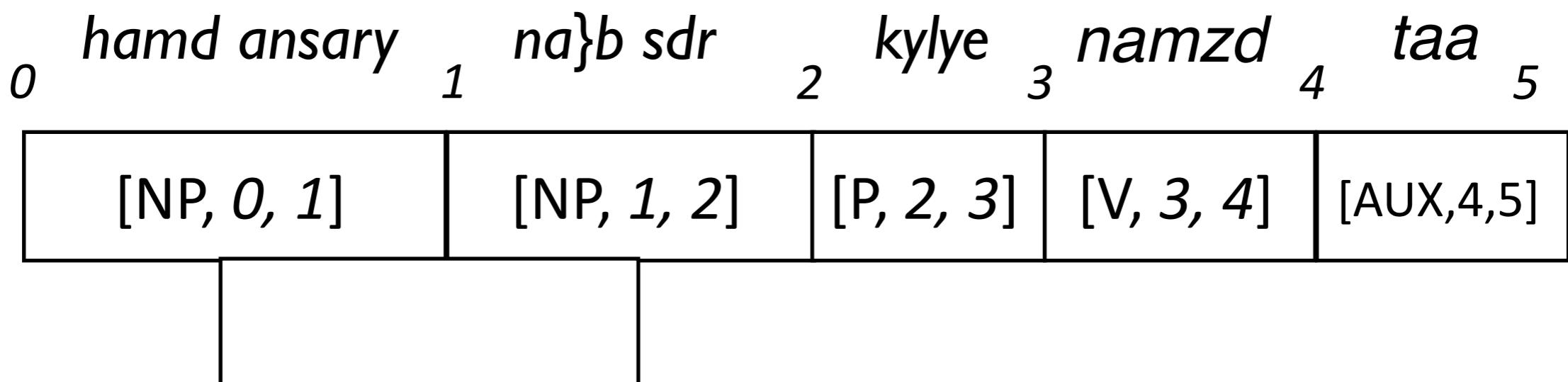
Axioms		Inference rule used	Goal
$S \rightarrow$	NP VP		
$VP \rightarrow$	PP VP		
$VP \rightarrow$	V AUX	<u>AUX <math>\rightarrow</math> taa<sub>5</sub></u>	[S, 0, 5]
$PP \rightarrow$	NP P		
$NP \rightarrow$	<i>hamd ansary</i>	[AUX, 4, 5]	
$NP \rightarrow$	<i>na}b sdr</i>		
$V \rightarrow$	<i>namzd</i>		
$P \rightarrow$	<i>kylye</i>		
$AUX \rightarrow$	<i>taa</i>		



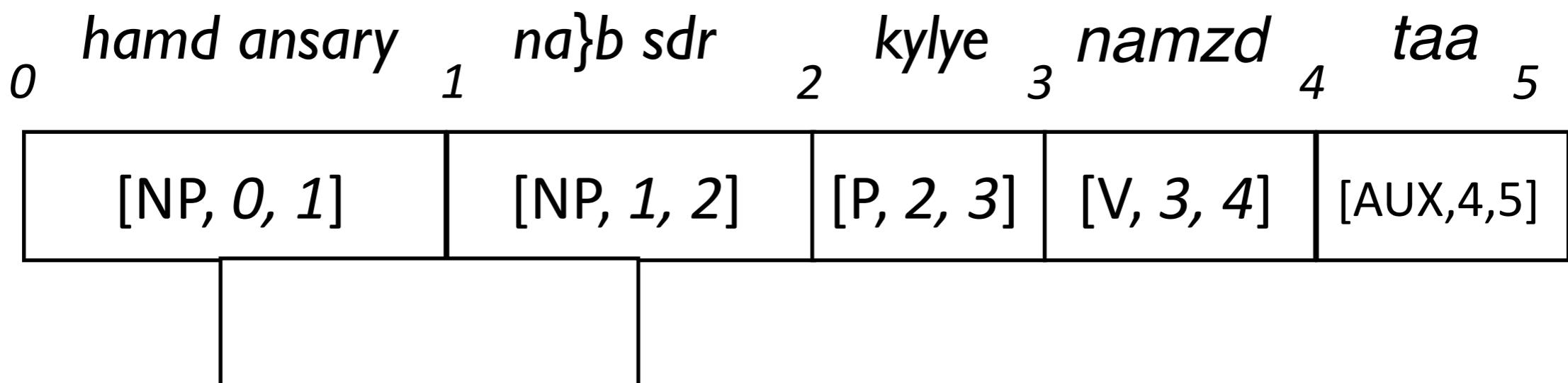
Axioms		Inference rule used	Goal
	$S \rightarrow NP VP$		
	$VP \rightarrow PP VP$		
	$VP \rightarrow V AUX$		$[S, 0, 5]$
	$PP \rightarrow NP P$		
	$NP \rightarrow hamd ansary$		
	$NP \rightarrow na}b sdr$		
	$V \rightarrow namzd$		
	$P \rightarrow kylye$		
	$AUX \rightarrow taa$		



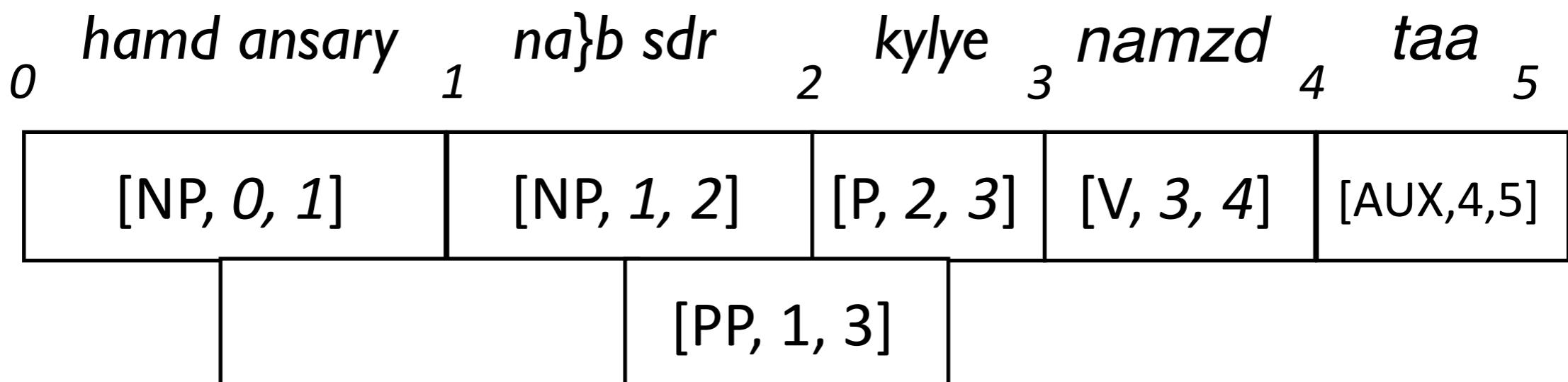
Axioms		Inference rule used	Goal
	$S \rightarrow NP VP$		
	$VP \rightarrow PP VP$		
	$VP \rightarrow V AUX$		$[S, 0, 5]$
	$PP \rightarrow NP P$		
	$NP \rightarrow hamd ansary$		
	$NP \rightarrow na}b sdr$		
	$V \rightarrow namzd$		
	$P \rightarrow kylye$		
	$AUX \rightarrow taa$		



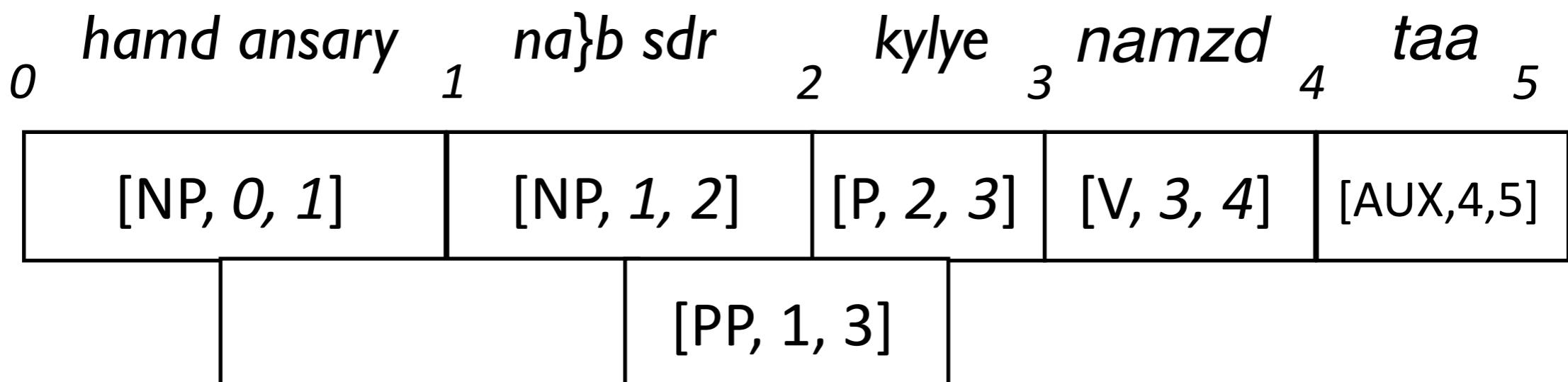
Axioms	S → NP VP	Inference rule used	Goal
VP →	PP VP		
VP →	V AUX	[NP, 1, 2] [P, 2, 3] PP → NP P	[S, 0, 5]
PP →	NP P		
NP →	<i>hamd ansary</i>	[PP, 1, 3]	
NP →	<i>na}b sdr</i>		
V →	<i>namzd</i>		
P →	<i>kylye</i>		
AUX →	<i>taa</i>		



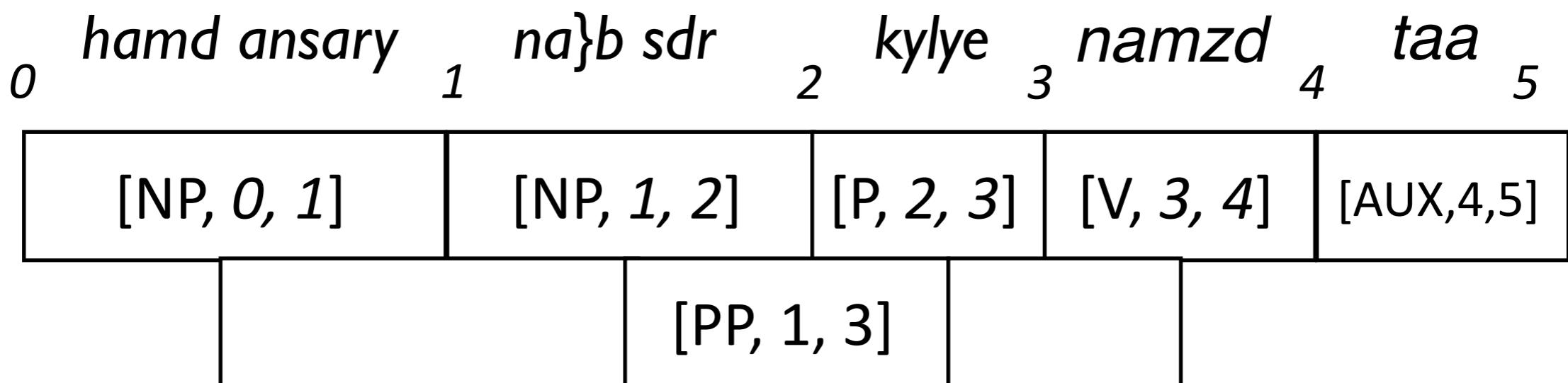
Axioms	S → NP VP	Inference rule used	Goal
VP →	PP VP		
VP →	V AUX	[NP, 1, 2] [P, 2, 3] PP → NP P	[S, 0, 5]
PP →	NP P		
NP →	<i>hamd ansary</i>	[PP, 1, 3]	
NP →	<i>na}b sdr</i>		
V →	<i>namzd</i>		
P →	<i>kylye</i>		
AUX →	<i>taa</i>		



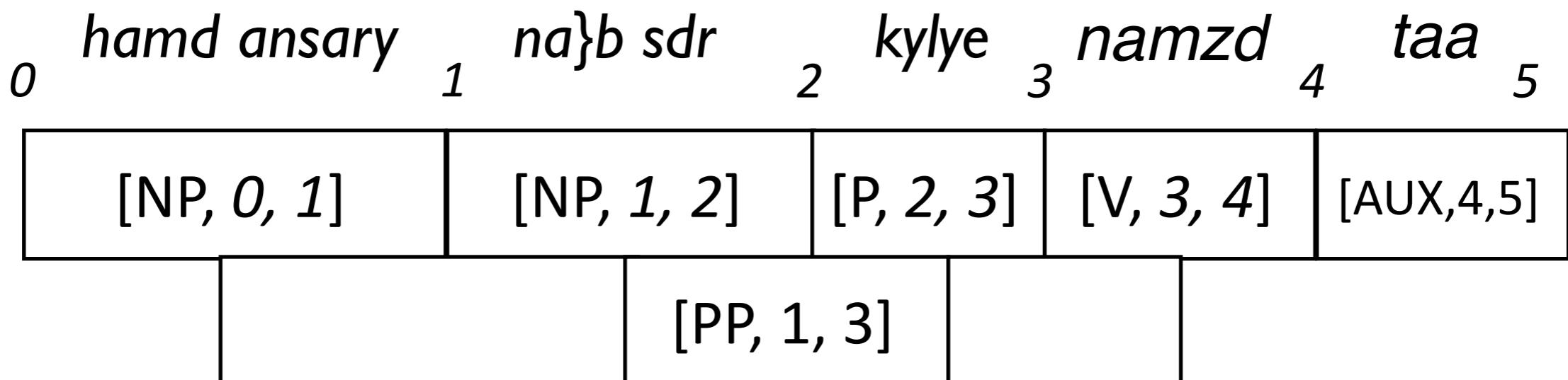
Axioms		Inference rule used	Goal
	$S \rightarrow NP VP$		
	$VP \rightarrow PP VP$		
	$VP \rightarrow V AUX$		$[S, 0, 5]$
	$PP \rightarrow NP P$		
	$NP \rightarrow hamd ansary$		
	$NP \rightarrow na}b sdr$		
	$V \rightarrow namzd$		
	$P \rightarrow kylye$		
	$AUX \rightarrow taa$		



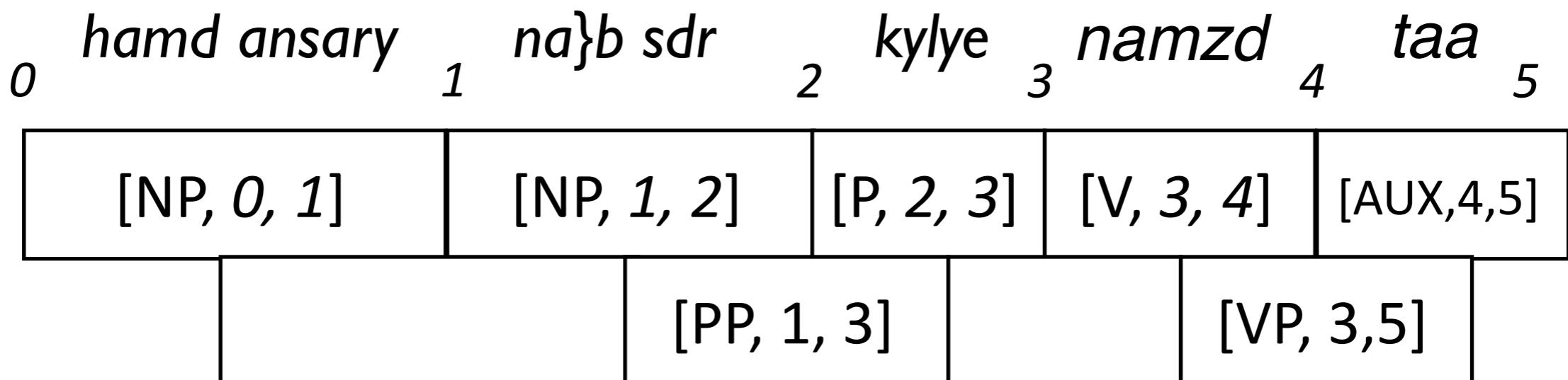
Axioms		Inference rule used	Goal
	$S \rightarrow NP VP$		
	$VP \rightarrow PP VP$		
	$VP \rightarrow V AUX$		$[S, 0, 5]$
	$PP \rightarrow NP P$		
	$NP \rightarrow hamd ansary$		
	$NP \rightarrow na}b sdr$		
	$V \rightarrow namzd$		
	$P \rightarrow kylye$		
	$AUX \rightarrow taa$		



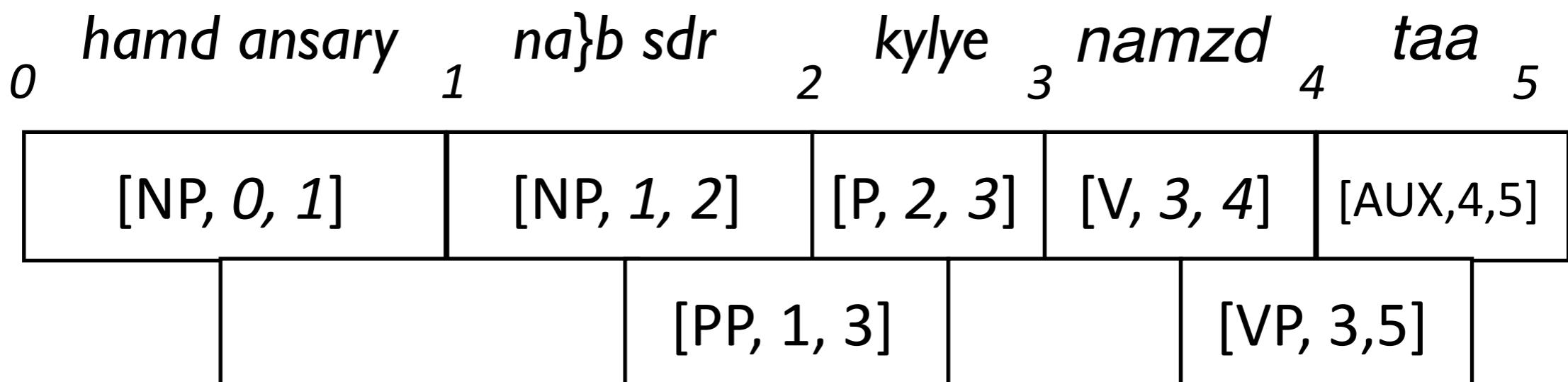
Axioms	S → NP VP	Inference rule used	Goal
	VP → PP VP		
	VP → V AUX	[V, 3, 4] [AUX, 4, 5]	VP → V AUX [S, 0, 5]
	PP → NP P		
NP → <i>hamd ansary</i>			[VP, 3, 5]
NP → <i>na}b sdr</i>			
V → <i>namzd</i>			
P → <i>kylye</i>			
AUX → <i>taa</i>			



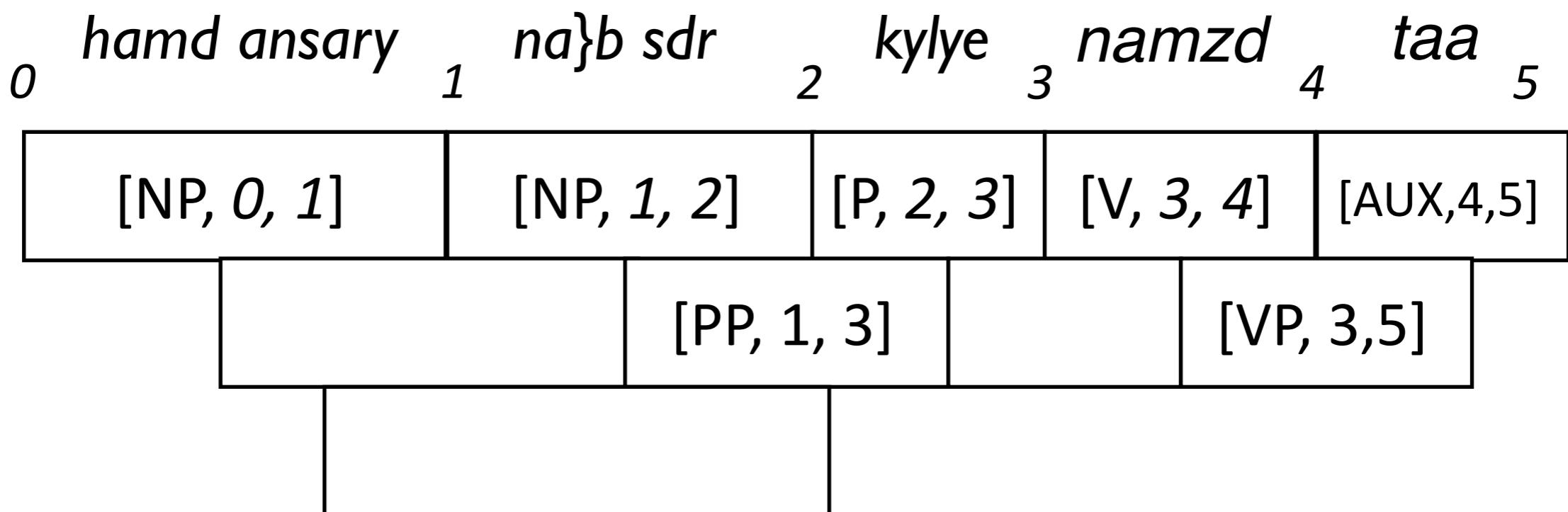
Axioms	S → NP VP	Inference rule used	Goal
	VP → PP VP		
	VP → V AUX	[V, 3, 4] [AUX, 4, 5]	VP → V AUX [S, 0, 5]
	PP → NP P		
NP → <i>hamd ansary</i>			[VP, 3, 5]
NP → <i>na}b sdr</i>			
V → <i>namzd</i>			
P → <i>kylye</i>			
AUX → <i>taa</i>			



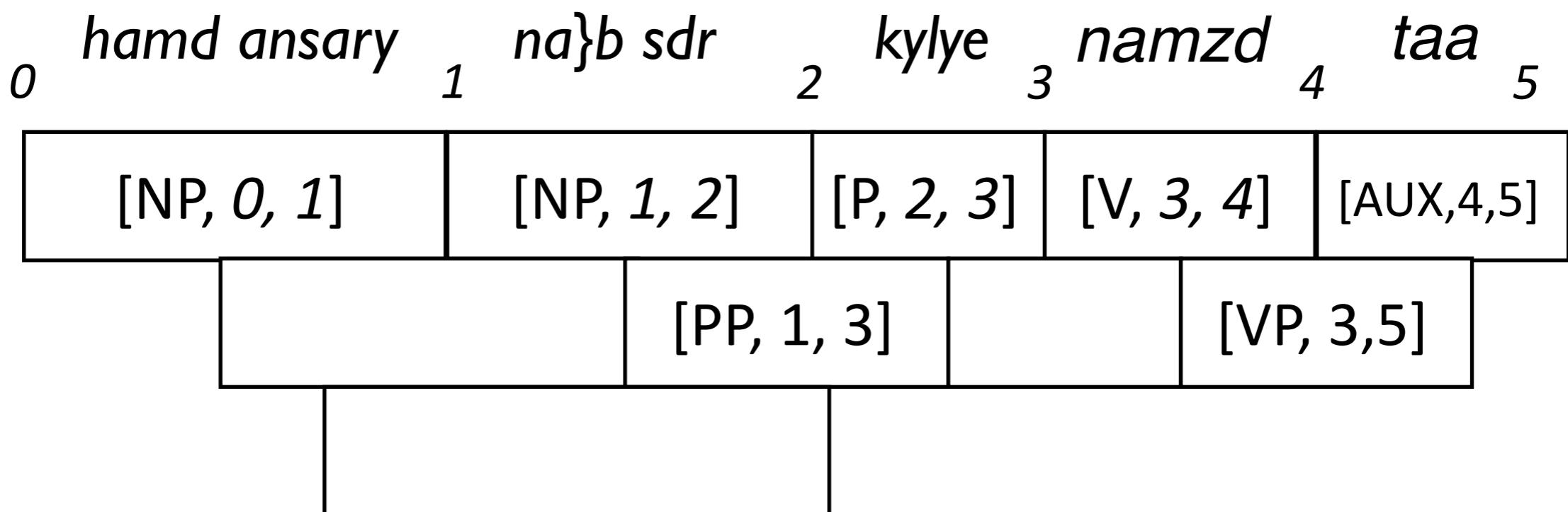
Axioms		Inference rule used	Goal
	$S \rightarrow NP VP$		
	$VP \rightarrow PP VP$		
	$VP \rightarrow V AUX$		$[S, 0, 5]$
	$PP \rightarrow NP P$		
	$NP \rightarrow hamd ansary$		
	$NP \rightarrow na}b sdr$		
	$V \rightarrow namzd$		
	$P \rightarrow kylye$		
	$AUX \rightarrow taa$		



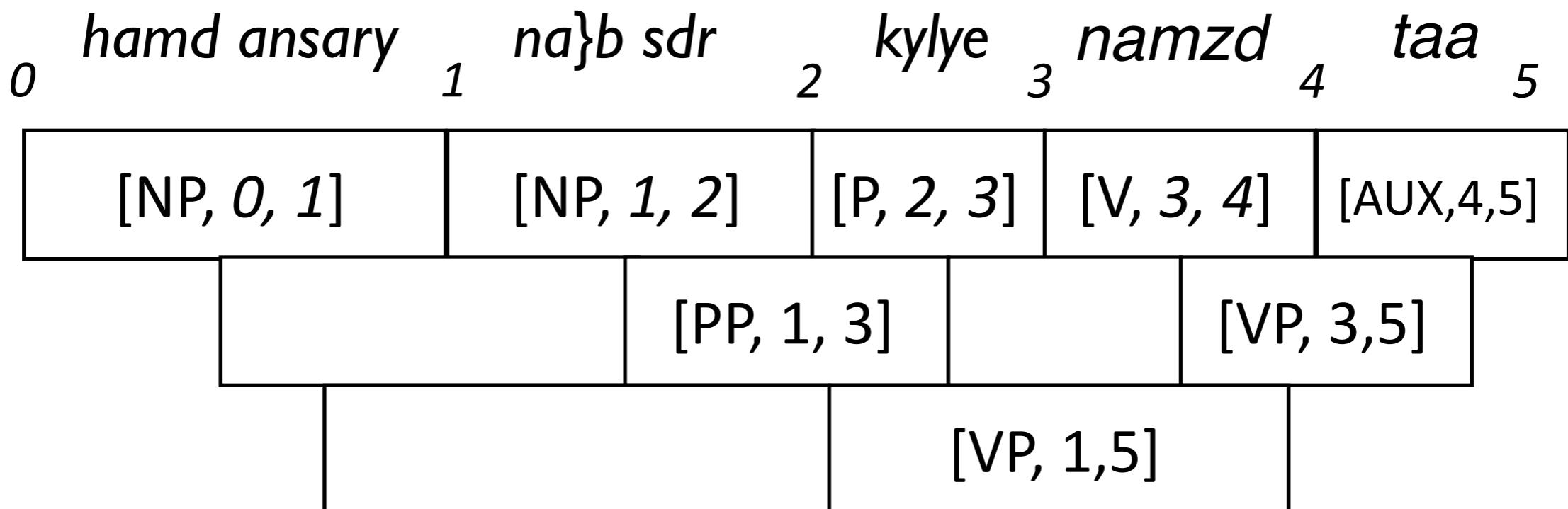
Axioms		Inference rule used	Goal
	$S \rightarrow NP VP$		
	$VP \rightarrow PP VP$		
	$VP \rightarrow V AUX$		$[S, 0, 5]$
	$PP \rightarrow NP P$		
	$NP \rightarrow hamd ansary$		
	$NP \rightarrow na}b sdr$		
	$V \rightarrow namzd$		
	$P \rightarrow kylye$		
	$AUX \rightarrow taa$		



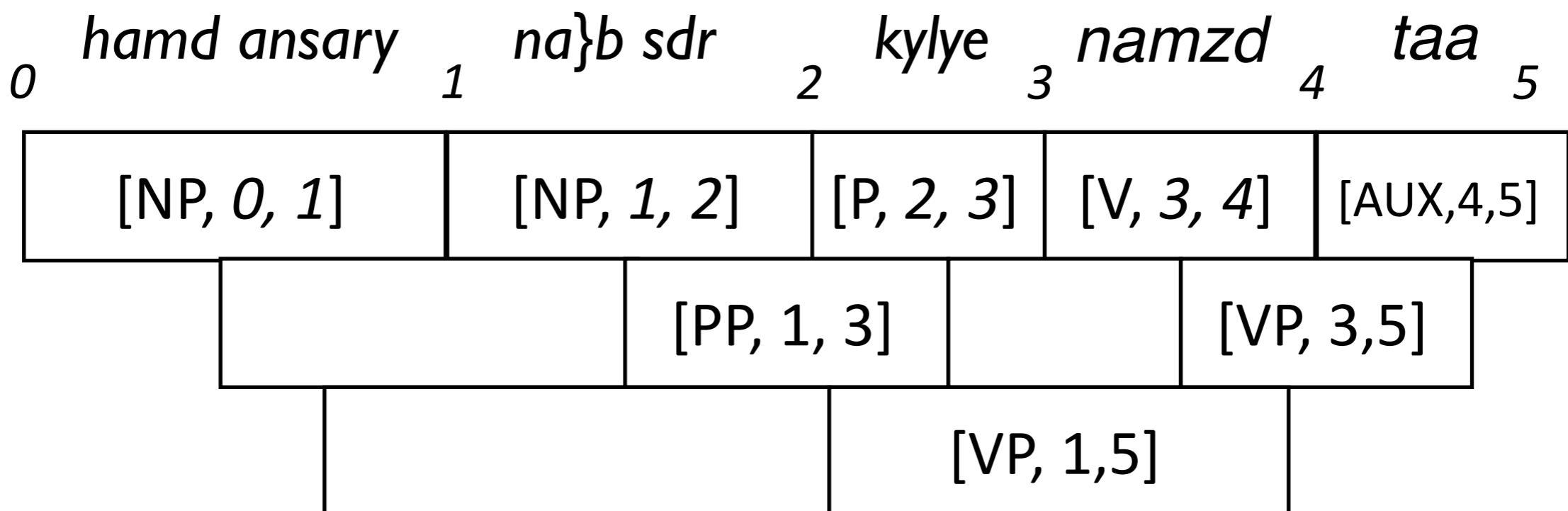
Axioms	S → NP VP	Inference rule used	Goal
	VP → PP VP		
	VP → V AUX	[PP, 1, 3] [VP, 3, 5] VP → PP CP	[S, 0, 5]
	PP → NP P		
NP → <i>hamd ansary</i>		[VP, 1, 5]	
NP → <i>na}b sdr</i>			
V → <i>namzd</i>			
P → <i>kylye</i>			
AUX → <i>taa</i>			



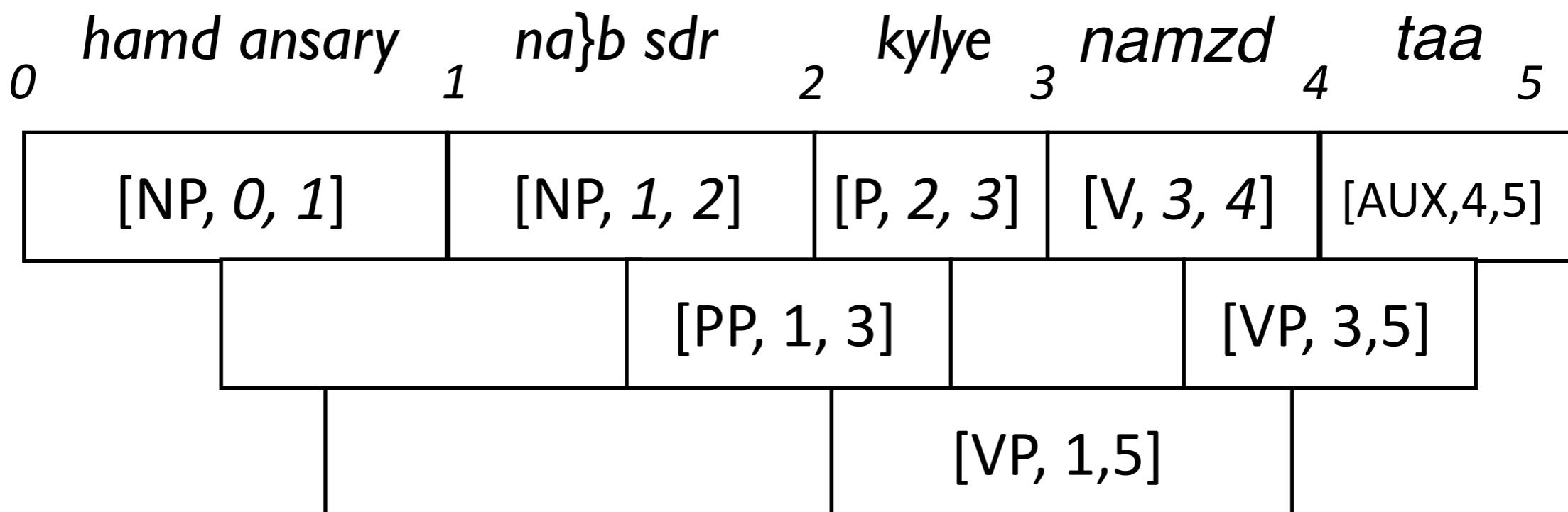
Axioms	S → NP VP	Inference rule used	Goal
	VP → PP VP		
	VP → V AUX	[PP, 1, 3] [VP, 3, 5] VP → PP CP	[S, 0, 5]
	PP → NP P		
NP → <i>hamd ansary</i>		[VP, 1, 5]	
NP → <i>na}b sdr</i>			
V → <i>namzd</i>			
P → <i>kylye</i>			
AUX → <i>taa</i>			



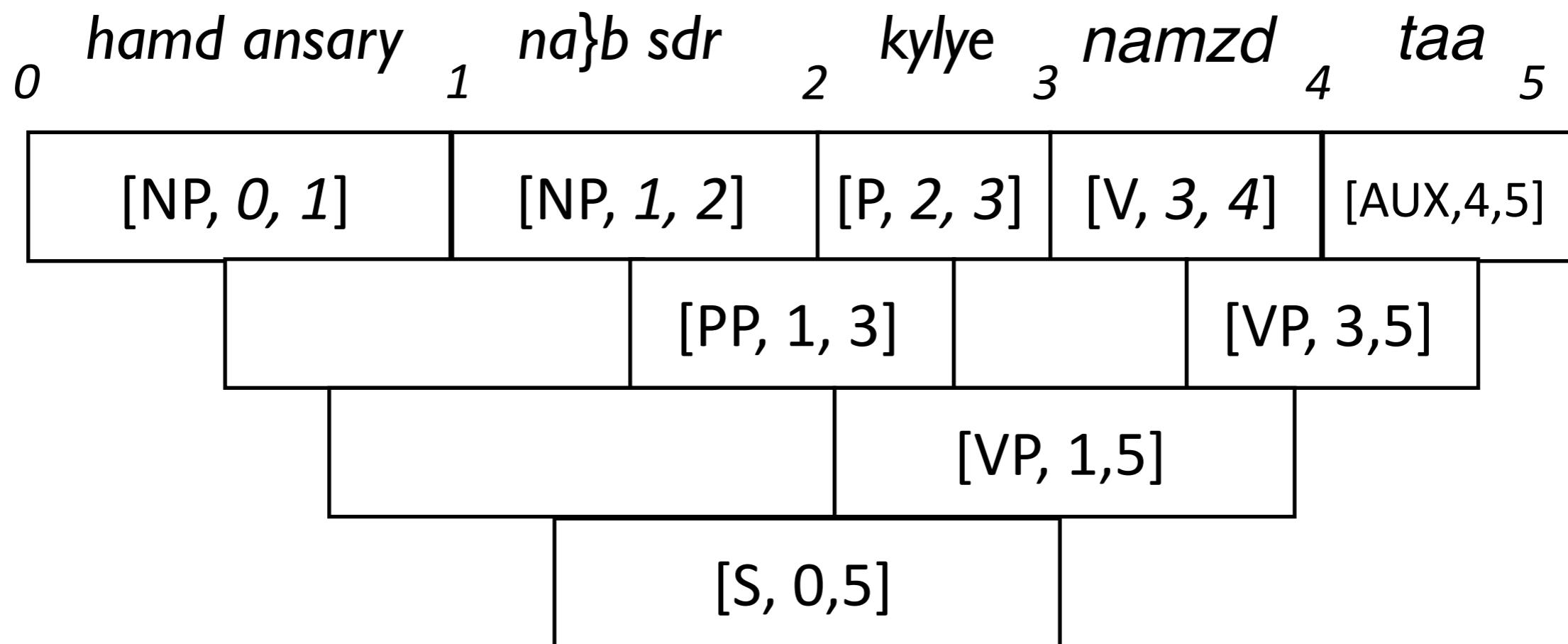
Axioms		Inference rule used	Goal
	$S \rightarrow NP VP$		
	$VP \rightarrow PP VP$		
	$VP \rightarrow V AUX$		$[S, 0, 5]$
	$PP \rightarrow NP P$		
	$NP \rightarrow hamd ansary$		
	$NP \rightarrow na}b sdr$		
	$V \rightarrow namzd$		
	$P \rightarrow kylye$		
	$AUX \rightarrow taa$		



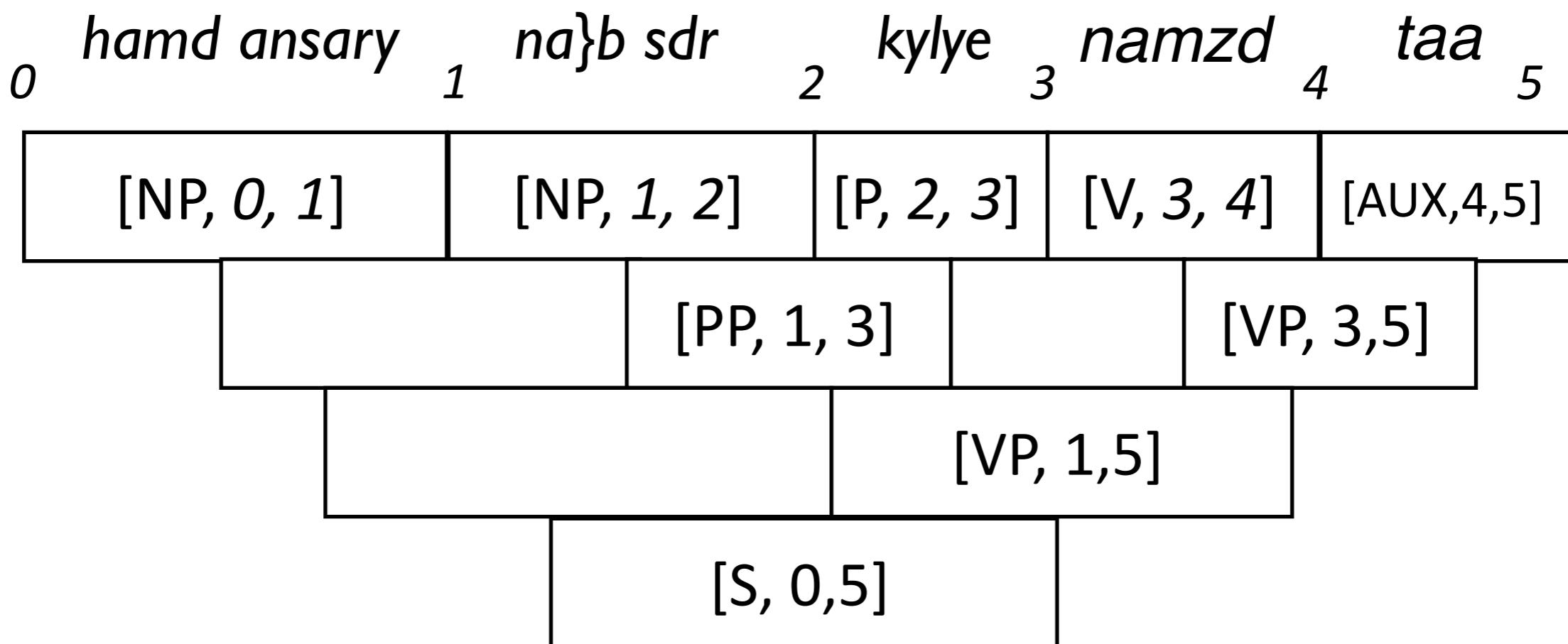
Axioms	S → NP VP	Inference rule used	Goal
VP →	PP VP		
VP →	V AUX	[NP, 0, 1] [VP, 1, 5] S → NP VP [S, 0, 5]	
PP →	NP P		
NP →	<i>hamd ansary</i>		[S, 0, 5]
NP →	<i>na}b sdr</i>		
V →	<i>namzd</i>		
P →	<i>kylye</i>		
AUX →	<i>taa</i>		



Axioms	S →	NP VP	Inference rule used	Goal
	VP →	PP VP		
	VP →	V AUX	<u>[NP, 0, 1] [VP, 1, 5] S → NP VP</u>	[S, 0, 5]
	PP →	NP P		
	NP →	<i>hamd ansary</i>		[S, 0, 5]
	NP →	<i>na}b sdr</i>		
	V →	<i>namzd</i>		
	P →	<i>kylye</i>		
	AUX →	<i>taa</i>		



Axioms		Inference rule used	Goal
	$S \rightarrow NP VP$		
	$VP \rightarrow PP VP$		
	$VP \rightarrow V AUX$		$[S, 0, 5]$
	$PP \rightarrow NP P$		
	$NP \rightarrow hamd ansary$		
	$NP \rightarrow na}b sdr$		
	$V \rightarrow namzd$		
	$P \rightarrow kylye$		
	$AUX \rightarrow taa$		



# The CKY Parsing Algorithm

Axioms	$\frac{}{A \rightarrow \alpha}$	for all $(A \rightarrow \alpha) \in R$
Inference rules	$\frac{A}{[A, i, i+1]}$ $\frac{[B, i, j] \ [C, j, k] \ A \rightarrow BC}{[A, i, k]}$	
Goal	$[S, 0, n]$	

# The CKY Translation Algorithm

Axioms	$\frac{}{A \rightarrow \alpha, \beta}$	for all $(A \rightarrow \alpha, \beta) \in R$
Inference rules	$\frac{A}{[A, i, i+1]}$ $\frac{[B, i, j] \ [C, j, k] \ A \rightarrow BC}{[A, i, k]}$	
Goal	$[S, 0, n]$	

# Where do grammars come from?

- Great! We now have
  - a formalism for describing the relationship between two languages,
  - an algorithm for producing translations
- All we need now is a synchronous grammar

# Where do grammars come from?

- Great! We now have
  - a formalism for describing the relationship between two languages,
  - an algorithm for producing translations
- All we need now is a synchronous grammar
- Where do grammars come from?
- Well, when two languages love each other very much...



# Data-driven grammar extraction

- Grammar rules are not written by hand, they are extracted from bilingual parallel corpora

## Arabic

فالتعذيب لا يزال يمارس على نطاق واسع

وتتم عمليات الاعتقال والاحتجاز دون سبب بصورة  
روتينية

وحان وقت التحلى بالبصيرة والشجاعة السياسية.

...

## English

Torture is still being practised on a wide scale.

Arrest and detention without cause take place routinely.

This is a time for vision and political courage

...

## Chinese

我国 能源 原材料 工业 生产 大幅度 增长 .

非国大 要求 阻止 更多 被 拘留 人员 死亡 .

...

## English

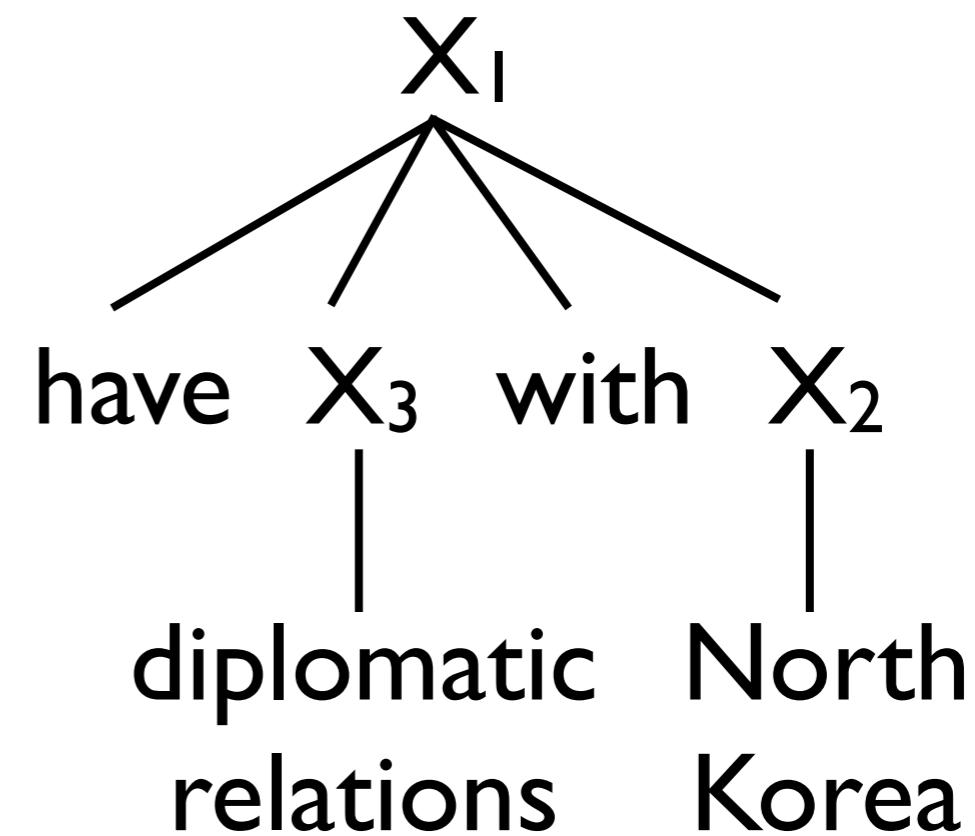
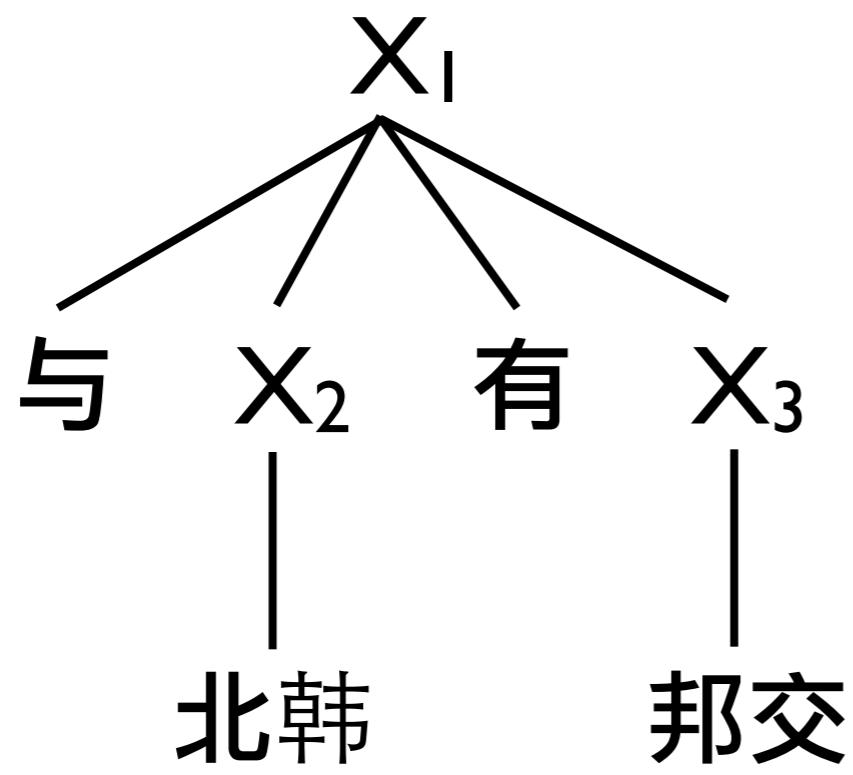
China's energy and raw materials production up.

ANC calls for steps to prevent deaths in police custody .

...

# Hiero-style SCFG rules

- Most common type of SCFG in SMT is Hiero which has rules w/one non-terminal symbol
- Not as nice as linguistically motivated rules, does not capture the reordering in Urdu



# Extracting Hiero rules

# 澳洲是与北韩有邦交的少数国家之一

# Extracting Hiero rules

# 澳洲是与北韩有邦交的少数国家之一

Australia  
is  
one  
of  
the  
few  
countries  
that  
have  
diplomatic  
relations  
with  
North  
Korea

X → 与 北 韩 有 邦 交,  
have diplomatic relations  
with North Korea

# Extracting Hiero rules

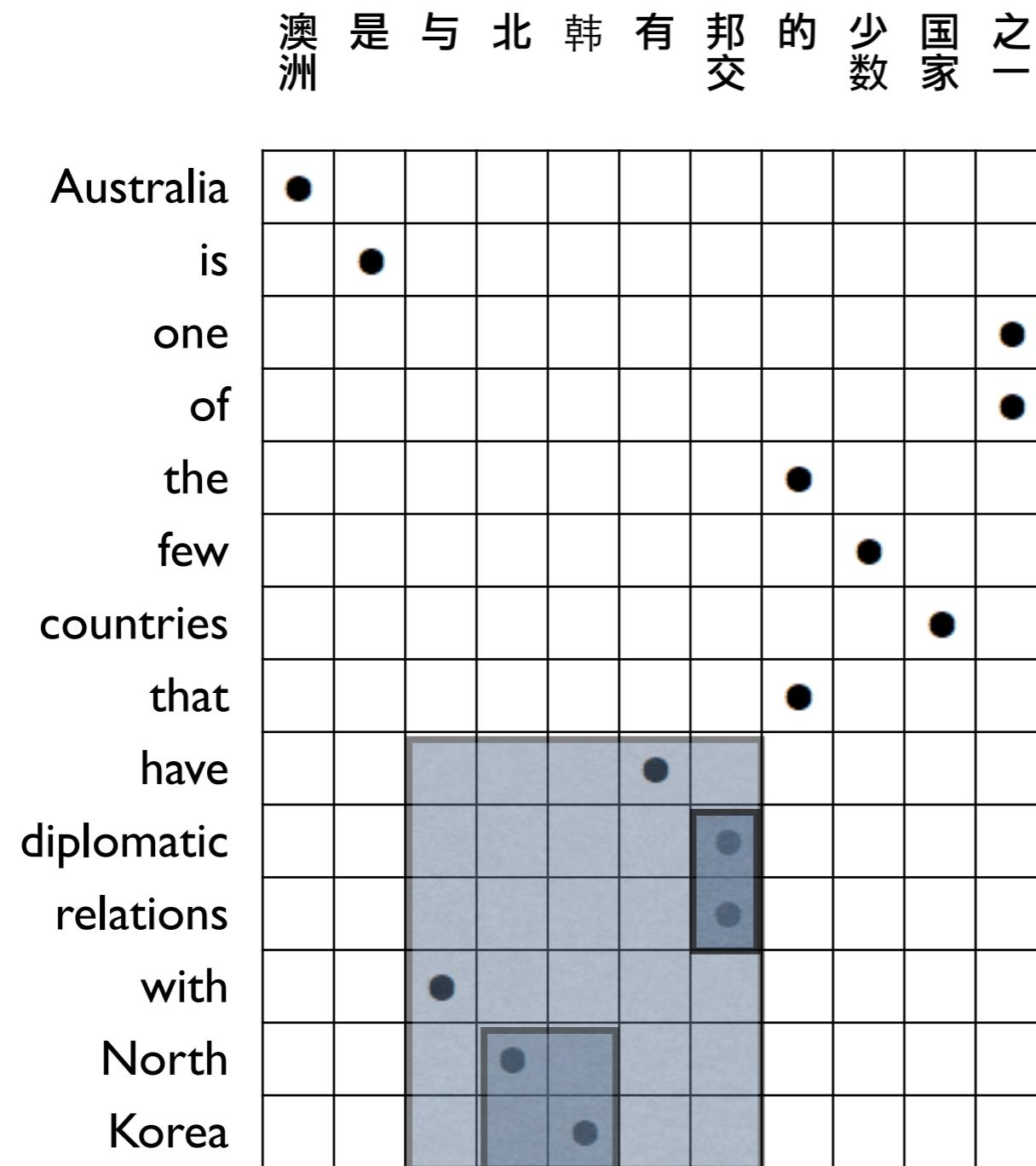
# 澳洲是与北韩有邦交的少数国家之一

Australia  
is  
one  
of  
the  
few  
countries  
that  
have  
diplomatic  
relations  
with  
North  
Korea

X → 与 北 韩 有 邦 交,  
have diplomatic relations  
with North Korea

X → 邦交,  
diplomatic relations

# Extracting Hiero rules

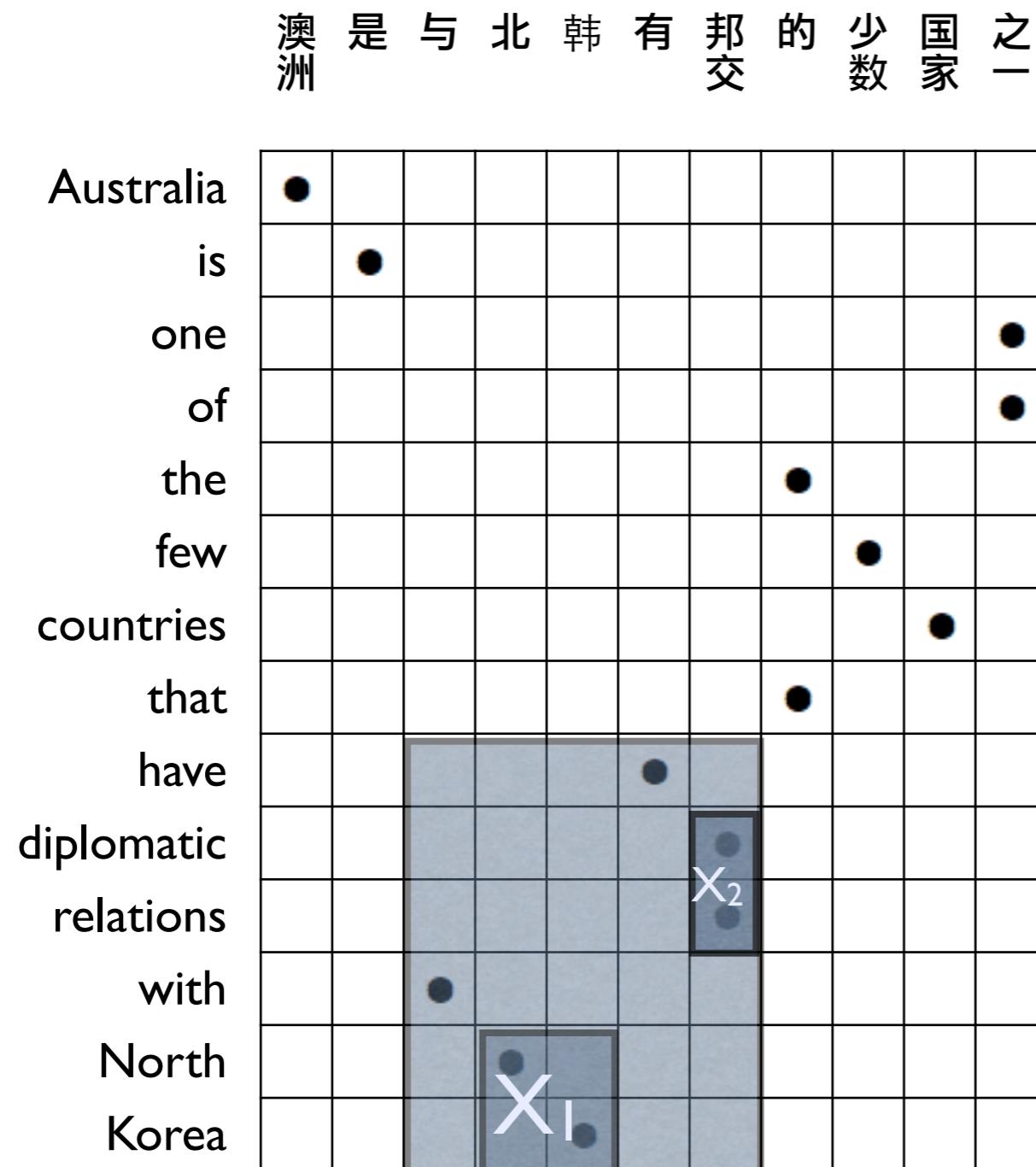


$\times \rightarrow$  与 北 韩 有 邦 交,  
have diplomatic relations  
with North Korea

$\times \rightarrow$  邦 交,  
diplomatic relations

$\times \rightarrow$  北 韩,  
North Korea

# Extracting Hiero rules

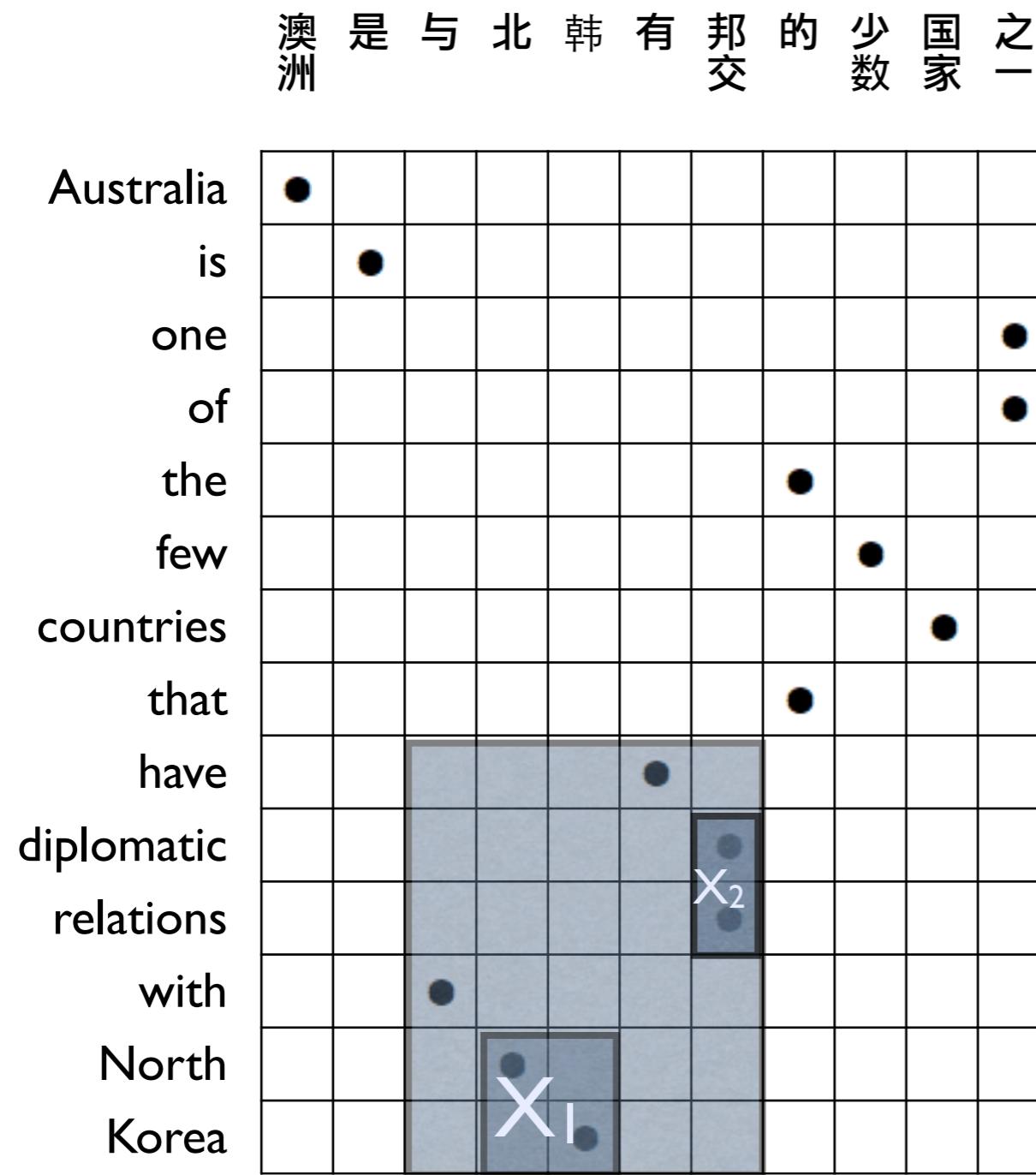


$\text{X} \rightarrow \text{与 北 韩 有 邦 交}$ ,  
have diplomatic relations  
with North Korea

$\text{X} \rightarrow \text{邦 交}$ ,  
diplomatic relations

$\text{X} \rightarrow \text{北 韩}$ ,  
North Korea

# Extracting Hiero rules



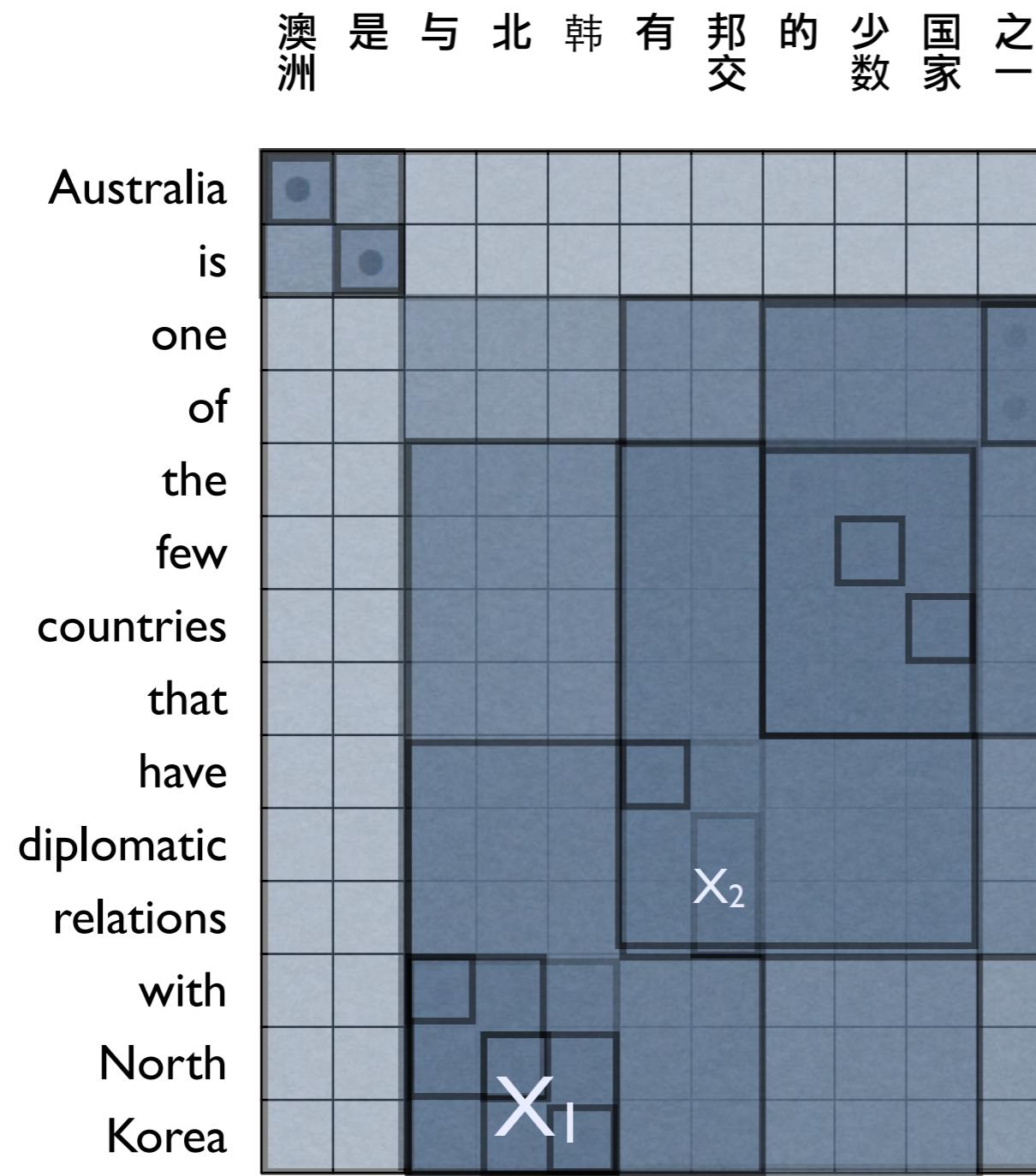
$X \rightarrow$  与 北 韩 有 邦 交,  
have diplomatic relations  
with North Korea

$X \rightarrow$  邦 交,  
diplomatic relations

$X \rightarrow$  北 韩,  
North Korea

$X \rightarrow$  与  $X_1$  有  $X_2$ ,  
have  $X_2$  with  $X_1$

# Extracting Hiero rules



$X \rightarrow$  与 北 韩 有 邦 交,  
have diplomatic relations  
with North Korea

$X \rightarrow$  邦 交,  
diplomatic relations

$X \rightarrow$  北 韩,  
North Korea

$X \rightarrow$  与  $X_1$  有  $X_2$ ,  
have  $X_2$  with  $X_1$

# Discussion: what do you think of Hiero?

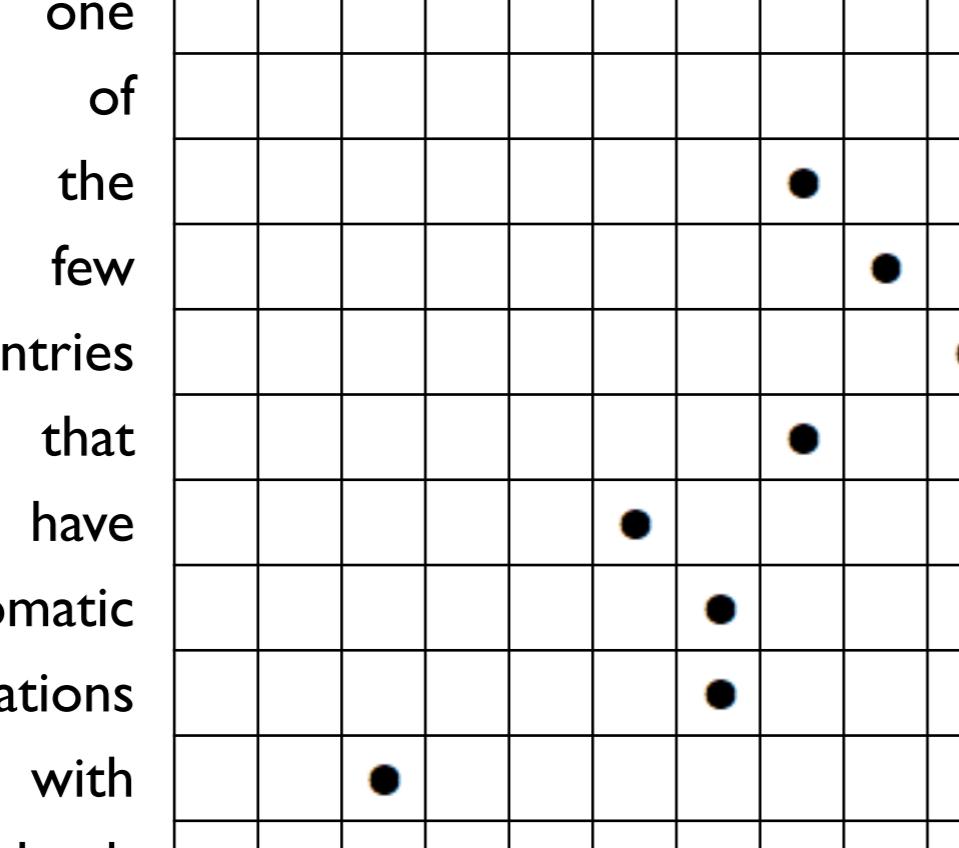
- So, we now have a way of extracting SCFGs from bitexts. Great! So what?
- Is this any better than the phrase based model?
- How?
- Do you feel that it is lacking anything?

(Discuss with your neighbor)

# Extracting Syntactic Rules

# 澳洲是与北韩有邦交的少数国家之一

Australia  
is  
one  
of  
the  
few  
countries  
that  
have  
diplomatic  
relations  
with  
North  
Korea

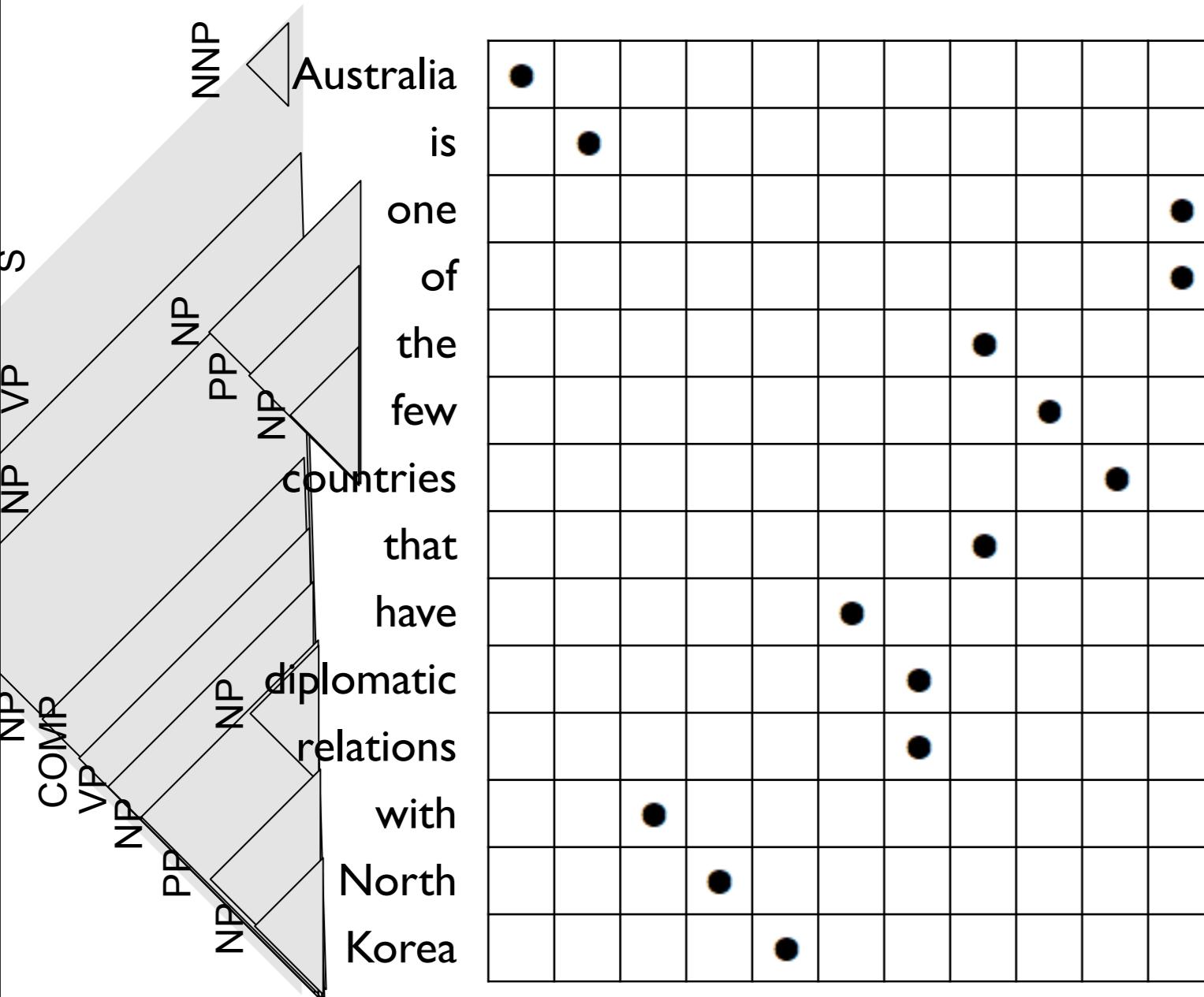


The grid has 10 columns and 10 rows. The words are placed as follows:

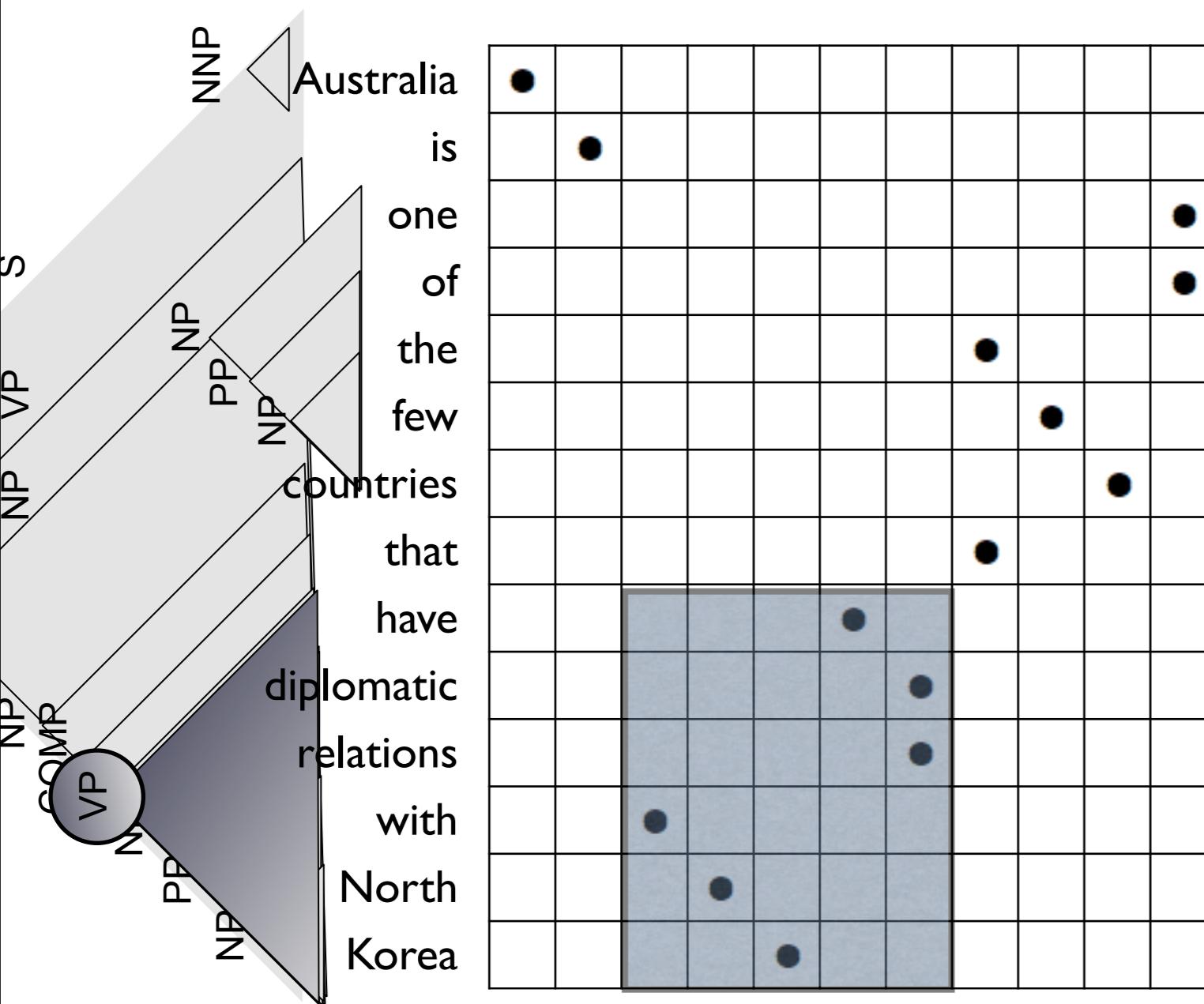
- "Australia" is at (1, 1)
- "is" is at (2, 2)
- "one" is at (3, 3)
- "of" is at (4, 4)
- "the" is at (5, 5)
- "few" is at (6, 6)
- "countries" is at (7, 7)
- "that" is at (8, 8)
- "have" is at (9, 9)
- "diplomatic" is at (10, 1)
- "relations" is at (11, 2)
- "with" is at (12, 3)
- "North" is at (13, 4)
- "Korea" is at (14, 5)

# Extracting Syntactic Rules

# 澳洲是与北韩有邦交的少数国家之一



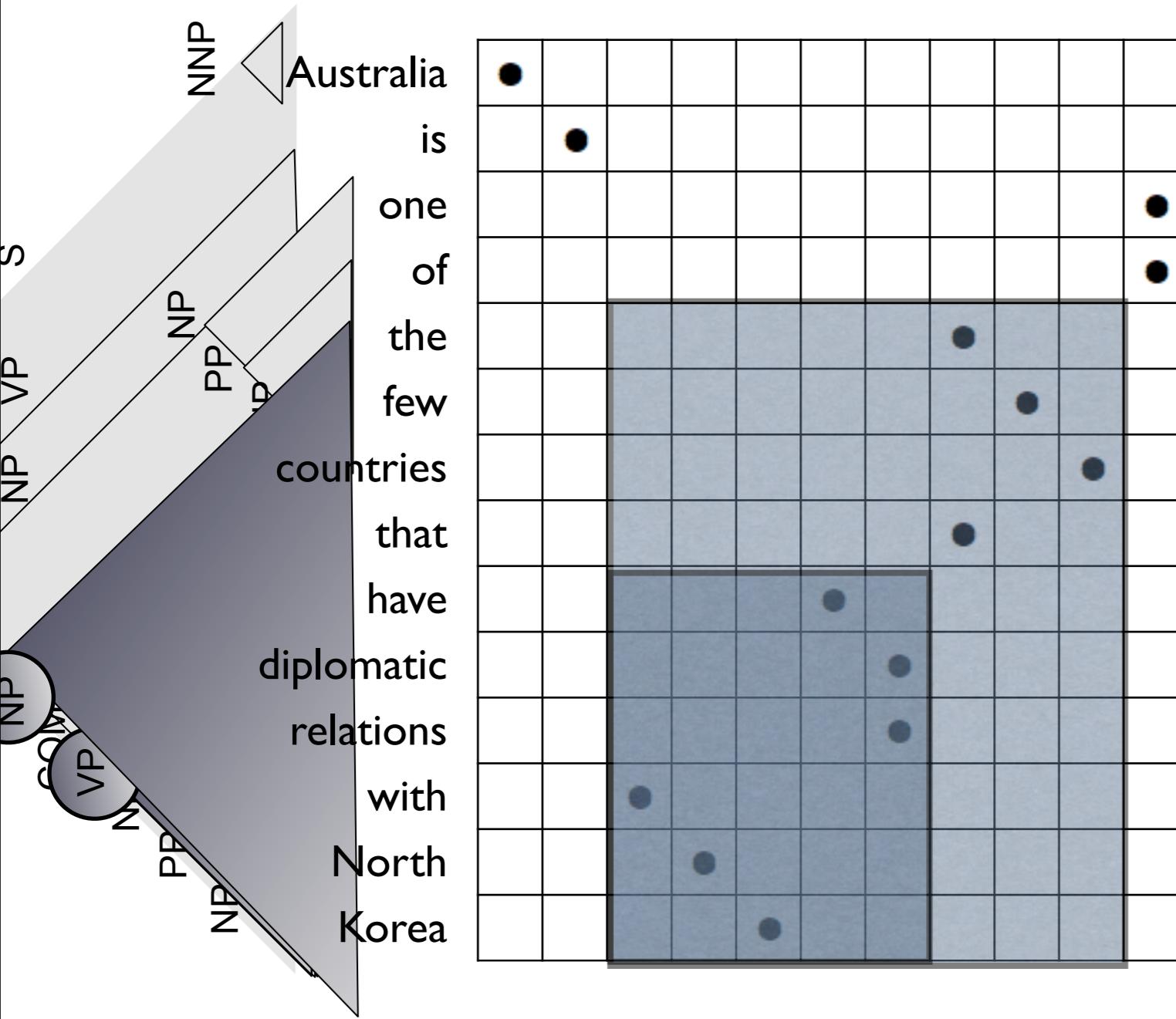
# Extracting Syntactic Rules



# 澳洲是与北韩有邦交的少数国家之一

VP → 与 北 韩 有 邦 交,  
have diplomatic relations  
with North Korea

# Extracting Syntactic Rules

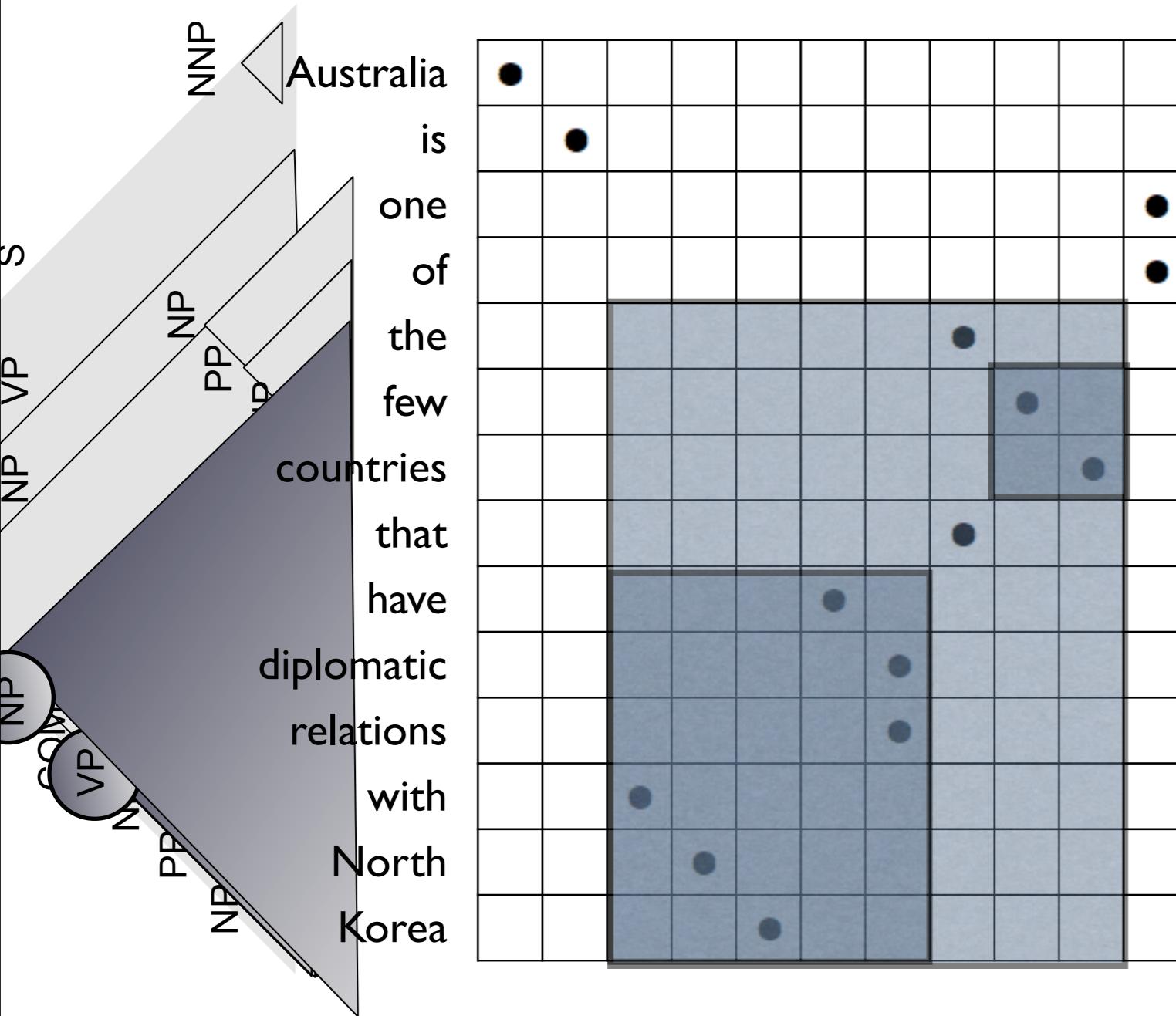


$VP \rightarrow \text{与北韩有邦交}$ ,  
have diplomatic relations  
with North Korea

$NP \rightarrow \text{与北韩有邦交}$   
的 少数 国家, the few  
countries that have  
diplomatic relations with  
North Korea

$NP \rightarrow VP$  的 少数 国家,  
the few countries that VP

# Extracting Syntactic Rules



$VP \rightarrow \text{与 北 韩 有 邦 交}$ ,  
have diplomatic relations  
with North Korea

$NP \rightarrow \text{与 北 韩 有 邦 交}$   
的 少 数 国 家, the few  
countries that have  
diplomatic relations with  
North Korea

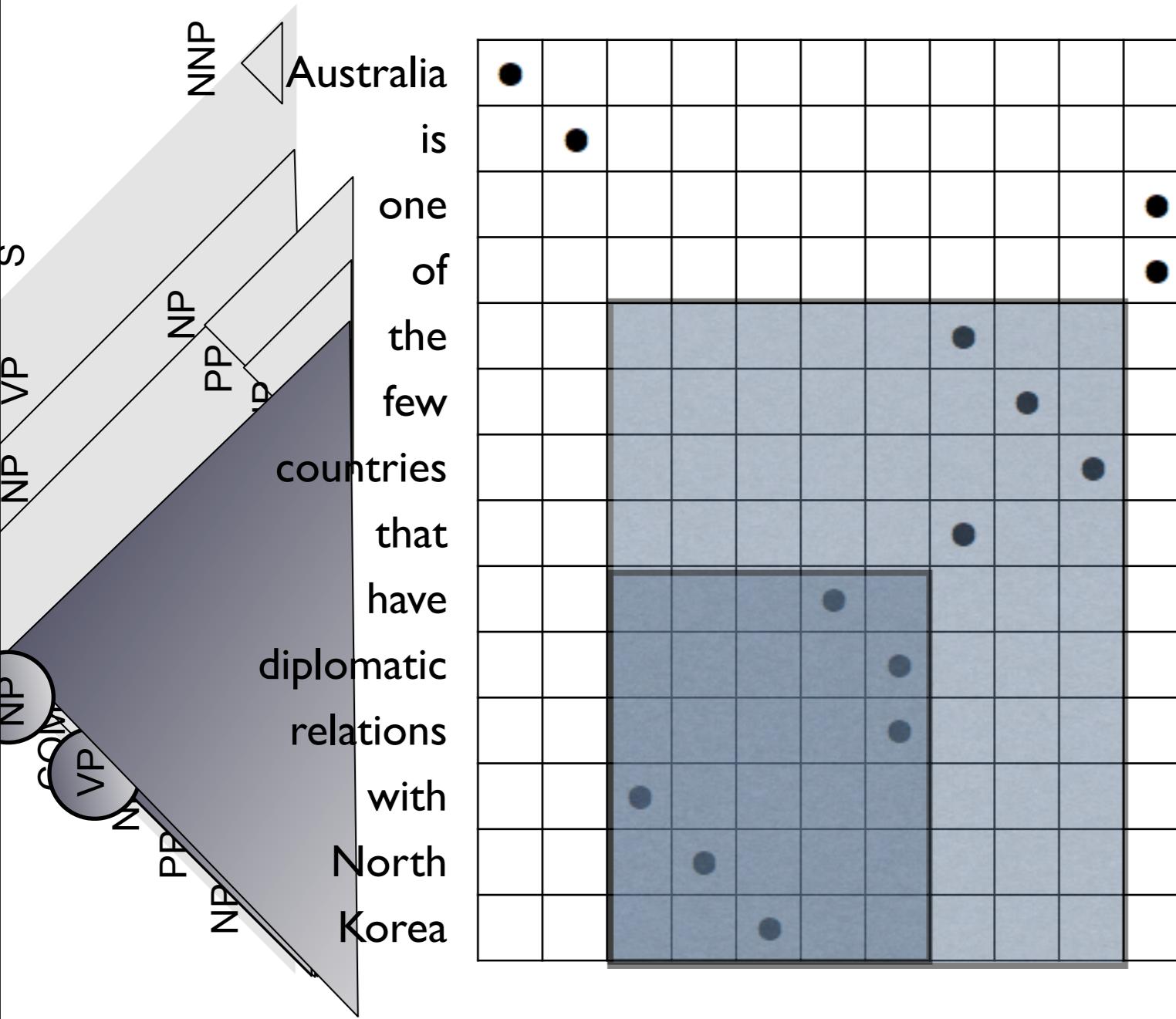
$NP \rightarrow VP$  的 少 数 国 家,  
the few countries that VP

$NP \rightarrow VP$  的 NP,  
the NP that VP

# Wait a minute...

- Didn't we see this earlier in Koehn's paper?
- Aren't we giving up a ton of rules that you said were valuable?
- Something about a reduced inventory because we got rid of non-constituent phrases?

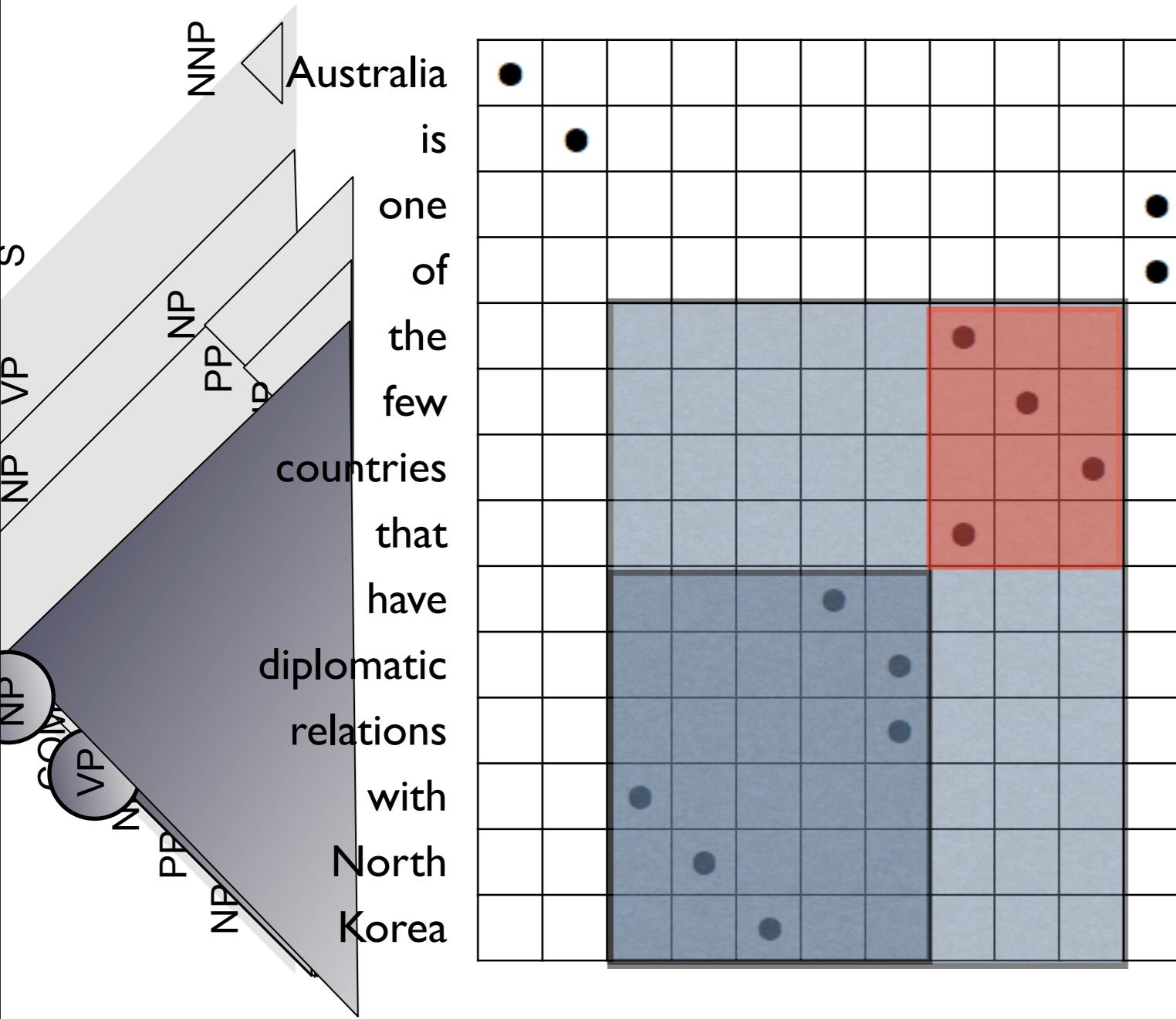
# Extracting Syntactic Rules



$VP \rightarrow \text{与北韩有邦交}$ ,  
have diplomatic relations  
with North Korea

$NP \rightarrow \text{与北韩有邦交}$   
的 少数 国家, the few  
countries that have  
diplomatic relations with  
North Korea

# Extracting Syntactic Rules

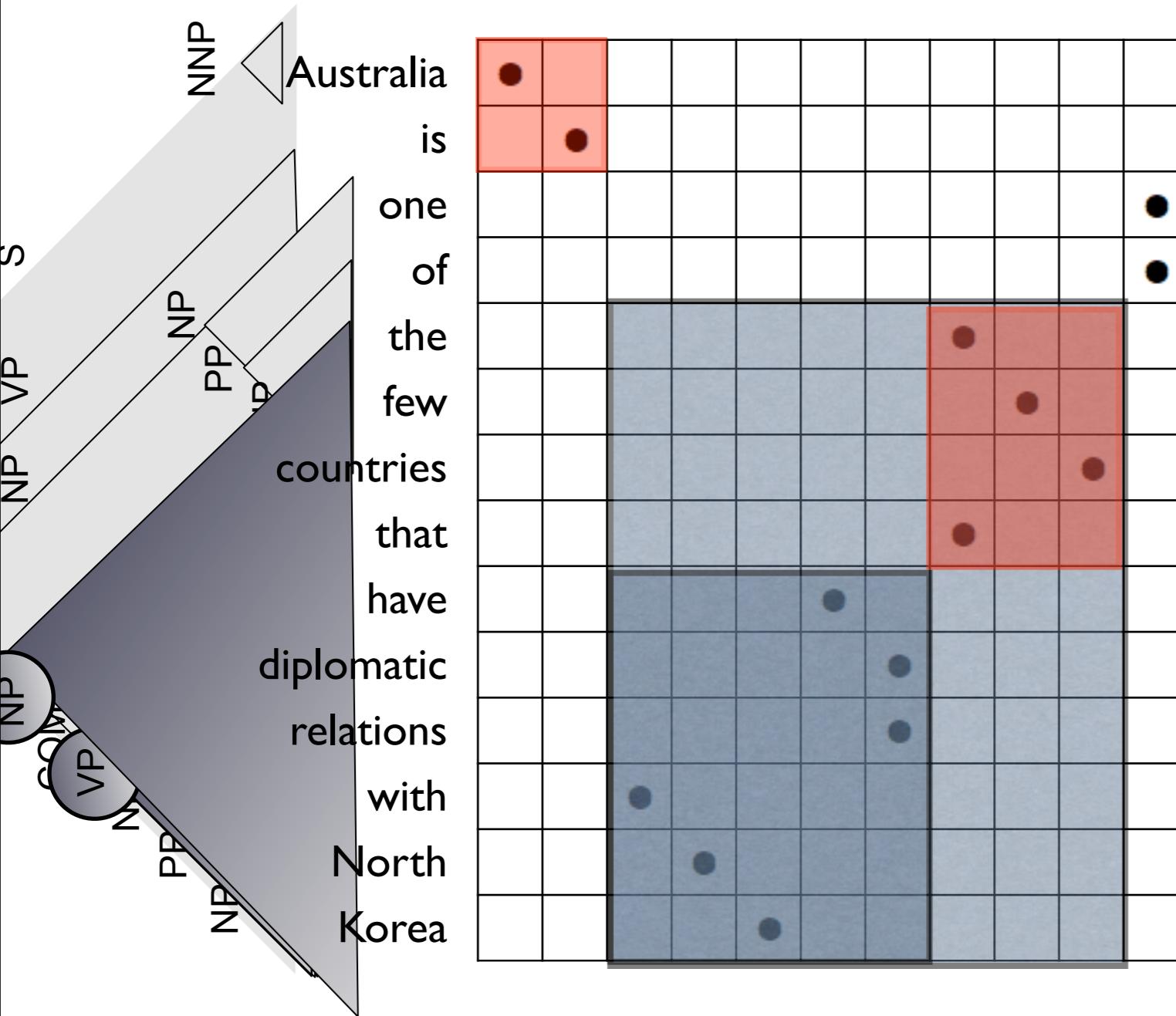


$VP \rightarrow \text{与北韩有邦交}$ ,  
have diplomatic relations  
with North Korea

$NP \rightarrow \text{与北韩有邦交}$   
的 少数 国家, the few  
countries that have  
diplomatic relations with  
North Korea

??? → 的 少数 国家,  
the few countries that

# Extracting Syntactic Rules



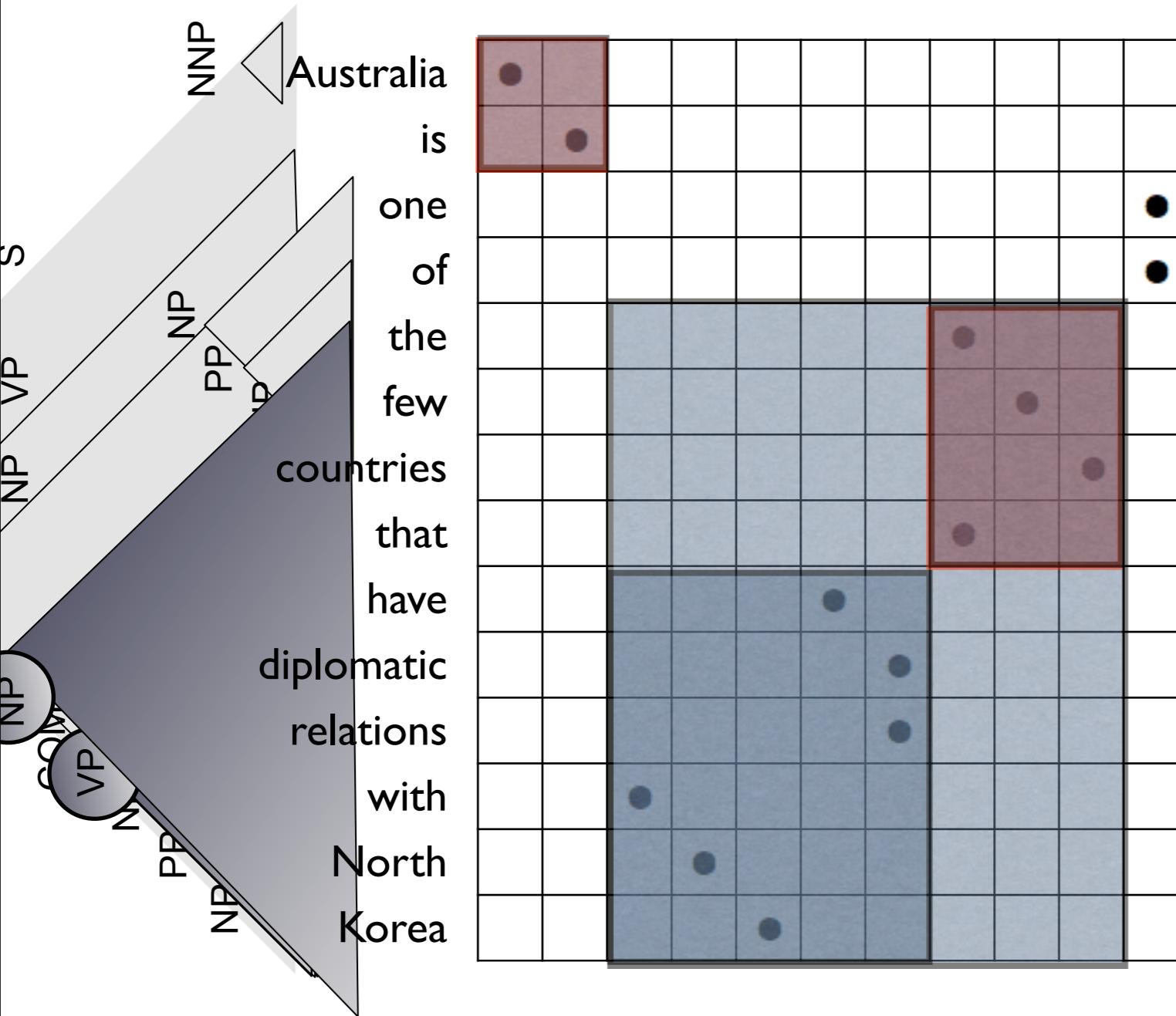
$VP \rightarrow \text{与北韩有邦交}$ ,  
have diplomatic relations  
with North Korea

$NP \rightarrow \text{与北韩有邦交}$   
的少数国家, the few  
countries that have  
diplomatic relations with  
North Korea

??? → 的少数国家,  
the few countries that

??? → 澳洲是,  
Australia is

# Extracting Syntactic Rules



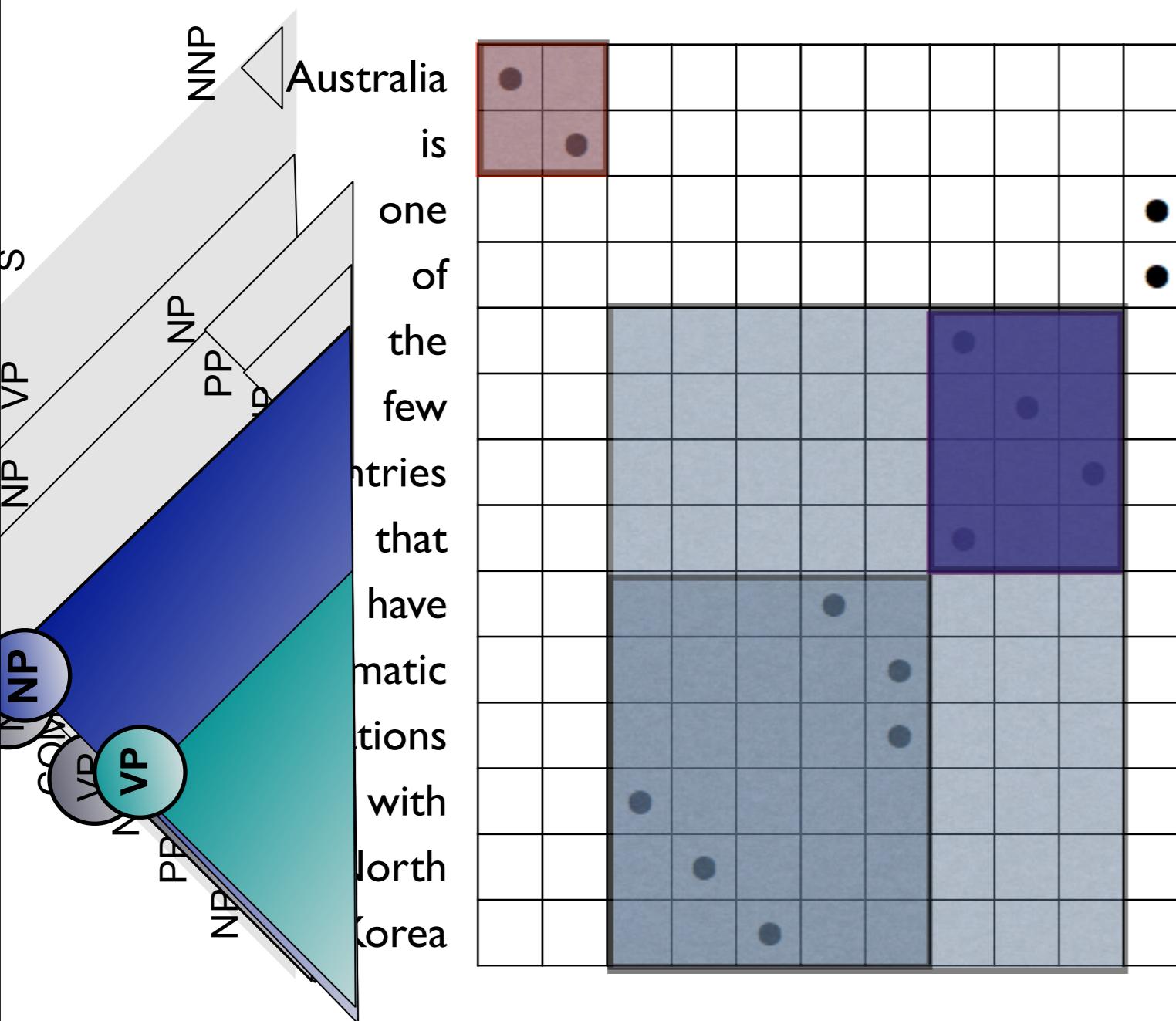
$VP \rightarrow \text{与 北 韩 有 邦 交}$ ,  
have diplomatic relations  
with North Korea

$NP \rightarrow \text{与 北 韩 有 邦 交}$   
的 少数 国家, the few  
countries that have  
diplomatic relations with  
North Korea

??? → 的 少数 国家,  
the few countries that

??? → 澳洲 是,  
Australia is

# Extracting Syntactic Rules



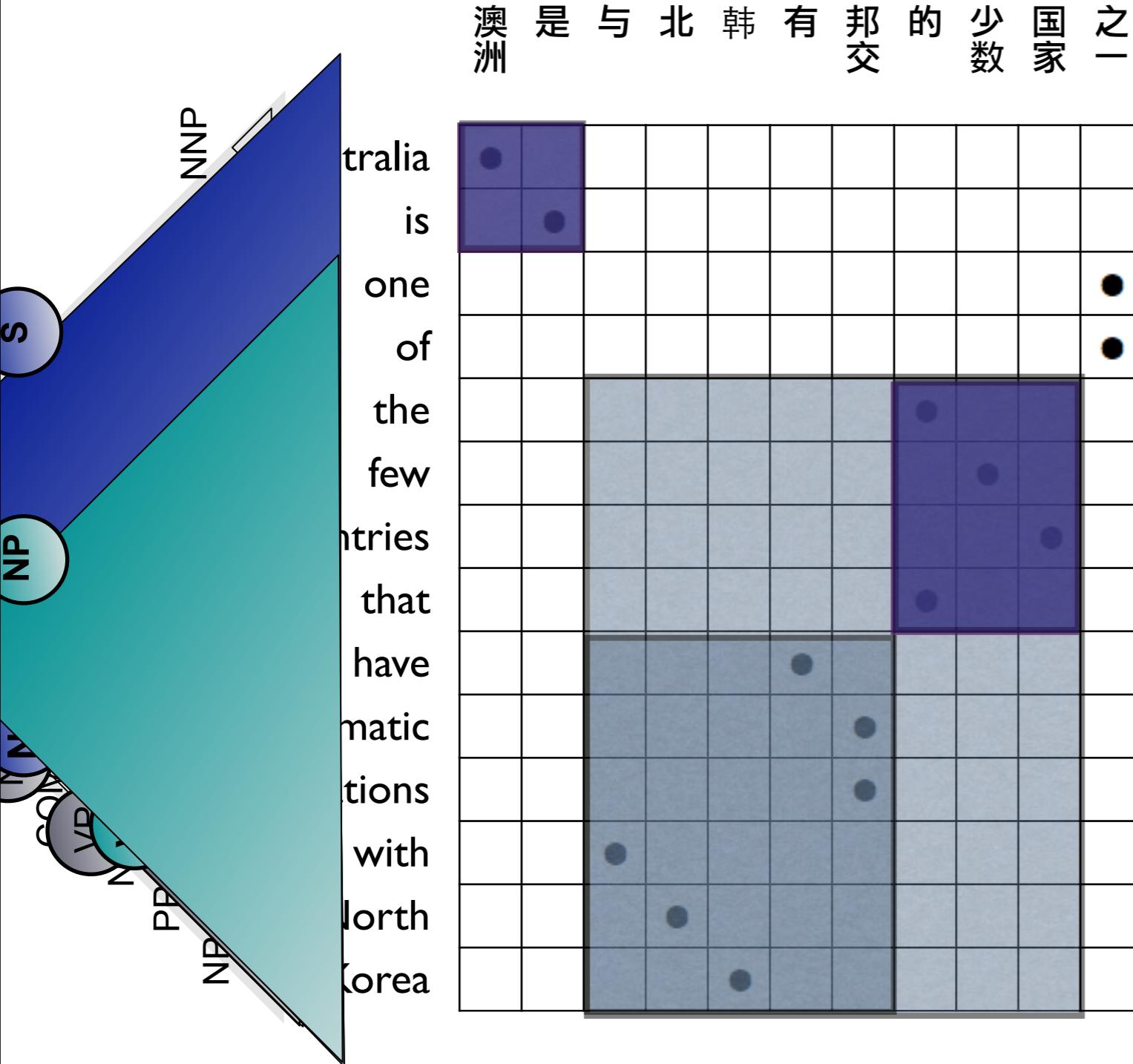
VP → 与 北 韩 有 邦 交,  
have diplomatic relations  
with North Korea

NP → 与 北 韩 有 邦 交  
的 少 数 国 家, the few  
countries that have  
diplomatic relations with  
North Korea

NP/ VP → 约 少 数 国 家,  
the few countries that

??? → 澳 洲 是,  
Australia is

# Extracting Syntactic Rules



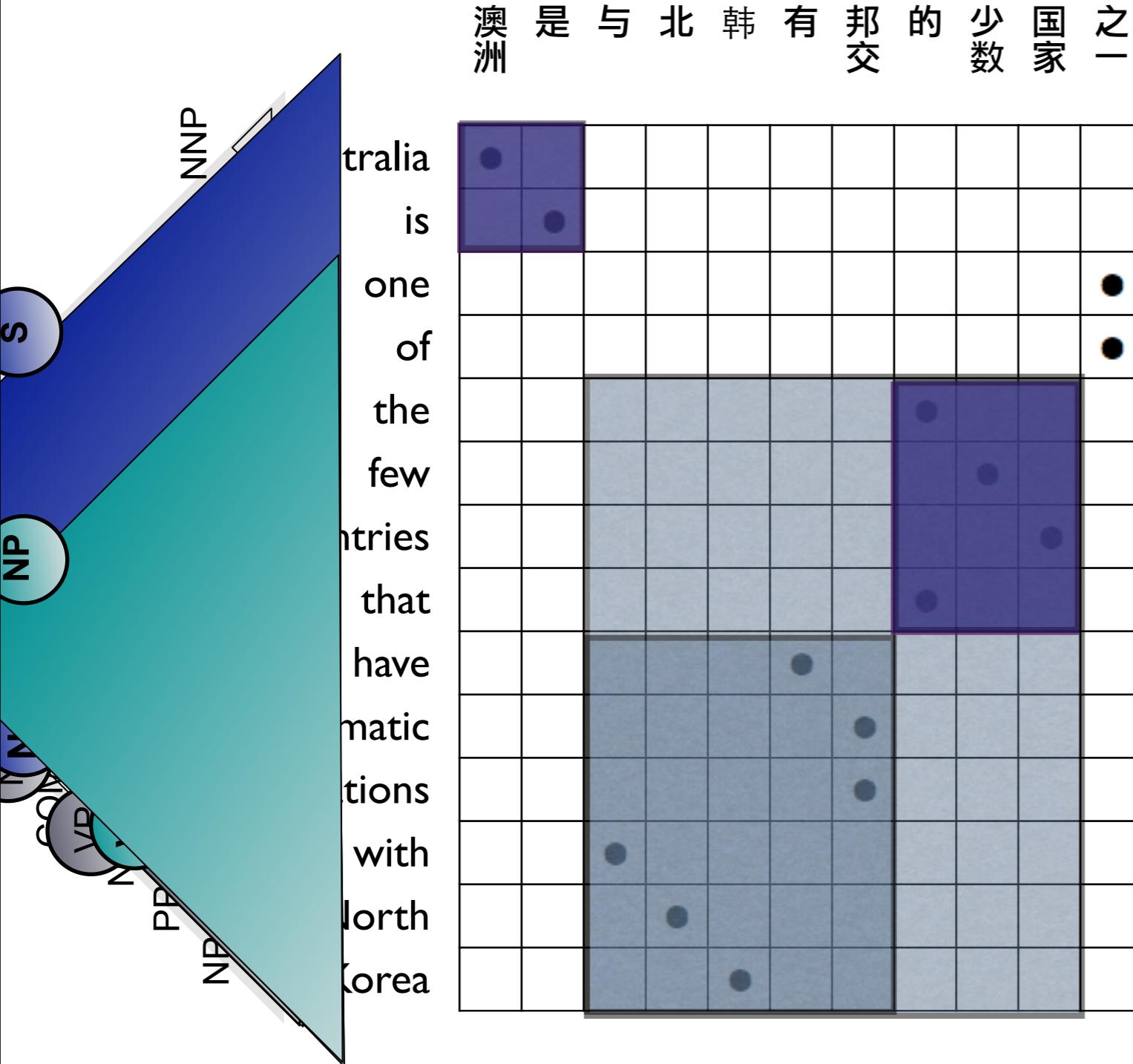
VP → 与 北 韩 有 邦 交,  
have diplomatic relations  
with North Korea

NP → 与 北 韩 有 邦 交  
的 少 数 国 家, the few  
countries that have  
diplomatic relations with  
North Korea

NP/ VP → 约 少 数 国 家,  
the few countries that

S/ NP → 澳 洲 是,  
Australia is

# Extracting Syntactic Rules



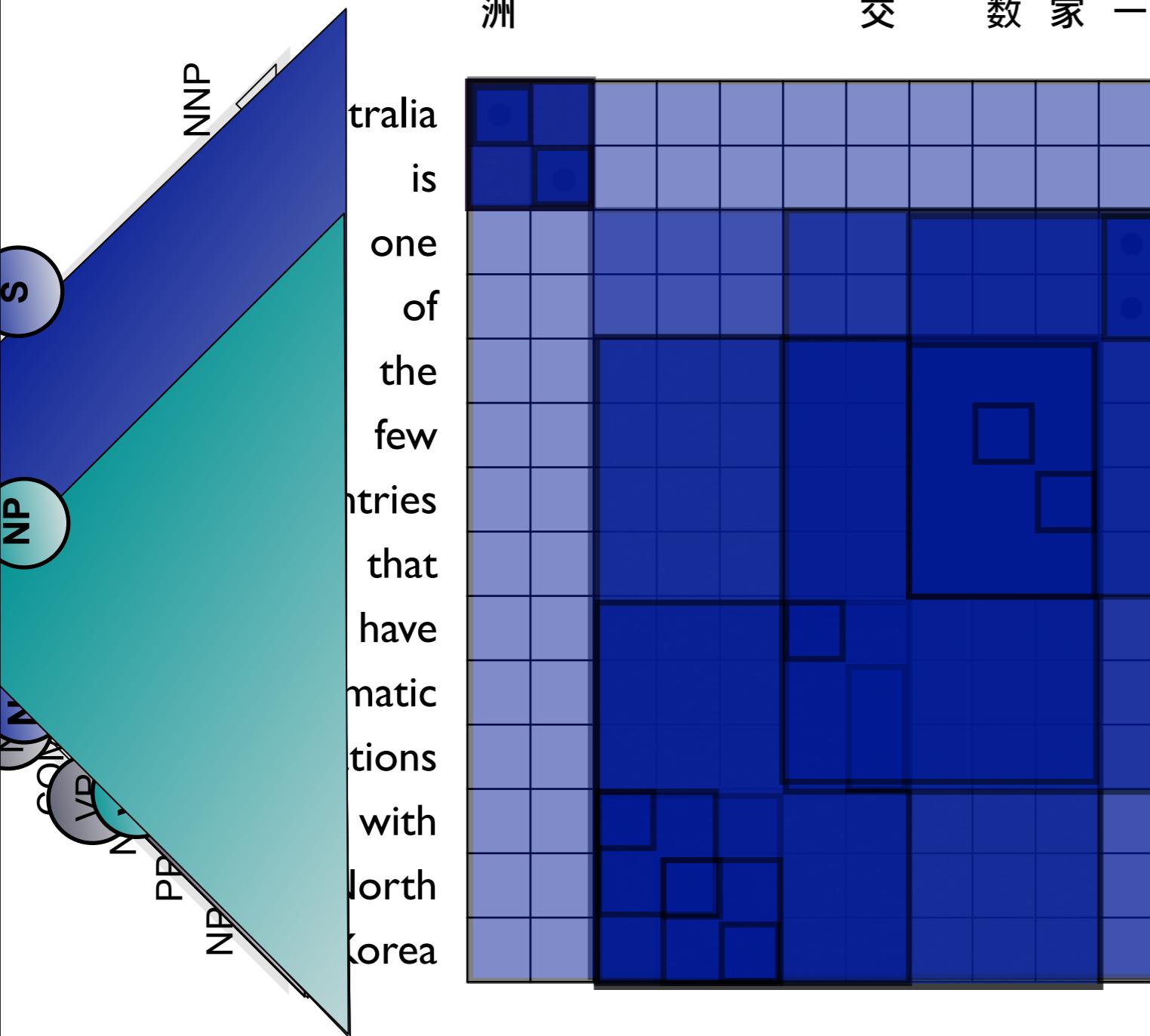
VP → 与 北 韩 有 邦 交,  
have diplomatic relations  
with North Korea

NP → 与 北 韩 有 邦 交  
的 少 数 国 家, the few  
countries that have  
diplomatic relations with  
North Korea

NP/ VP → 约 少 数 国 家,  
the few countries that

S/ NP → 澳 洲 是,  
Australia is

# Extracting Syntactic Rules



澳洲 是 与 北 韩 有 邦 交 的 少 数 国 家 之

$VP \rightarrow \text{与 北 韩 有 邦 交}$ ,  
have diplomatic relations  
with North Korea

$NP \rightarrow \text{与 北 韩 有 邦 交}$   
的 少 数 国 家, the few  
countries that have  
diplomatic relations with  
North Korea

$NP/VP \rightarrow \text{的 少 数 国 家}$ ,  
the few countries that

$S/ NP \rightarrow \text{澳洲 是}$ ,  
Australia is

# Discussion: Is this better?

- What do you think of this flavor of SCFGs?
- What are its limitations?
- Do you think that it is better or worse than Hiero?
- How would you prove it?

(Discuss with your neighbors)

# New training paradigm

- Training data: word-aligned bilingual parallel corpus, with **parse trees**
  - No need to parse the Urdu, just parse the English
  - Method is therefore transferable to other resource poor languages
- Extract SCFG rules with **syntactic nonterminals**
- For **non-constituent phrases** use CCG-style nonterminals
- **Same coverage** as Hiero model

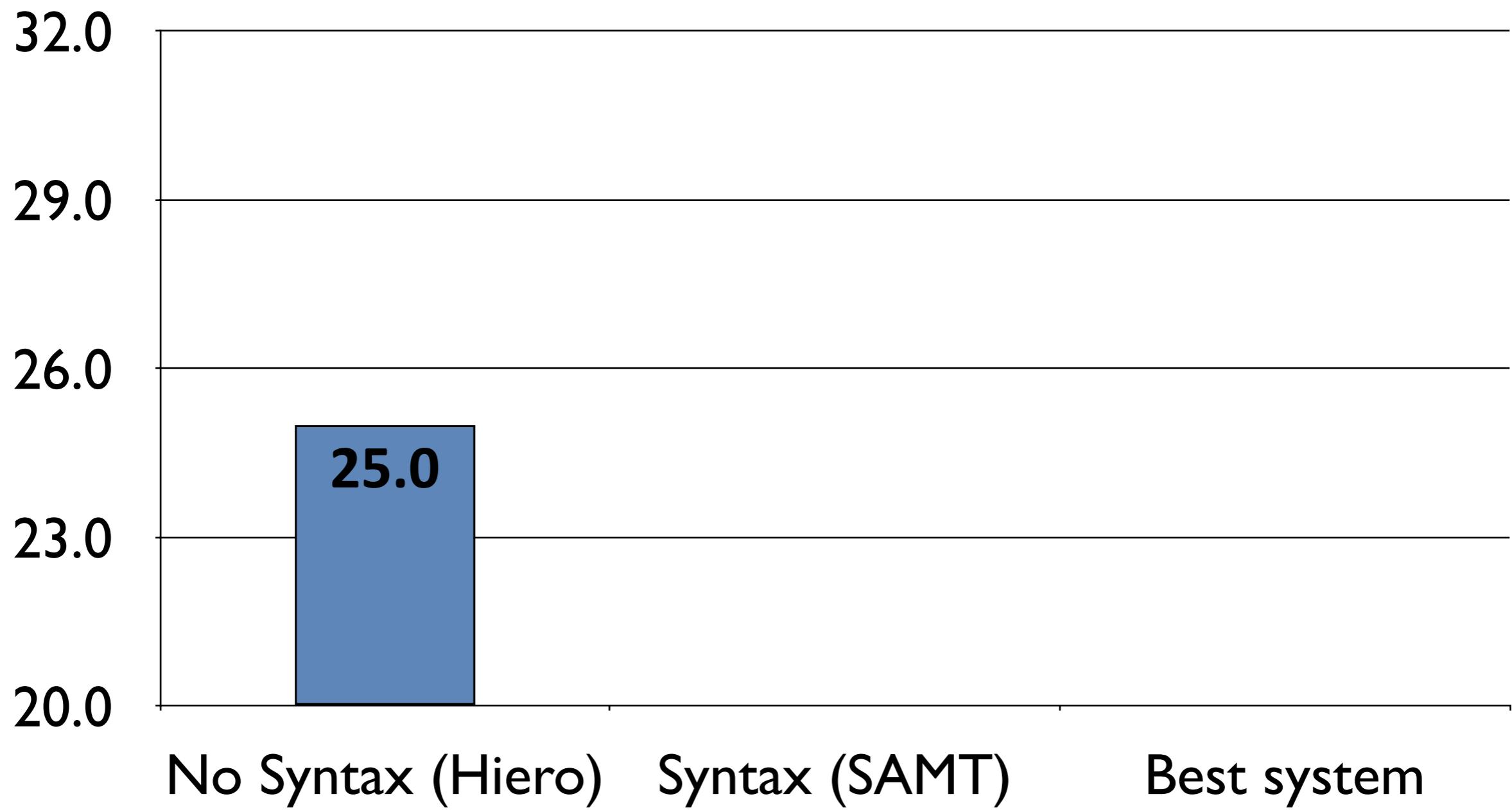
# Does it work?

- Tested for Urdu-English MT
- 1.5 Million word parallel corpus
- Two contrastive systems, with different grammar extraction mechanism
  - Hiero
  - Syntax-augmented grammars
- Used same decoder in both cases
- Tested results in a blind test set administered by the National Institute for Standards in Technology

# Syntax v. no Syntax

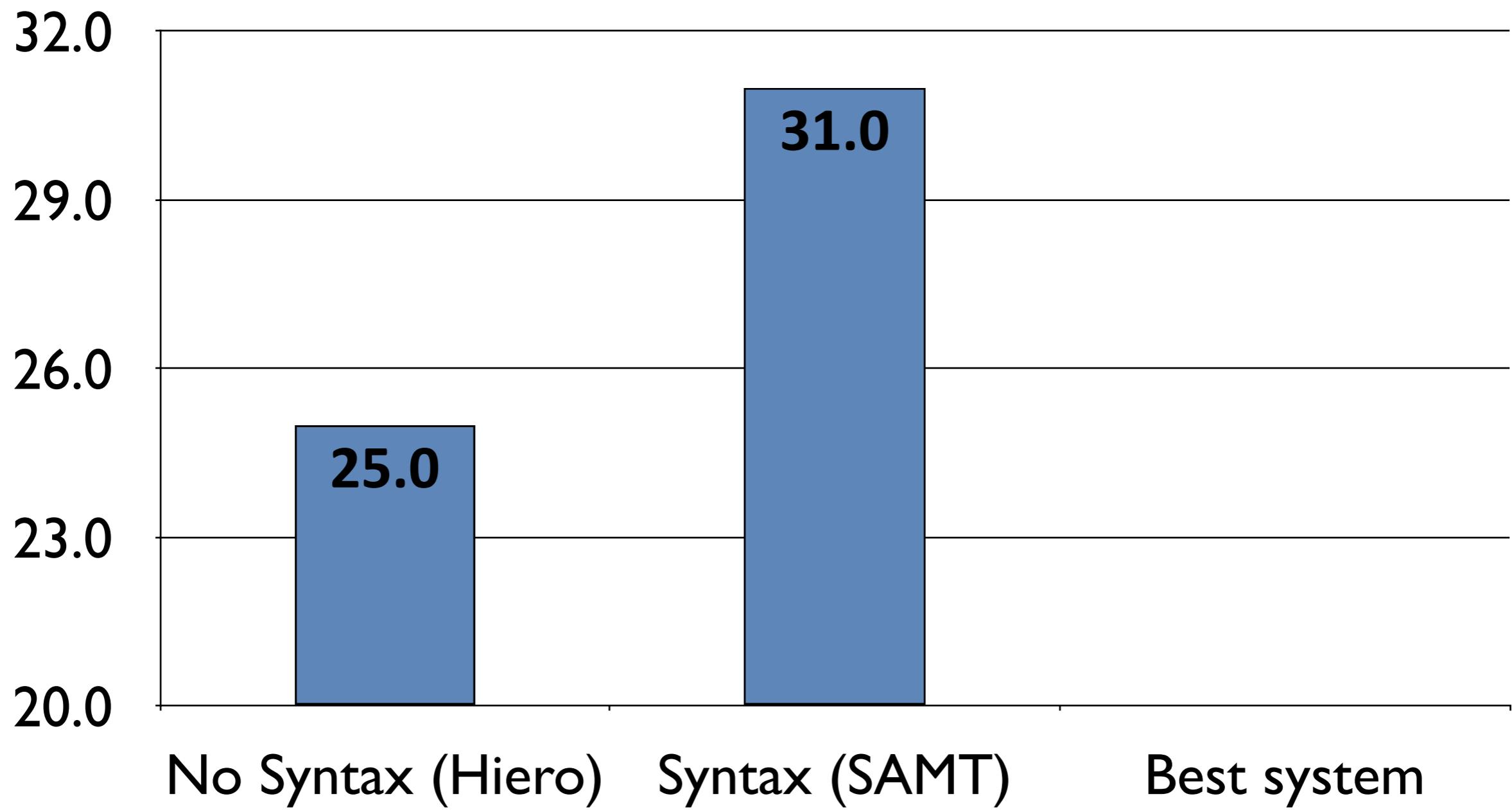
# Syntax v. no Syntax

Bleu score on blind NIST Urdu-English test set



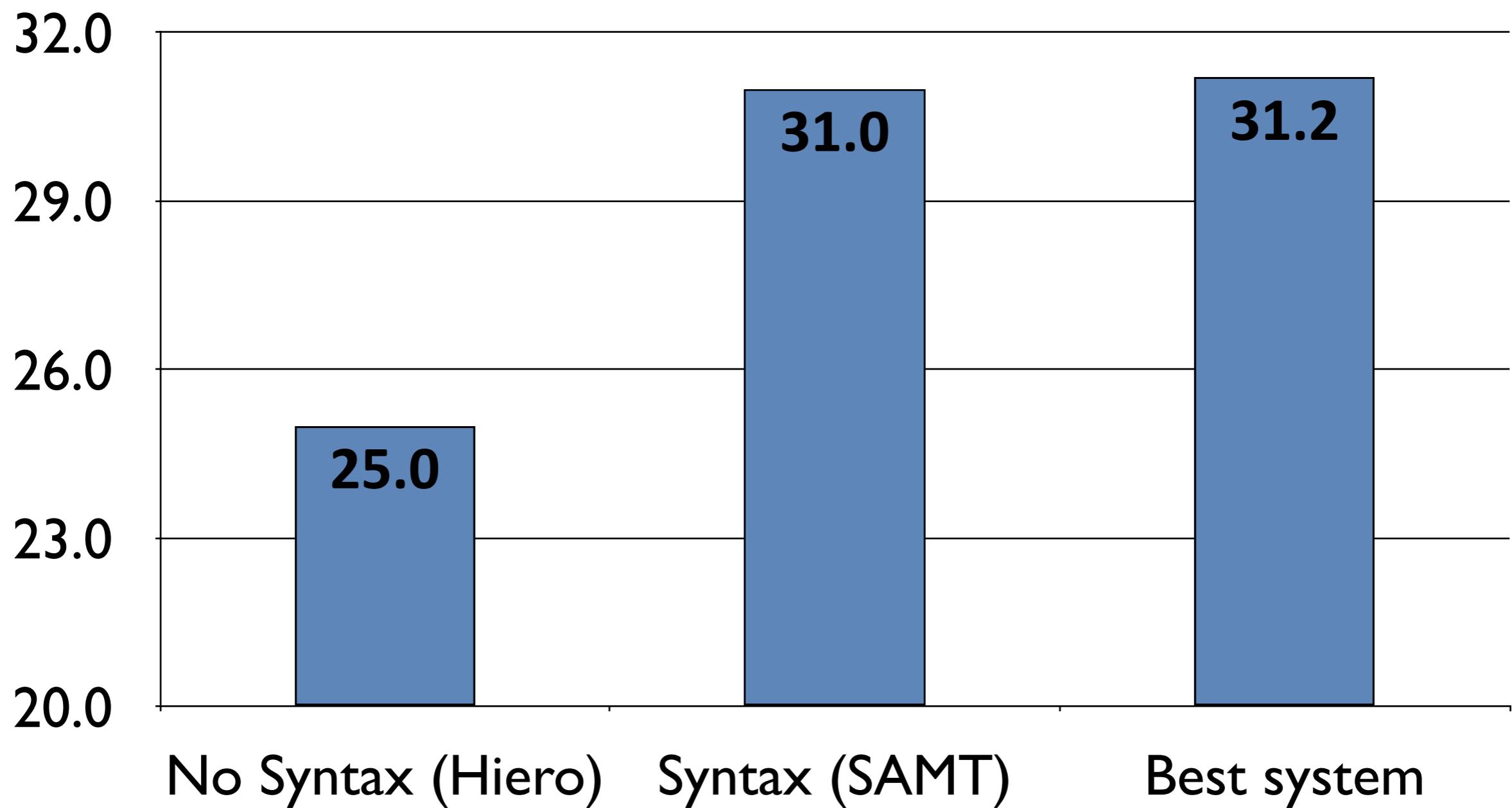
# Syntax v. no Syntax

Bleu score on blind NIST Urdu-English test set



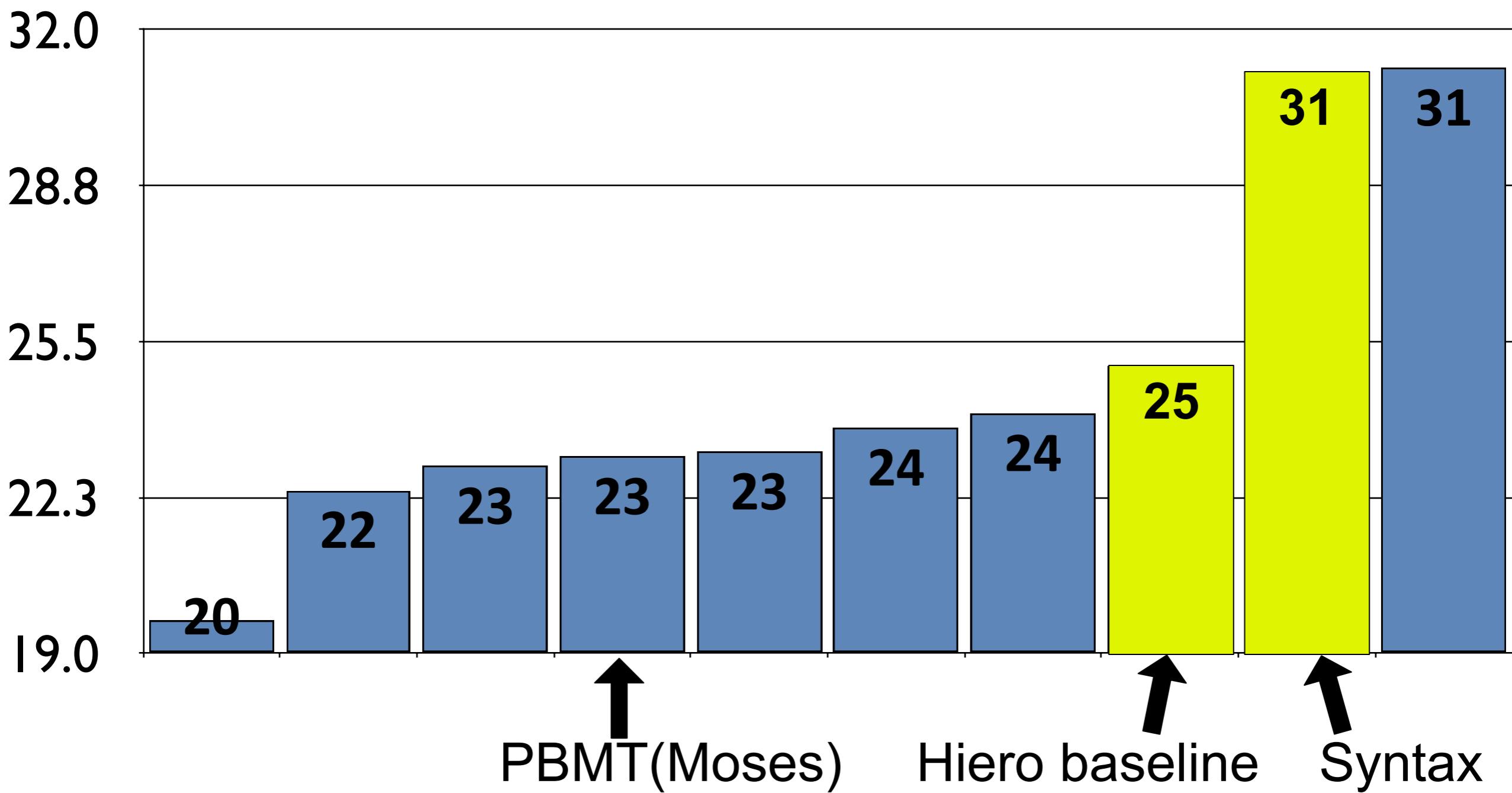
# Syntax v. no Syntax

Bleu score on blind NIST Urdu-English test set



# State of the Art Urdu Results

All system scores on NIST09 Urdu-English constrained task



# Translation improvements

'first nuclear experiment in 1990  
was'

Thomas red Unilever National Laboratory of the United States in designer، are already working on the book of Los ایلوموس National Laboratory ڈینی، former director of the technical انجینئرنگ written with the cooperation of سٹلمین.

This book 'nuclear express: political history and the expansion of bomb' has been written, and the two writers have also claimed that the country has made nuclear bomb is he or any other country's nuclear secrets to چرائے or that of any other nuclear power cooperation is achieved.

**The First Nuclear Test Was in 1990.**

Thomas red of the United States, the National Laboratory in designer are already working on the book of Los Alamos National Laboratory, former director of the technical intelligence, with the cooperation of Diana steelman wrote.

This book under the title of the spread of nuclear expressway: the political history of the bomb and this has been written and the two writers have claimed that the country also has made nuclear bomb or any other country, Korea nuclear secrets, or any of the other nuclear power cooperation.

# Who did what to whom?

## Baseline

He said that China, North Korea, Iran, Syria, Pakistan, through Egypt, Libya and Yemen is to provide nuclear technology.

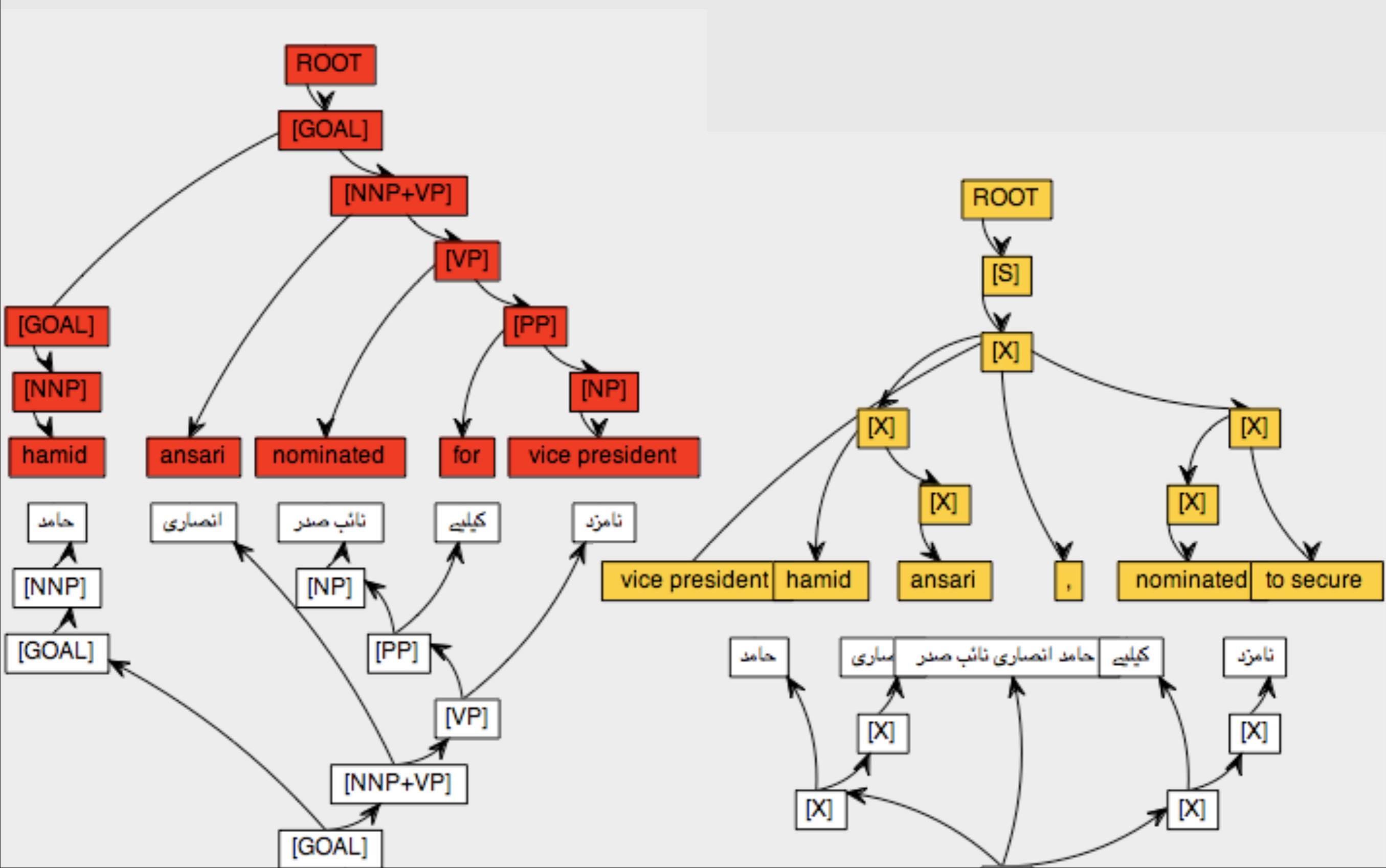
Thomas was red when this question why China has provided the nuclear technology to Pakistan, In response, He said as China and India was joint enemy of Pakistan.

## SCALE final system

He said that China would provide nuclear technology to North Korea, Iran, Syria, Pakistan, Egypt, Libya and Yemen.

Thomas red when was this question why China has provided to Pakistan nuclear technology, he said in response to China, Pakistan and India as a common enemy.

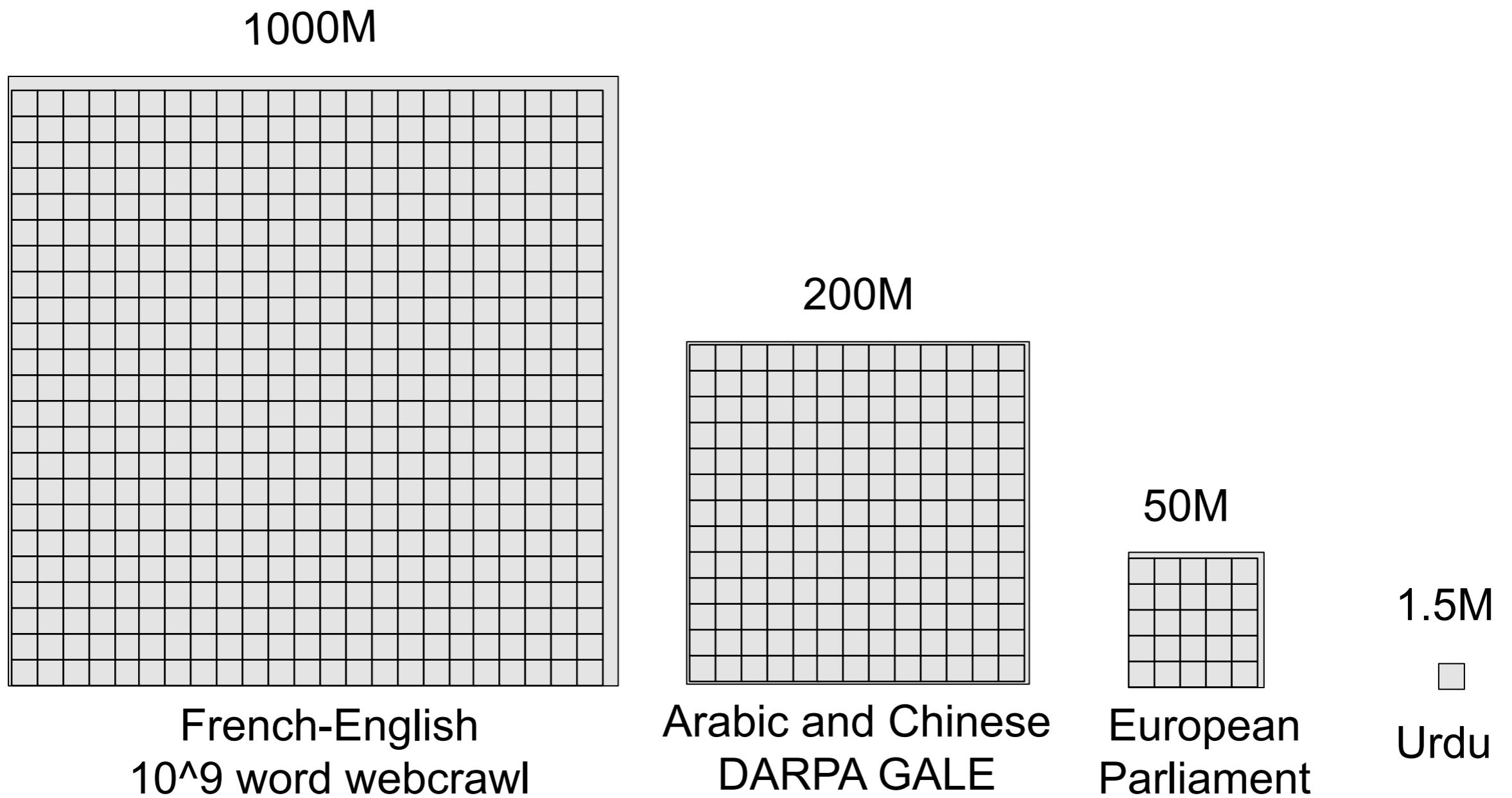
# Syntax captures Urdu reordering



# Why did this work?

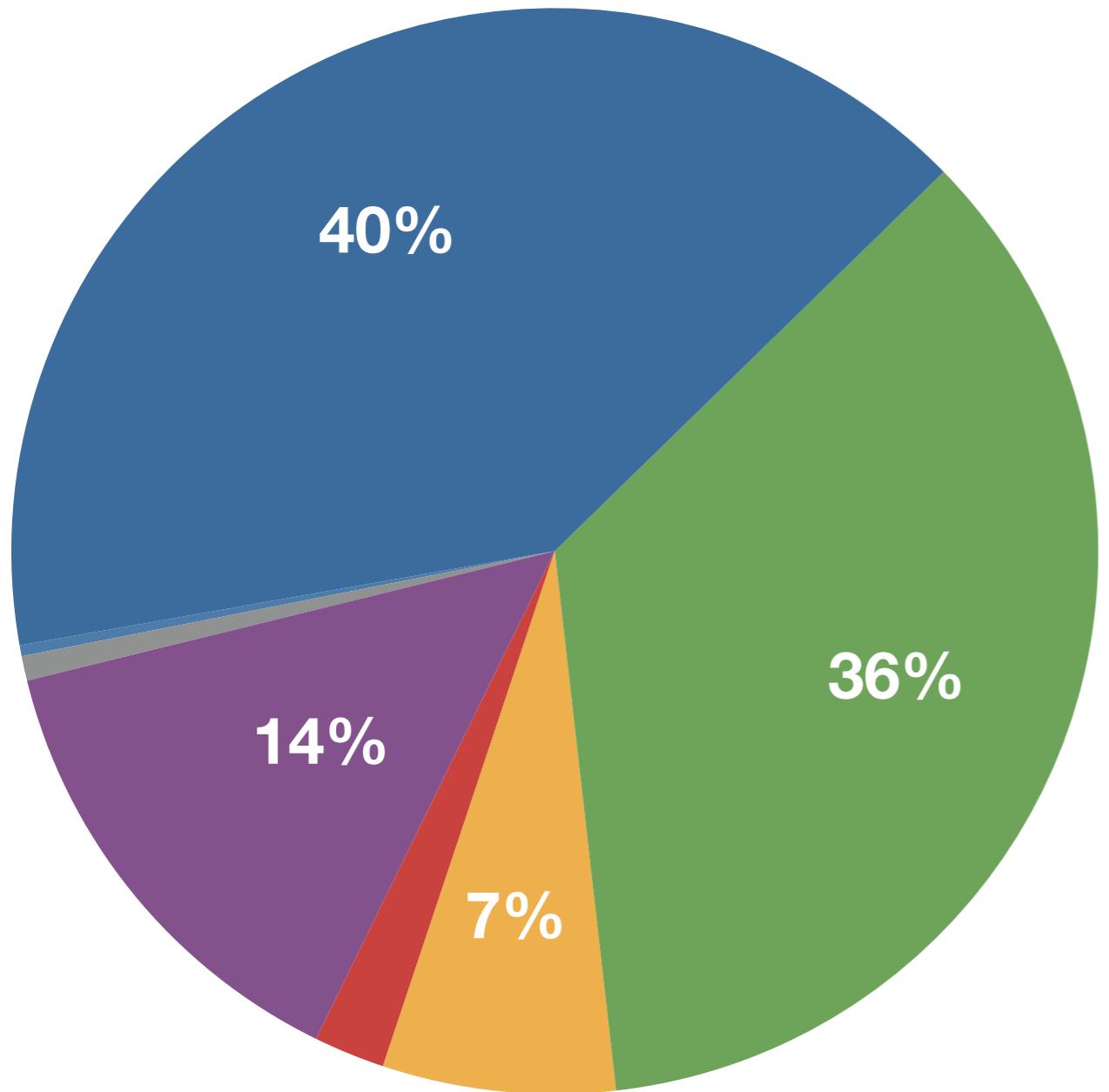
- Using **syntax-based translation models** resulted in huge improvements in quality
- Previous work on syntax did not show significant gains, so why did it work here?
- Urdu is an **ideal language** to show off the advantages of syntax
  - Very **small amount** of training data
  - Very **different word order** than English
- Can't simply **memorize** translations of phrases
- Must **generalize**

# Training data for MT Research



# Distribution of Word Orders

All Languages

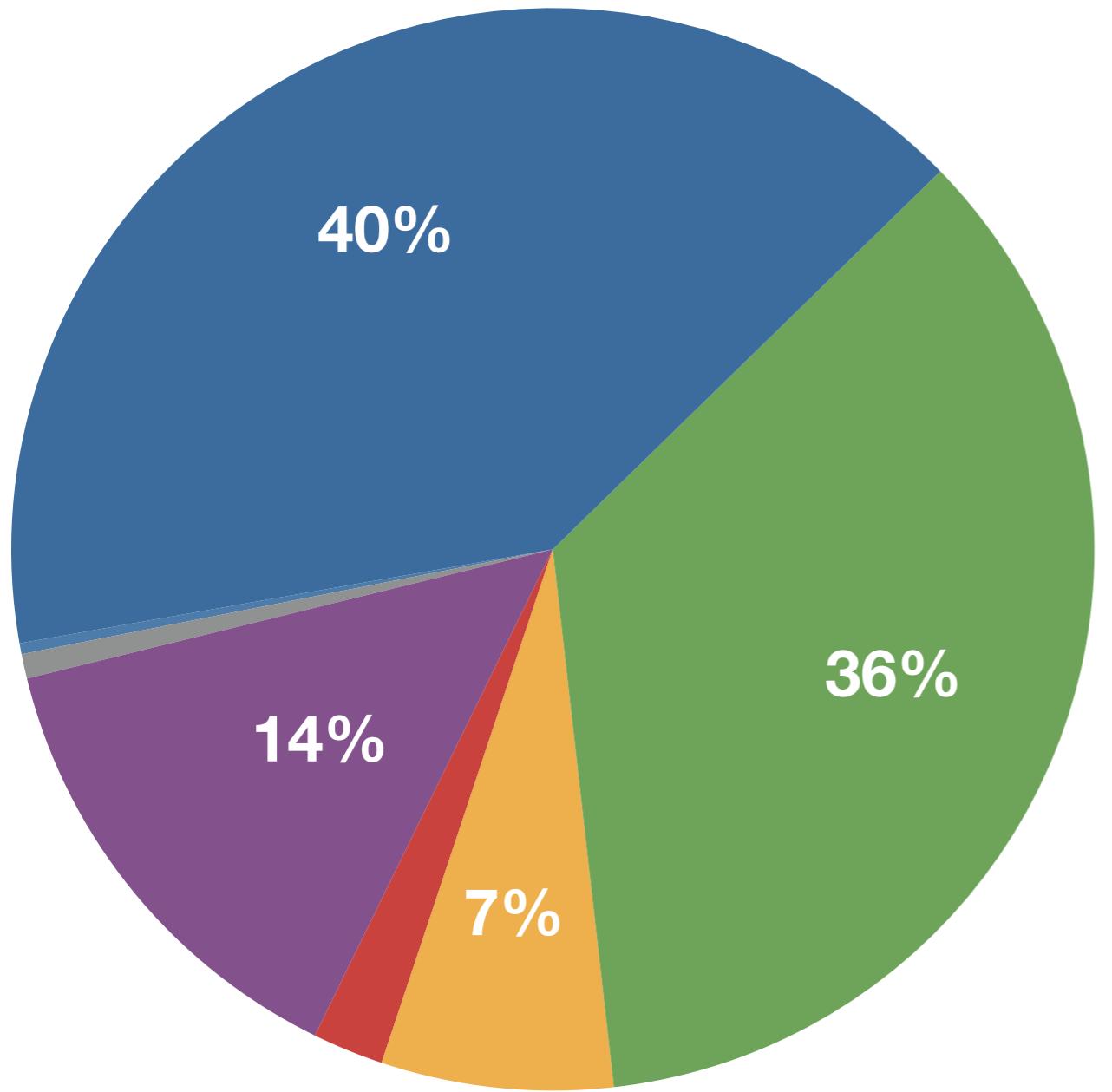


SMT Languages

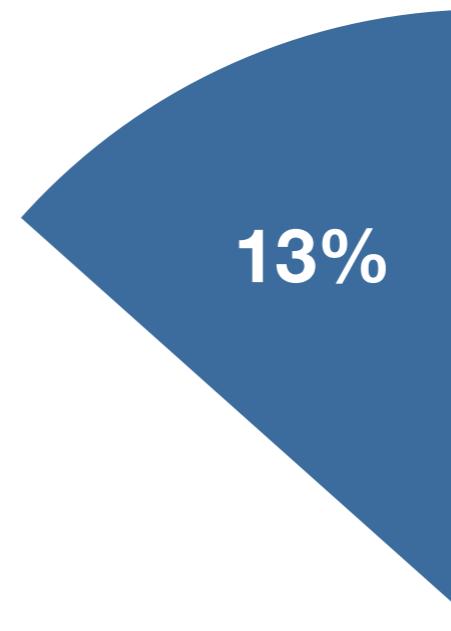
- SOV
- SVO
- VSO
- VOS
- No dominant order

# Distribution of Word Orders

All Languages



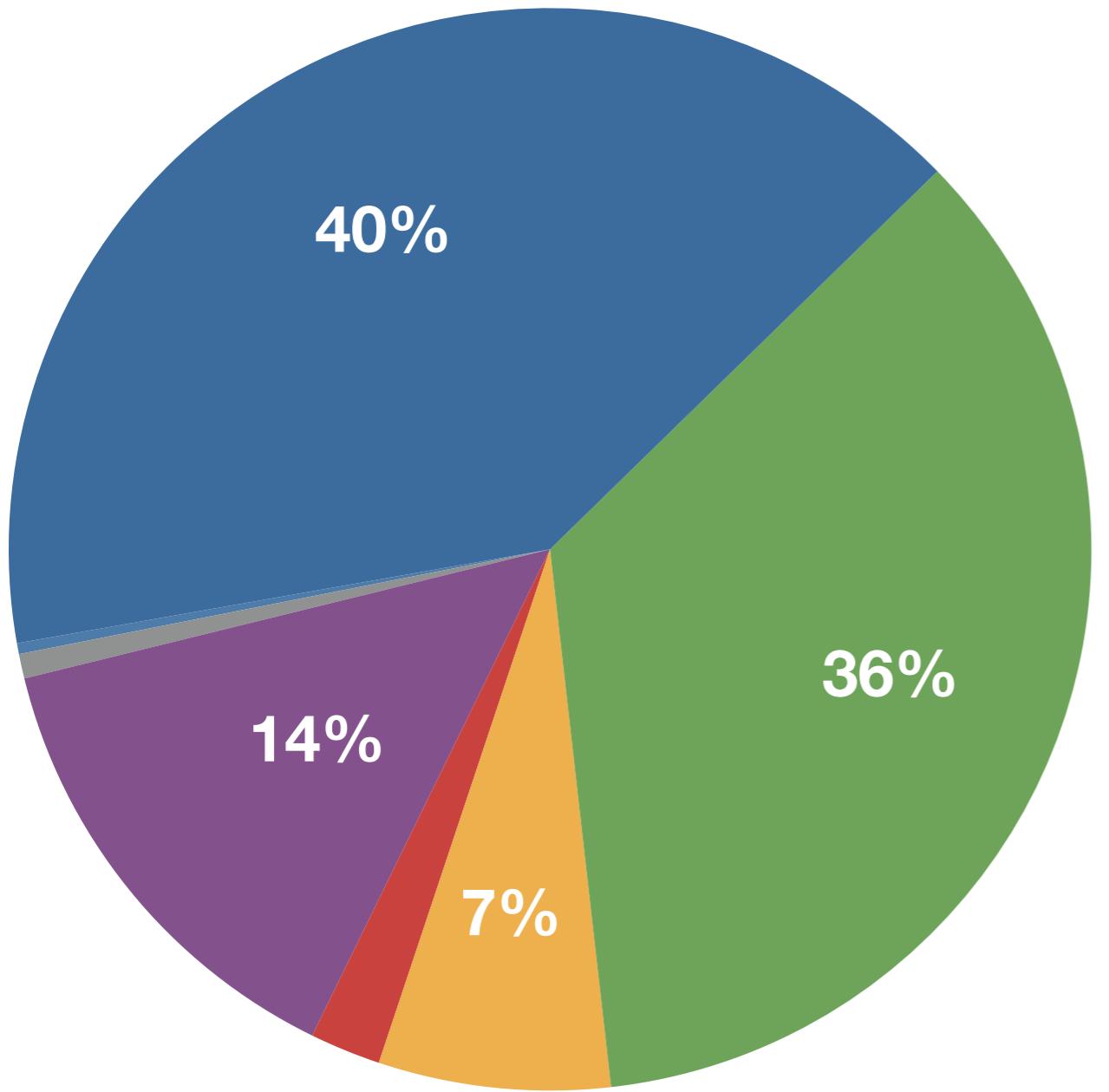
SMT Languages



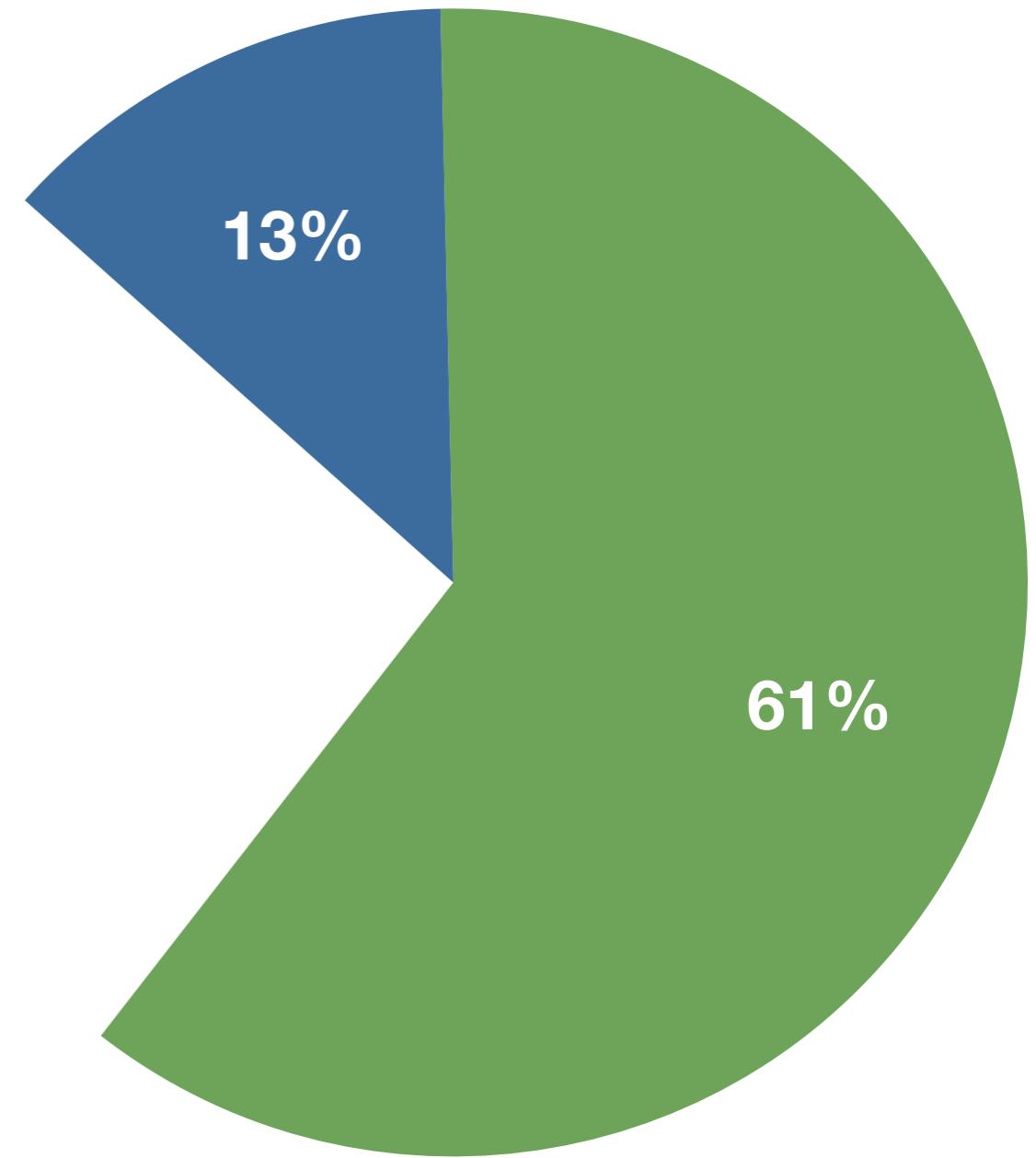
● SOV   ● SVO   ● VSO   ● VOS   ● No dominant order

# Distribution of Word Orders

All Languages



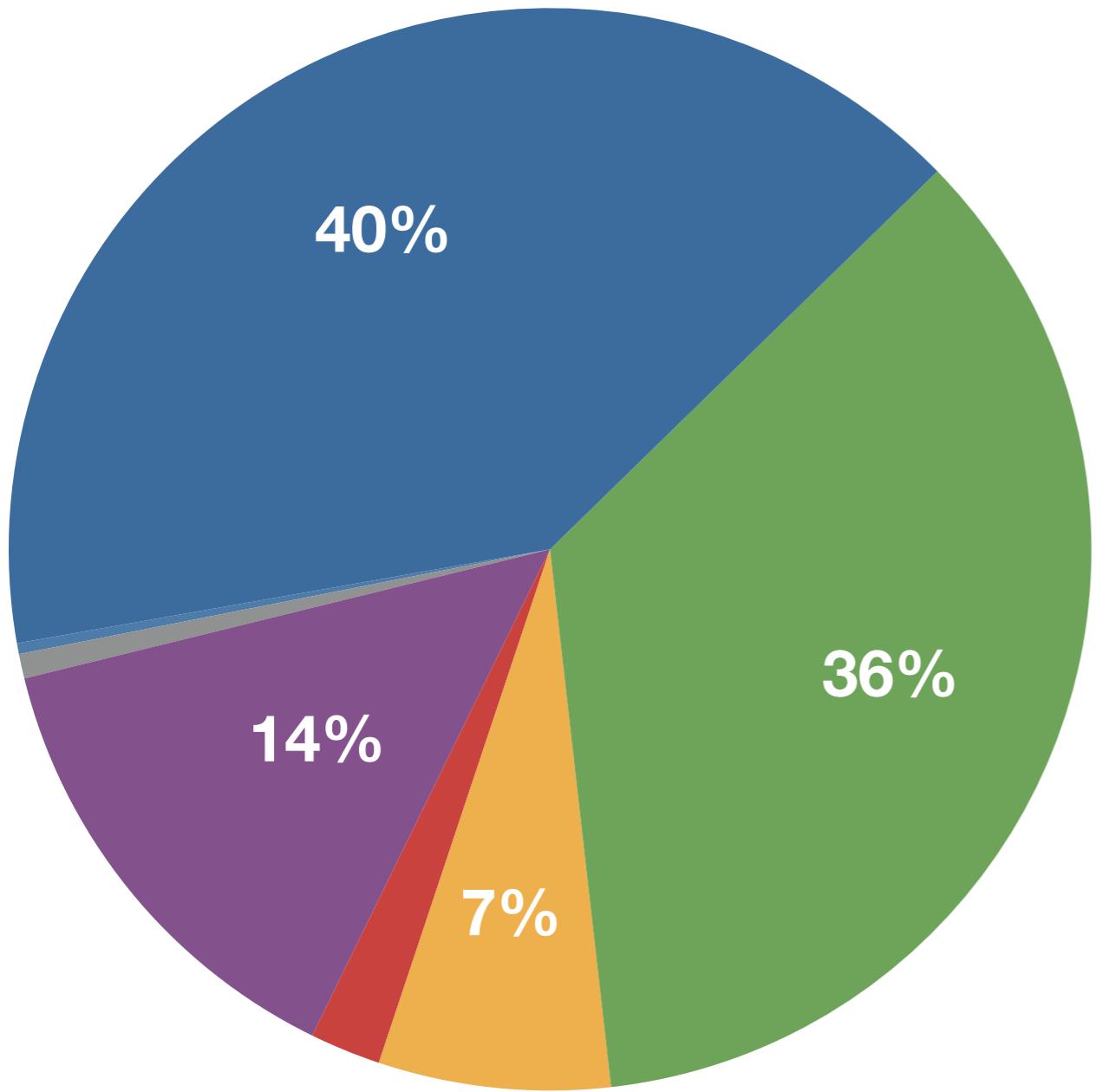
SMT Languages



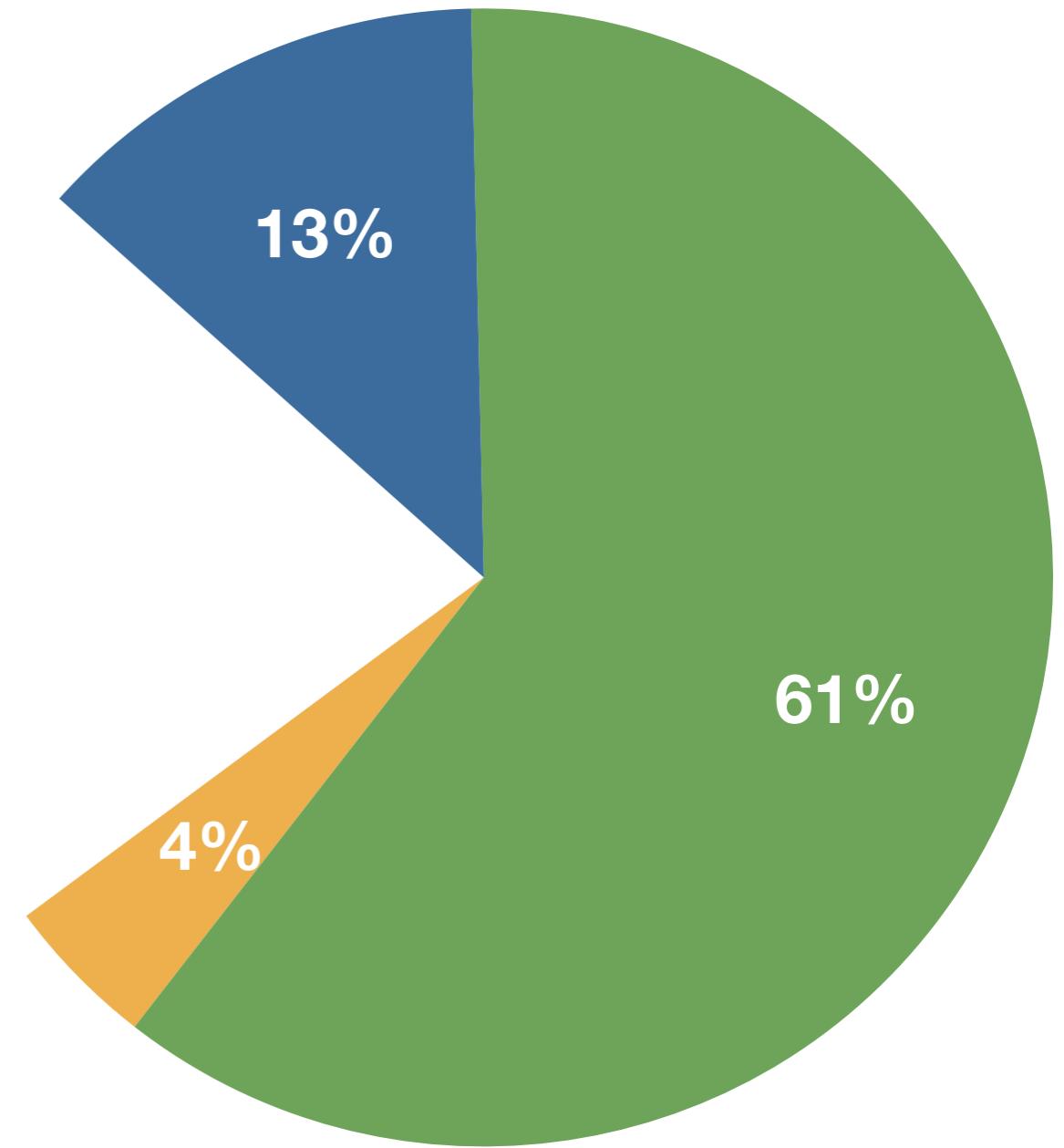
● SOV ● SVO ● VSO ● VOS ● No dominant order

# Distribution of Word Orders

All Languages



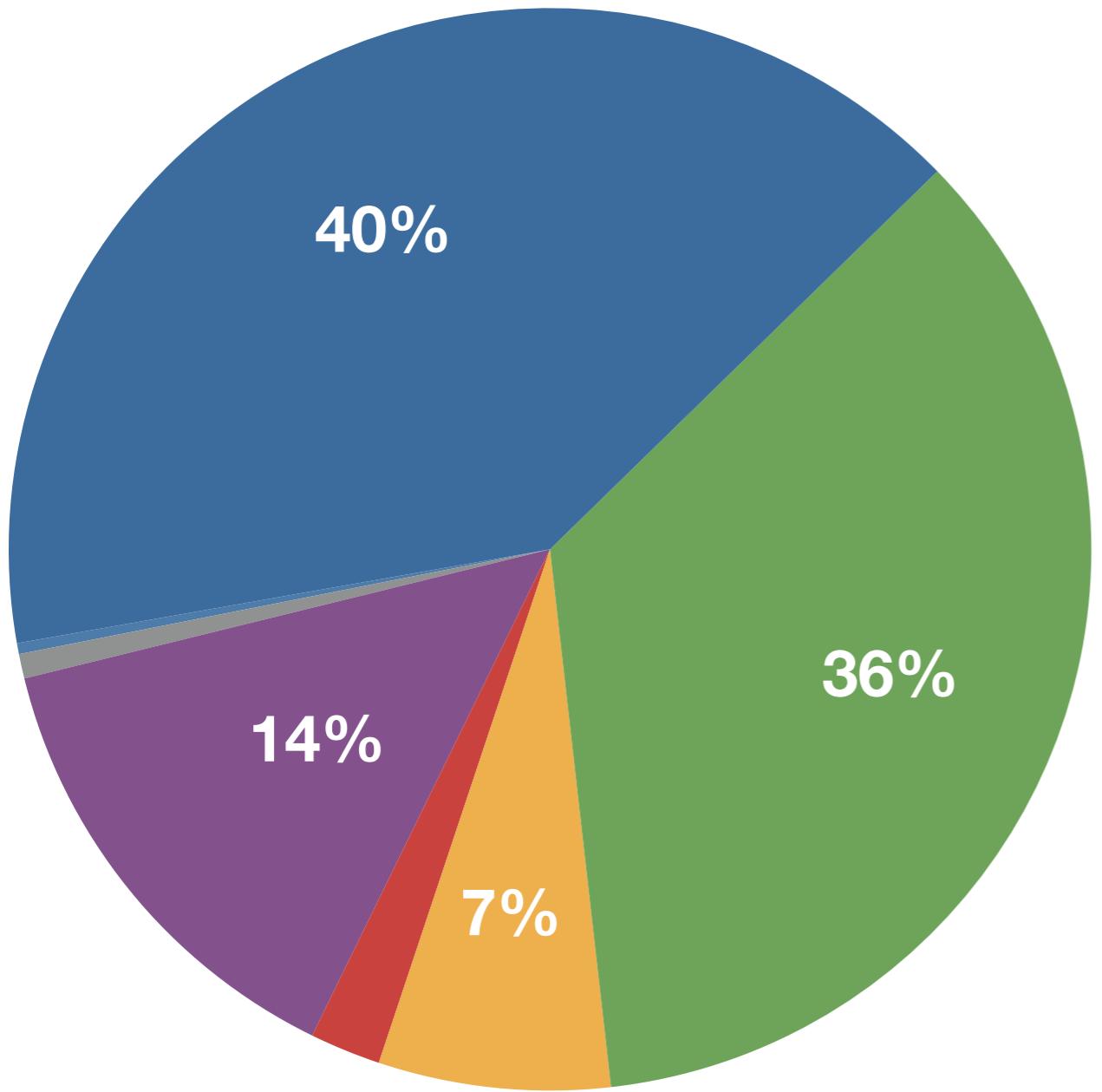
SMT Languages



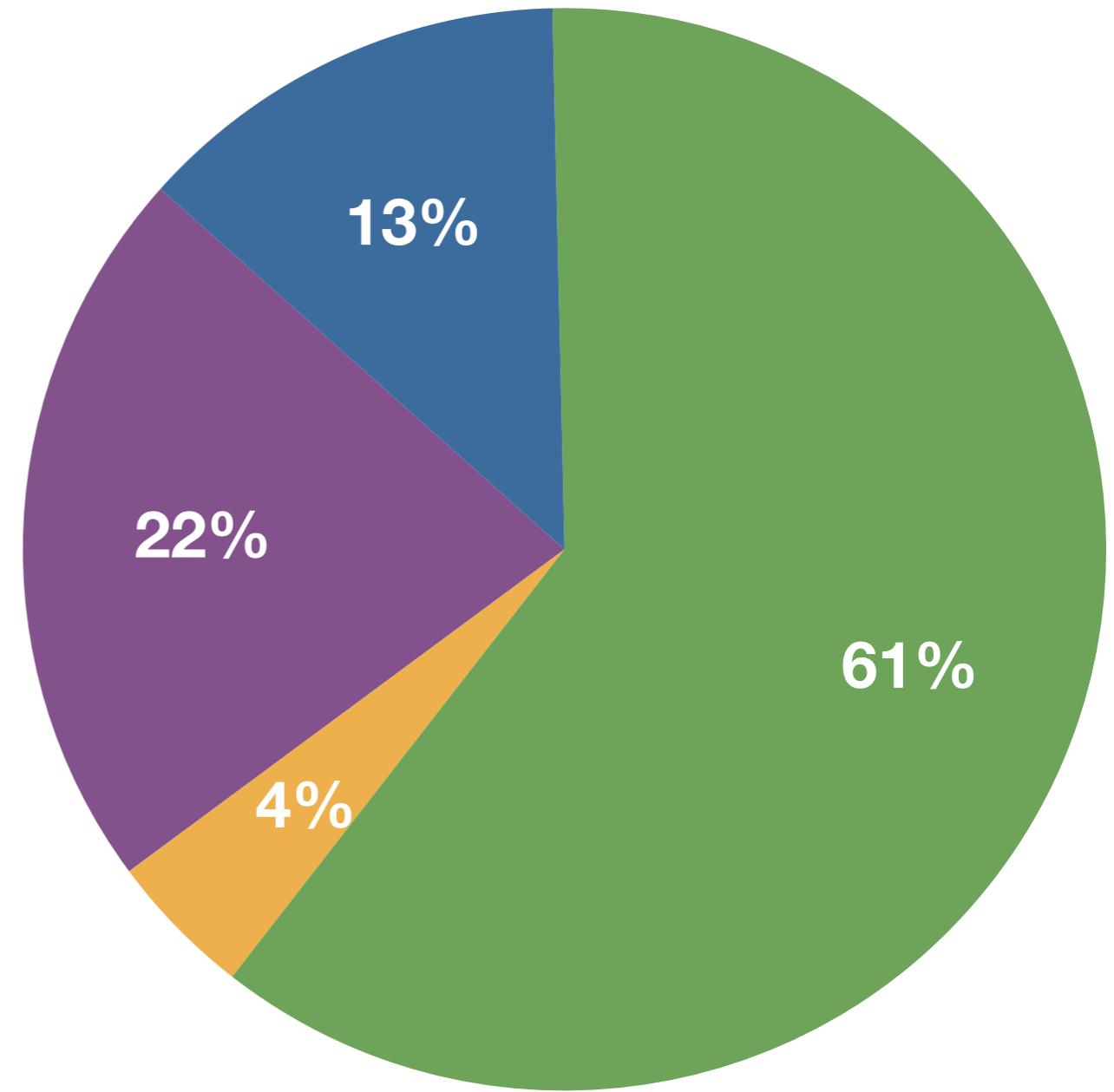
● SOV ● SVO ● VSO ● VOS ● No dominant order

# Distribution of Word Orders

All Languages



SMT Languages



● SOV ● SVO ● VSO ● VOS ● No dominant order

# Joshua Decoder



- An open source decoder
- Uses synchronous context free grammars to translate
- Implements all algorithms needed for translating with SCFGs
  - grammar extraction (Thrax!)
  - chart-parsing
  - n-gram language model integration
  - pruning, and k-best extraction<sup>51</sup>

# Joshua Decoder

- Download it from
  - <http://joshua-decoder.org>



# Joshua Decoder

- Download it from
  - <http://joshua-decoder.org>
- Brownie points if you use it in your final projects

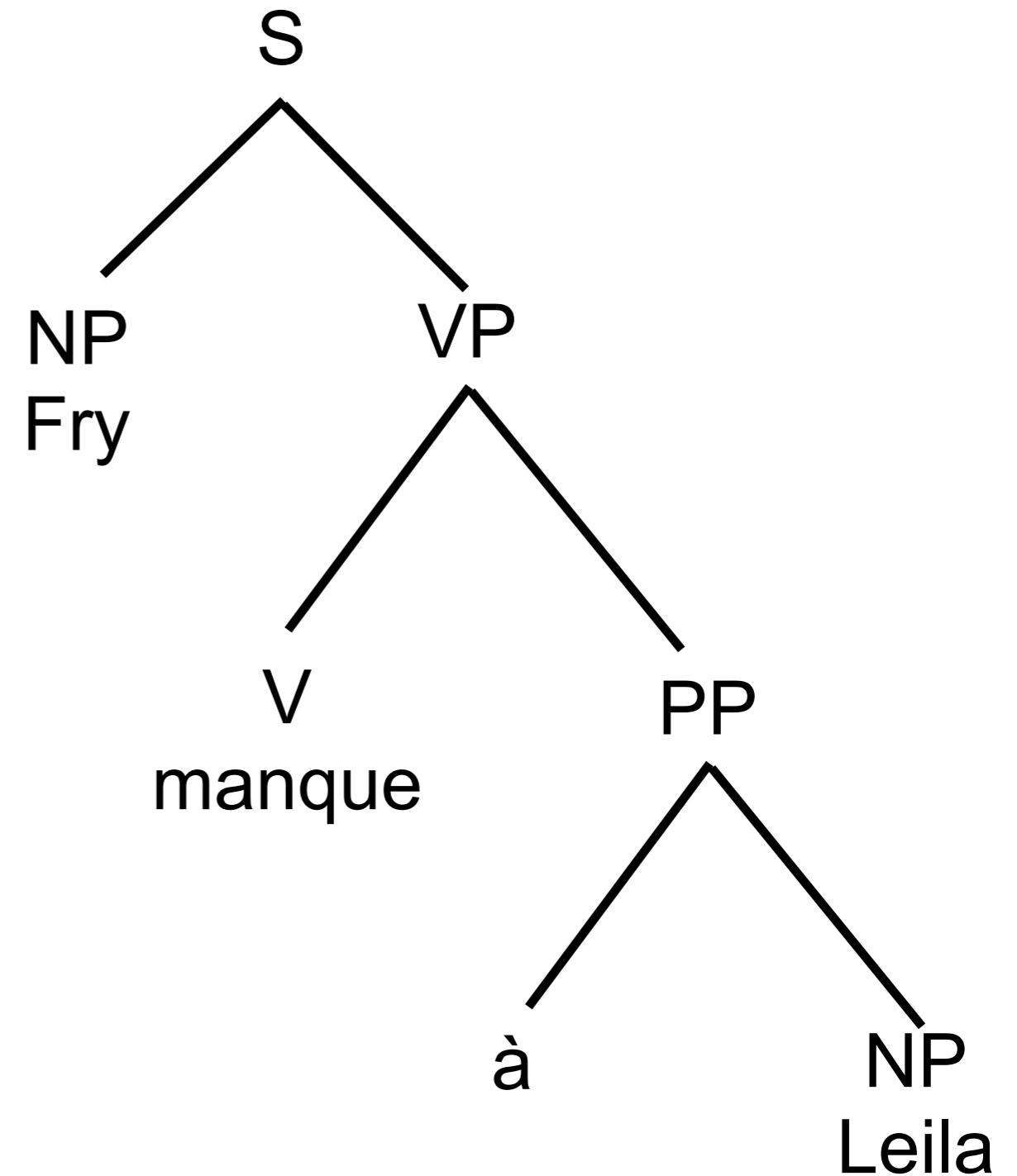
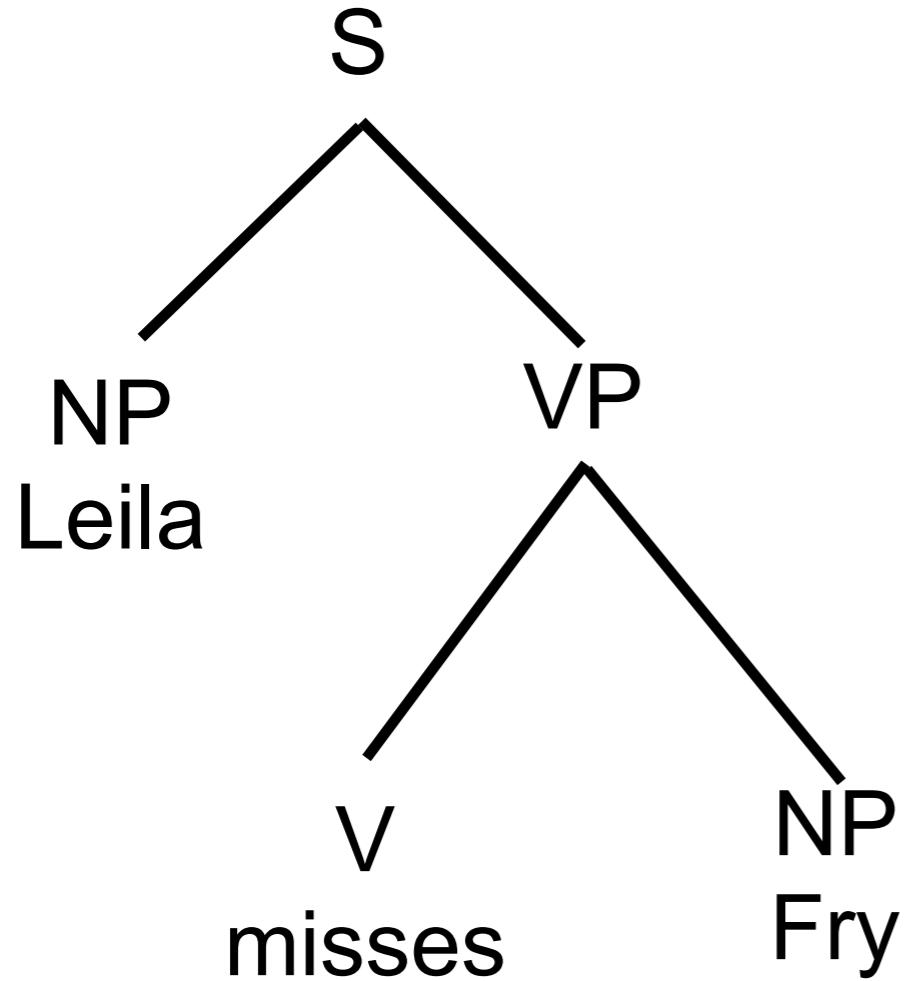


# Joshua Decoder

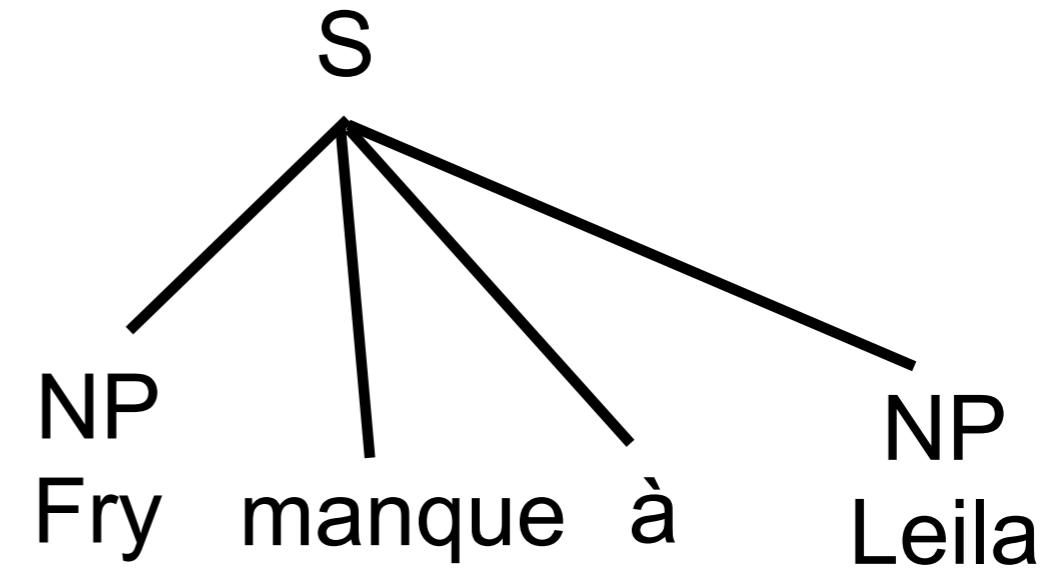
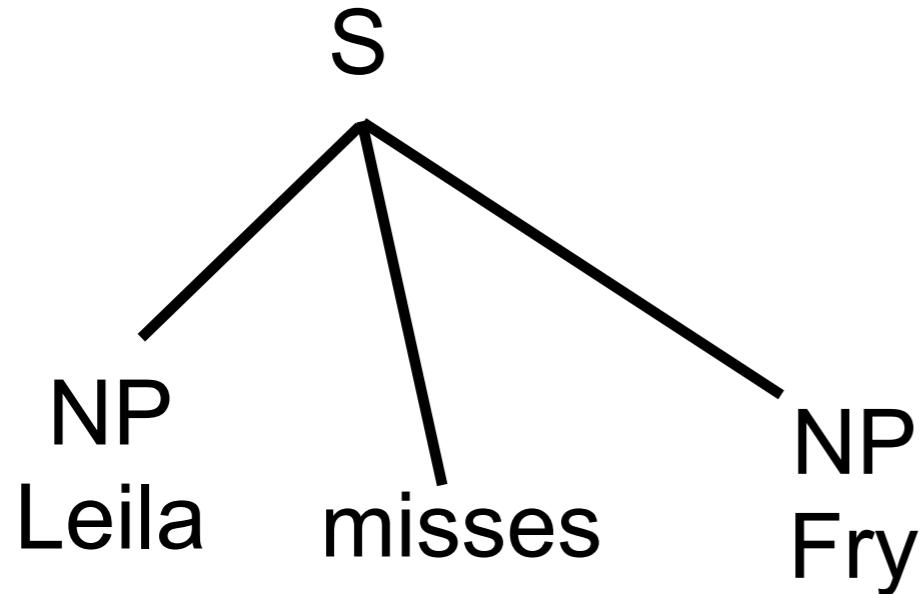


- Download it from
  - <http://joshua-decoder.org>
- Brownie points if you use it in your final projects
- Use Jonny's Thrax grammar extractor to test different kinds of SFCGs for your problems

# Dealing with language mismatches



# Dealing with language mismatches

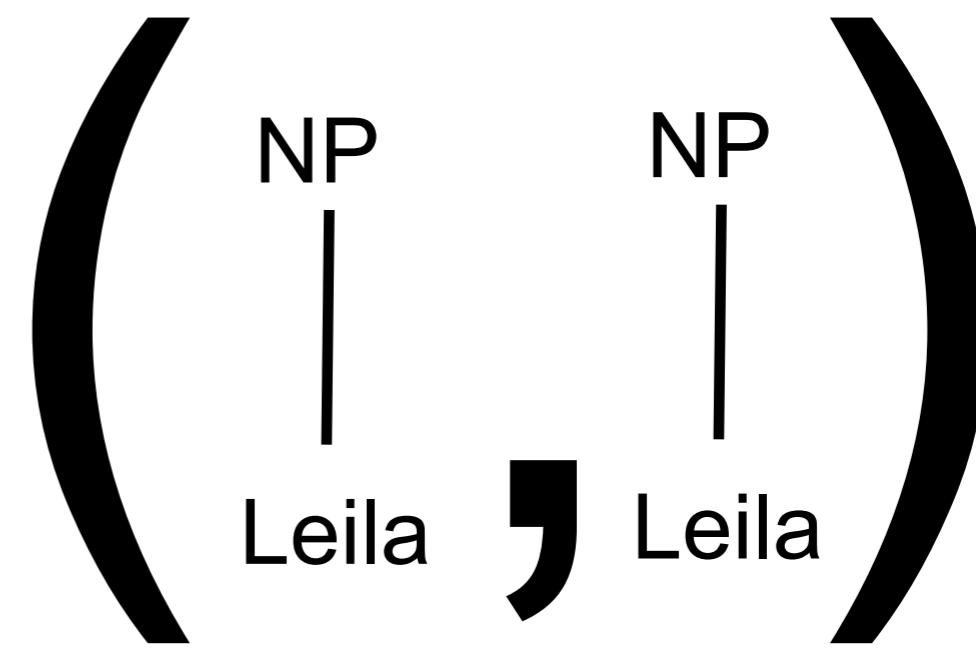
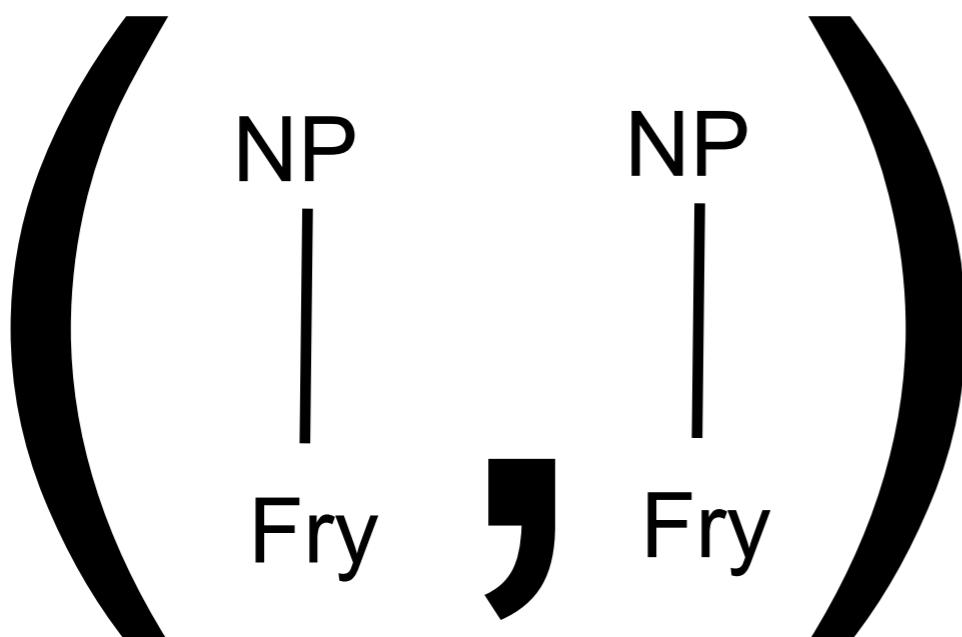
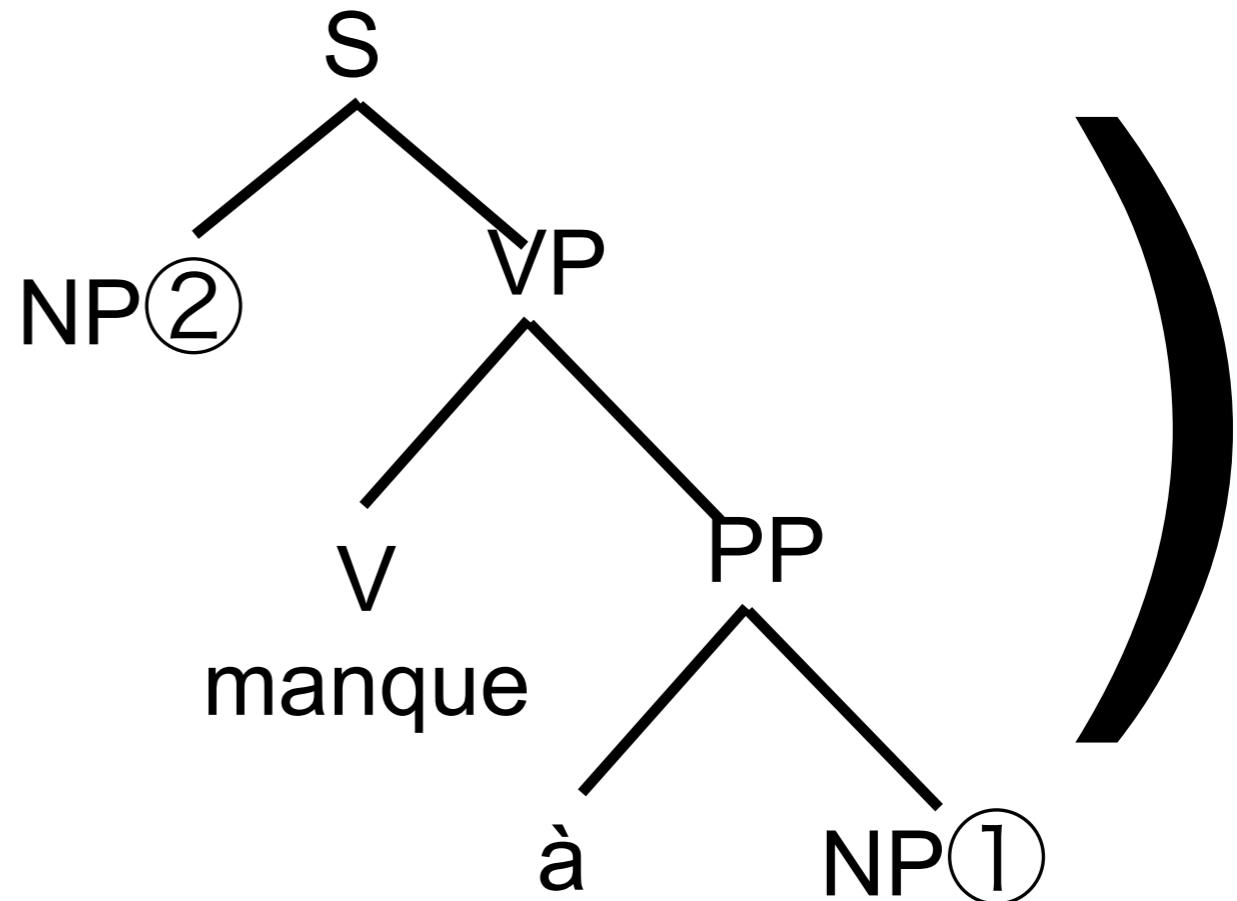
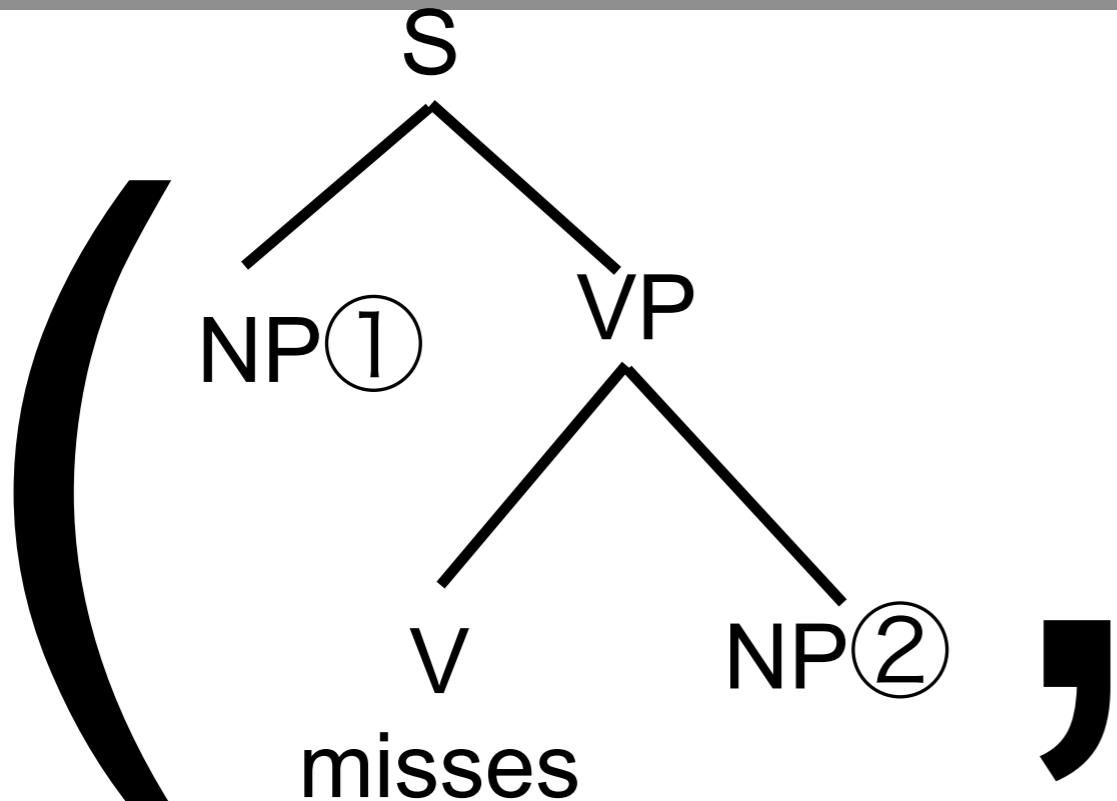


$S \rightarrow NP① \text{ misses } NP② \quad NP② \text{ manque à } NP①$

$NP \rightarrow \text{ Fry }$

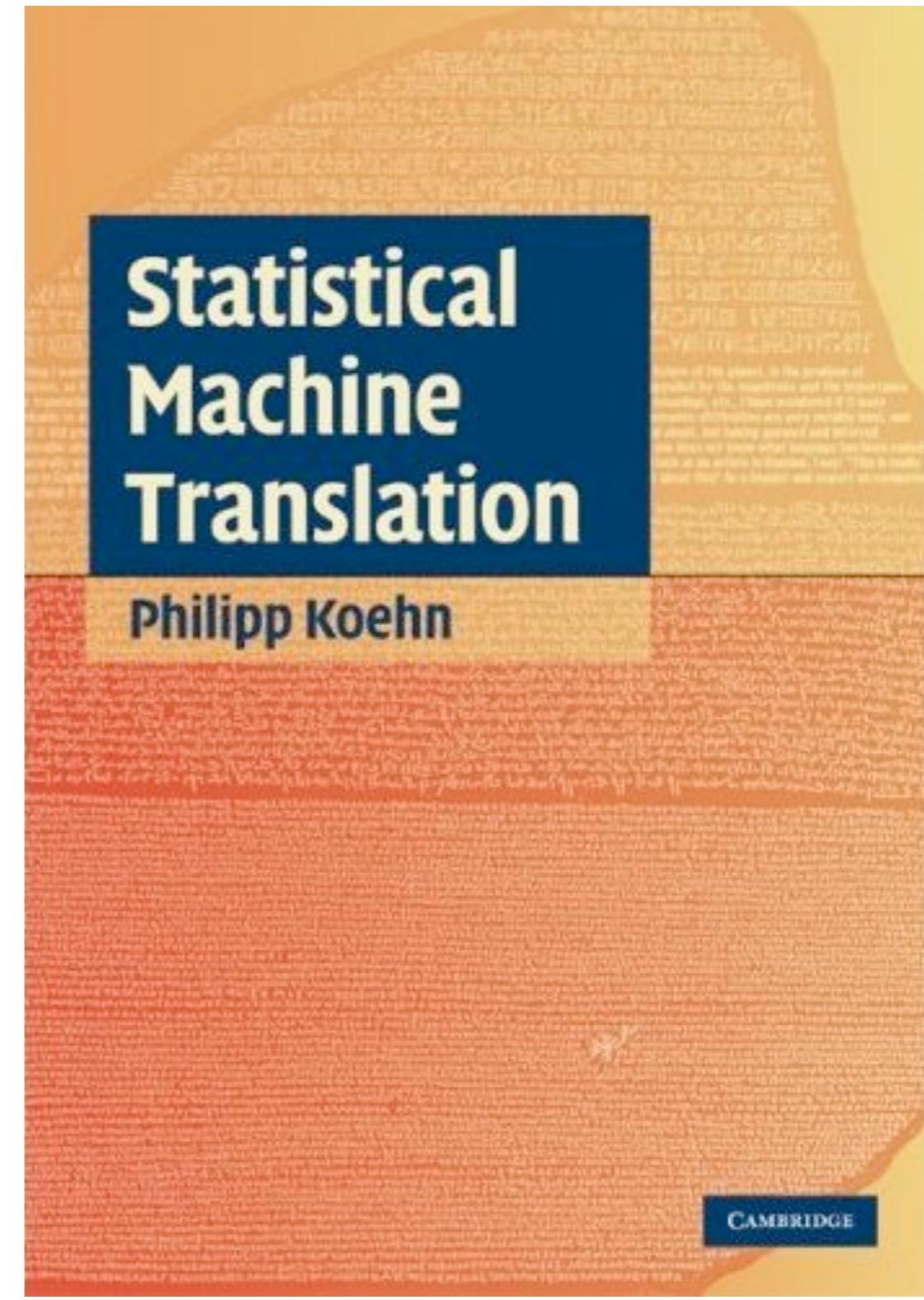
$NP \rightarrow \text{ Leila }$

# Synchronous Tree Substitution



# Reading

- Read Chapter 11 from the textbook



# Announcements

- Next week:
  - On Thursday 3/27 you have two items due
  - HW4 is due
  - You must send the training/development/test data that you will use for your term project to Jonny
- Language in 10 minutes:
  - Tuesday: Jessy - Serbian and Rigel - Javanese
  - Thursday: Kai - Italian