Statistical Machine Translation
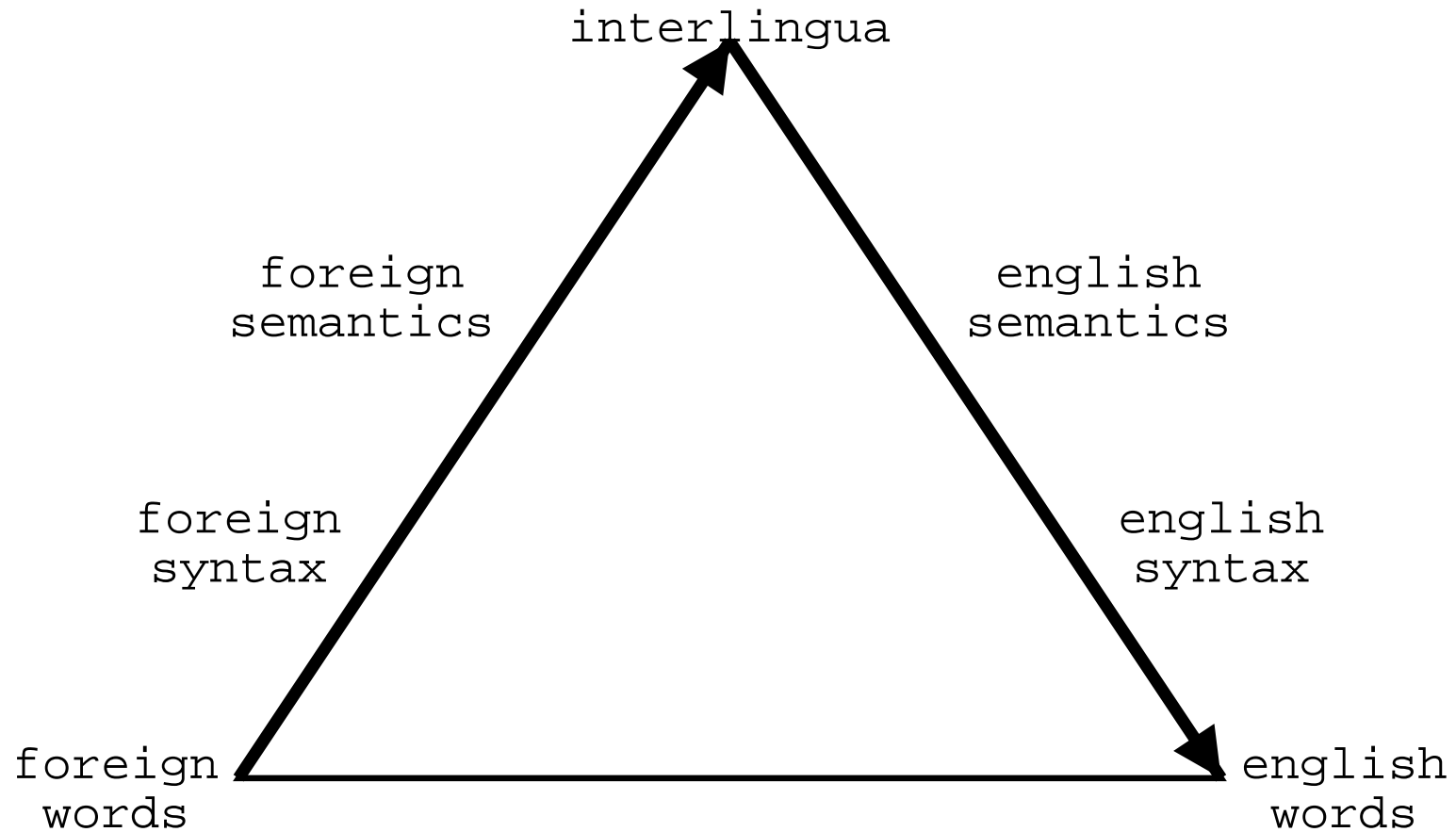# Lecture 5

Syntax-Based Models

Stephen Clark

(based on slides by Philipp Koehn)

# The Challenge of Syntax

- The classical machine translation pyramid

Advantages of Syntax-Based Translation

- Reordering for syntactic reasons

    – e.g., move German object to end of sentence

- Conditioning on syntactically related words

    – translation of verb may depend on subject or object

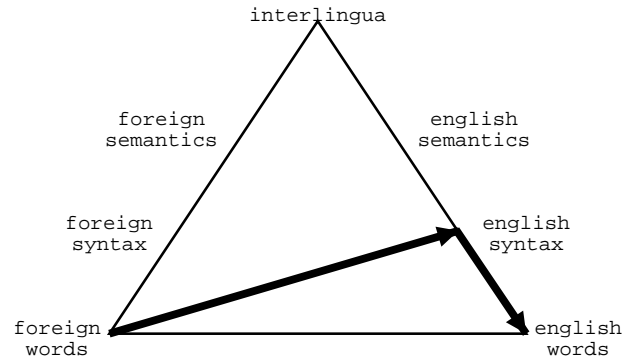- Use of syntactic language models

Syntactic Language Model

- Good syntax tree $\rightarrow$ good English
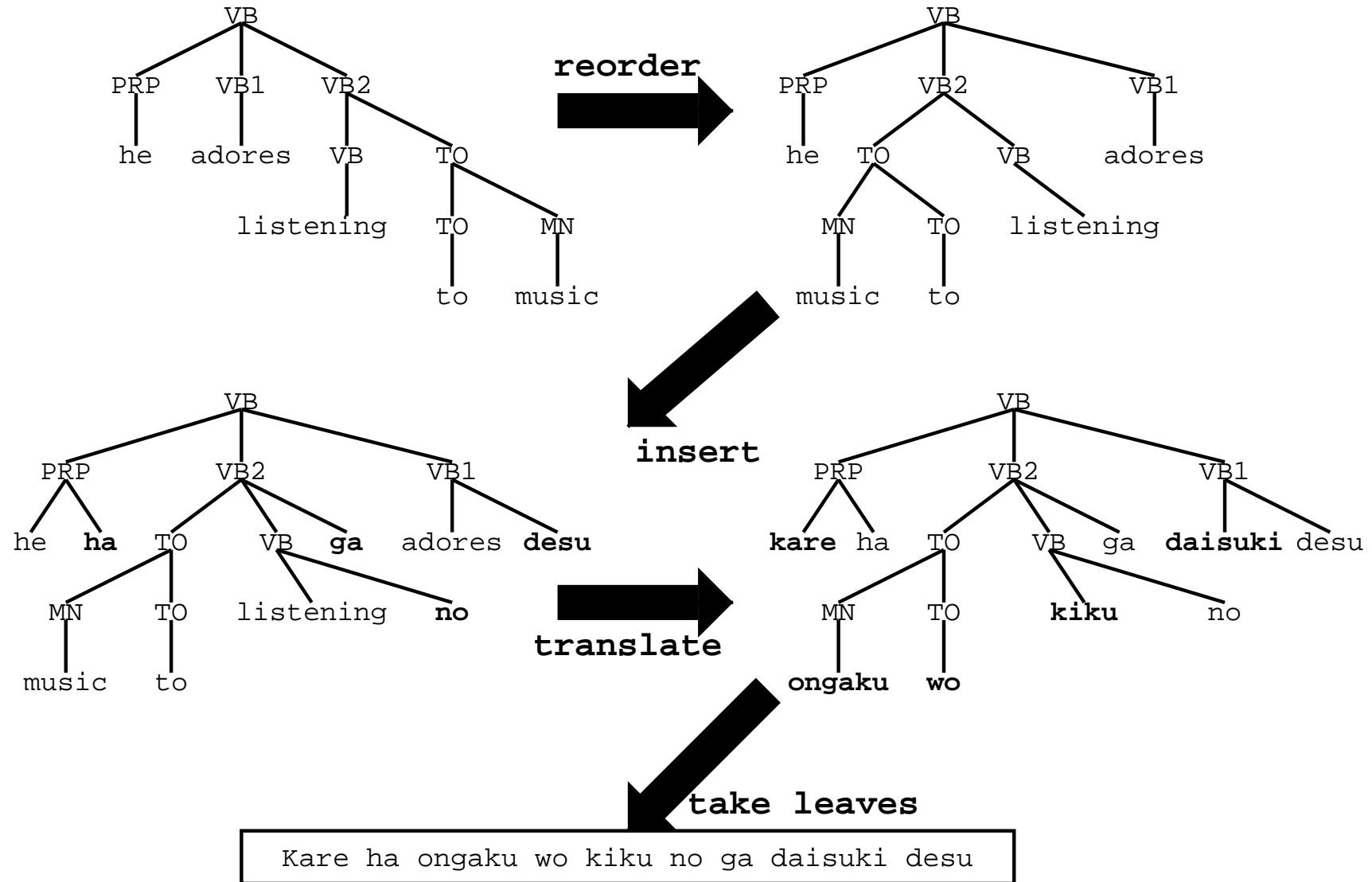
- Allows for long distance constraints
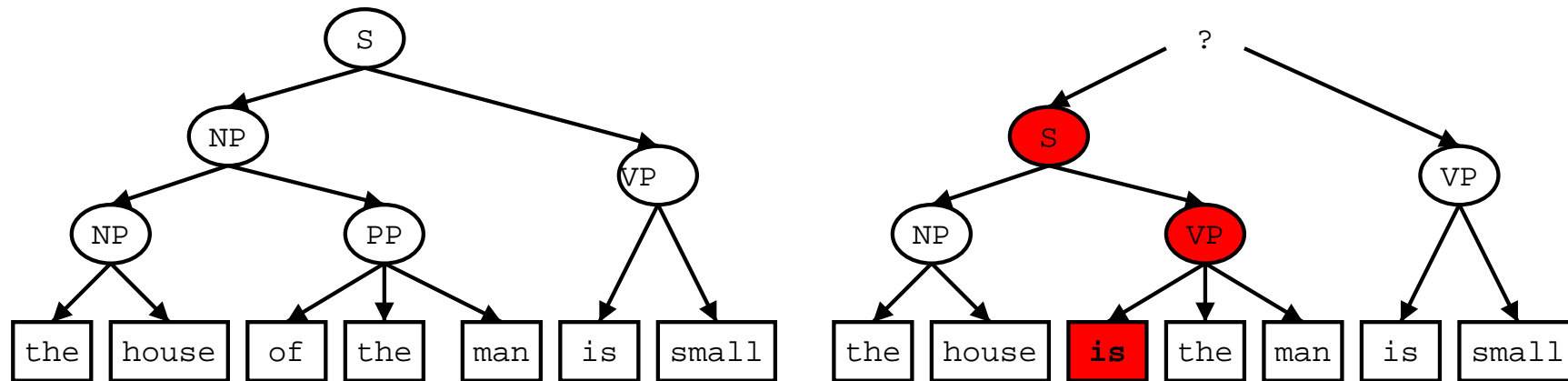
- Left translation preferred by syntactic LM

String to Tree Translation

- **Use of English syntax trees [Yamada and Knight, 2001]**

    – exploit rich resources on the English side

    – obtained with statistical parser [Collins, 1997]

    – flattened tree to allow more reorderings

   – works well with syntactic language model

Yamada and Knight [2001]

Reordering Table

Original Order	Reordering	$p(\text{reorder}|\text{original})$
PRP VB1 VB2	PRP VB1 VB2	0.074
PRP VB1 VB2	**PRP VB2 VB1**	**0.723**
PRP VB1 VB2	VB1 PRP VB2	0.061
PRP VB1 VB2	VB1 VB2 PRP	0.037
PRP VB1 VB2	VB2 PRP VB1	0.083
PRP VB1 VB2	VB2 VB1 PRP	0.021
VB TO	VB TO	0.107
VB TO	**TO VB**	**0.893**
TO NN	TO NN	0.251
TO NN	**NN TO**	**0.749**

Decoding as Parsing
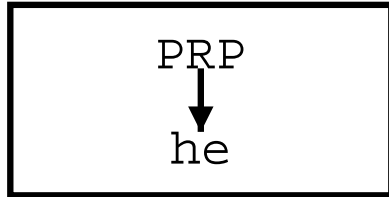
```
┌─────────────────────────┐
│        PRP              │
│         ↓               │
│         he              │
└─────────────────────────┘
```

kare   ha   ongaku   wo   kiku   no   ga daisuki desu

- Pick Japanese words

- Translate into tree stumps

Decoding as Parsing

```
┌─────────────────┐  ┌──────────┐  ┌────────┐
│      PRP        │  │   NN     │  │  TO    │
│       ↓         │  │    ↓     │  │   ↓    │
│      he         │  │  music   │  │  to    │
└─────────────────┘  └──────────┘  └────────┘
   kare     ha    ongaku   wo    kiku    no     ga  daisuki desu
```
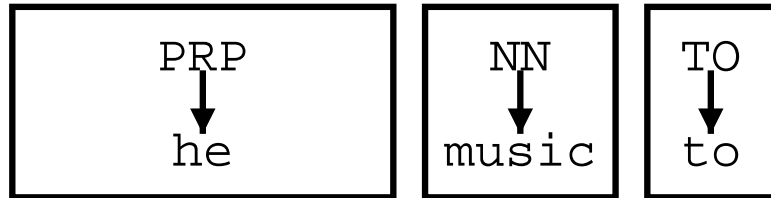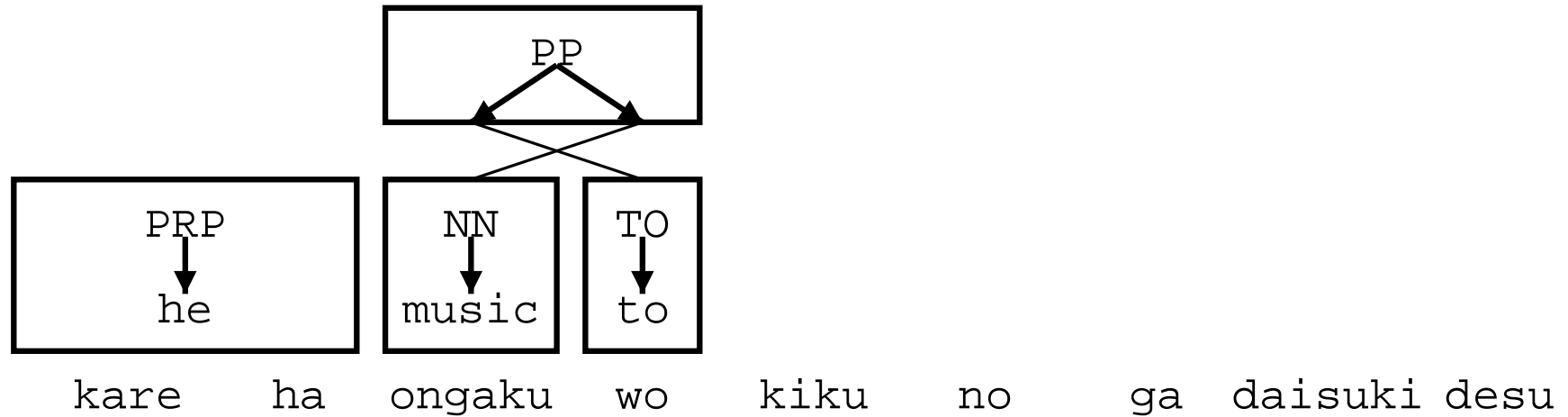
- Pick Japanese words

- Translate into tree stumps

Decoding as Parsing

```
                    ┌──────────────┐
                    │      PP      │
                    │   ◢     ◣    │
                    └──╱───╲──╱────┘
                      ╱  ╲ ╳ ╱
      ┌──────────┐ ┌──────┐┌──────┐
      │   PRP    │ │  NN  ││  TO  │
      │    ↓     │ │  ↓   ││  ↓   │
      │   he     │ │music ││  to  │
      └──────────┘ └──────┘└──────┘

   kare     ha   ongaku   wo   kiku   no    ga  daisuki desu
```
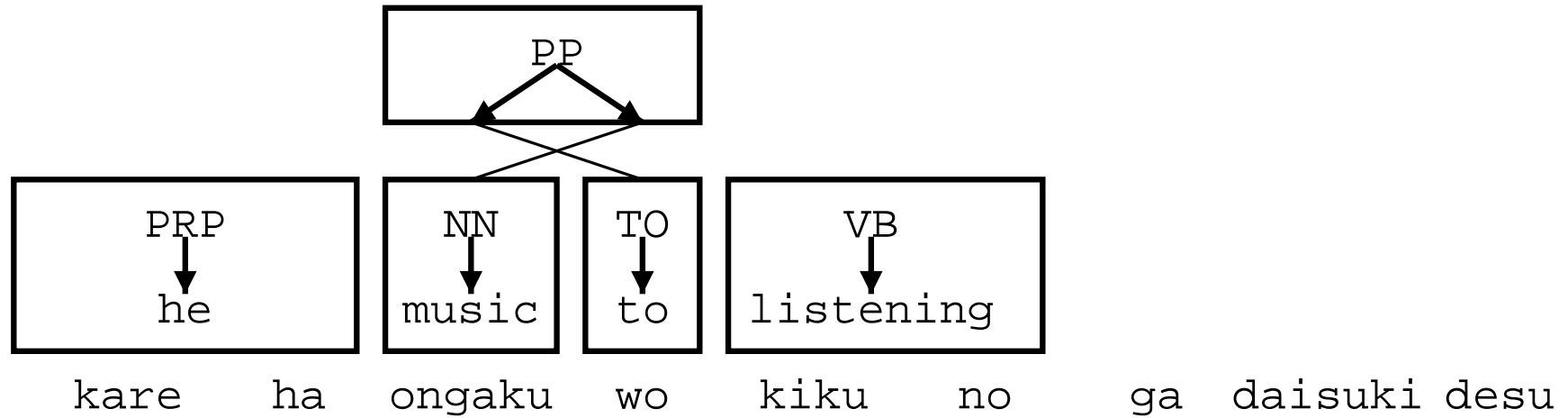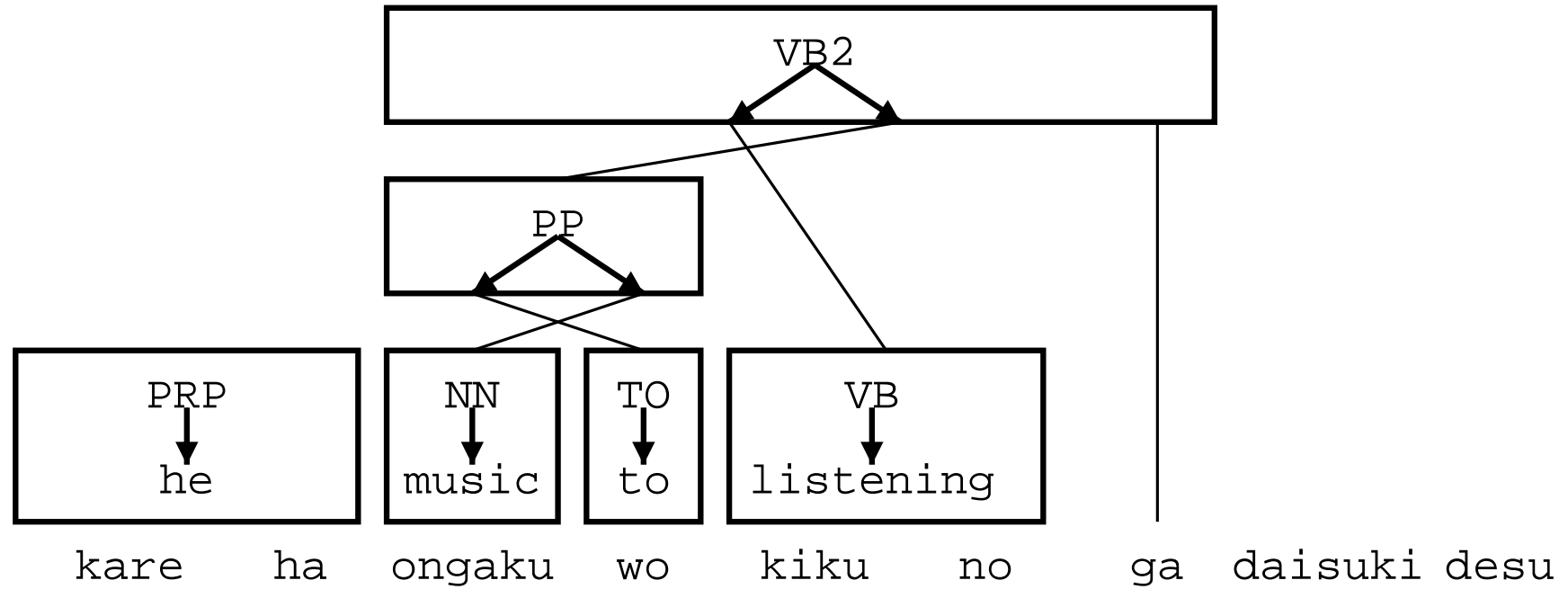
- Apply grammar rules (with reordering)

Decoding as Parsing

```
                    ┌─────────────────┐
                    │       PP        │
                    │                 │
                    └─────────────────┘
         ┌──────────┐  ┌──────┐┌─────┐ ┌─────────────┐
         │   PRP    │  │  NN  ││ TO  │ │     VB      │
         │    ↓     │  │  ↓   ││ ↓   │ │     ↓       │
         │   he     │  │music ││ to  │ │  listening  │
         └──────────┘  └──────┘└─────┘ └─────────────┘
   kare      ha    ongaku   wo    kiku    no    ga  daisuki desu
```
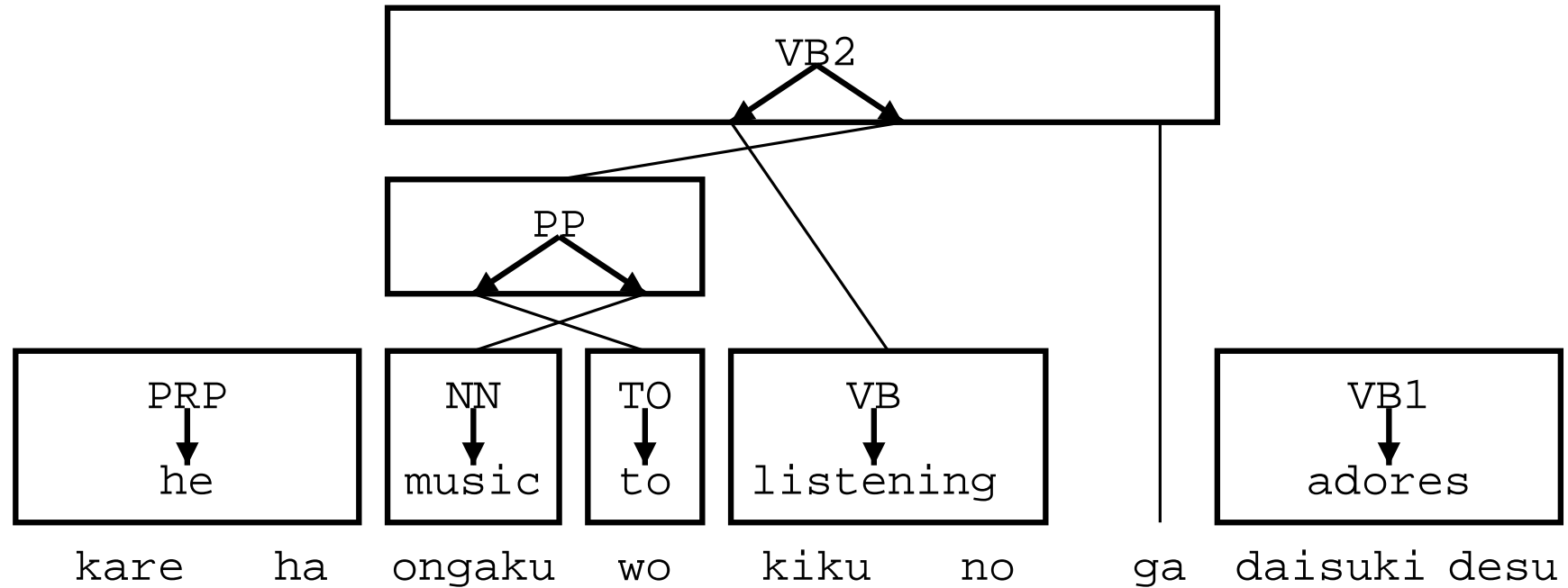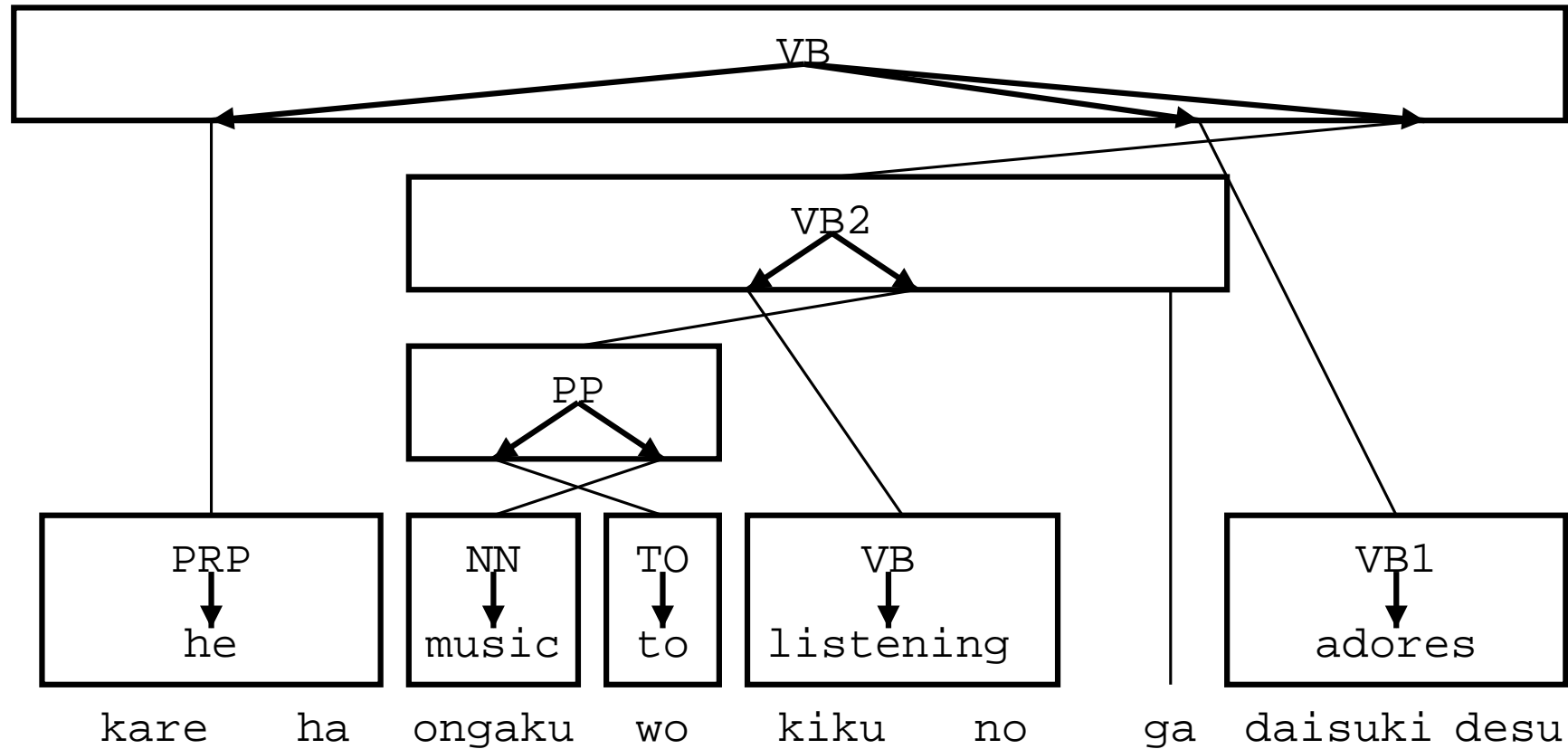
Decoding as Parsing

Decoding as Parsing

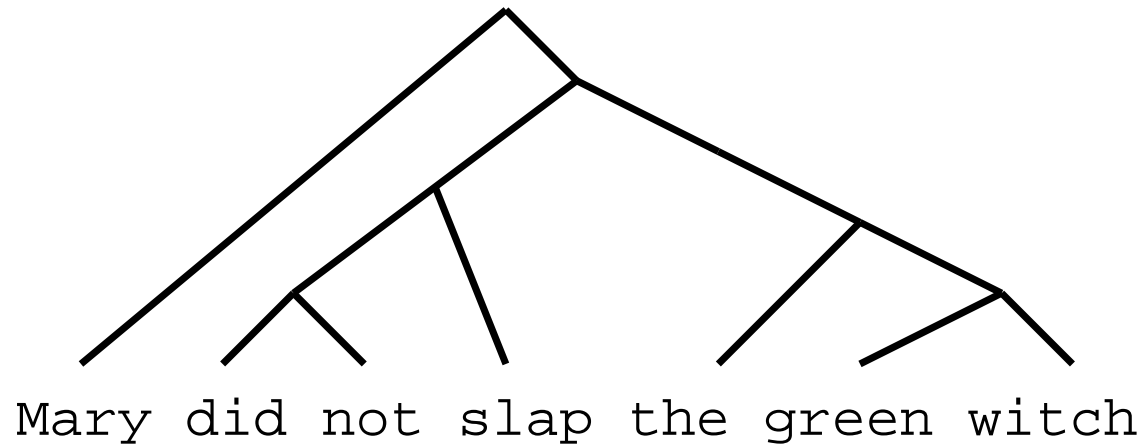# Decoding as Parsing

- Finished when all foreign words covered

Inversion Transduction Grammars

- Generation of both English and foreign trees [Wu, 1997]

- Rules (binary and unary)

  - $A \rightarrow A_1 A_2 \| A_1 A_2$

  - $A \rightarrow A_1 A_2 \| A_2 A_1$

  - $A \rightarrow e \| f$

  - $A \rightarrow e \| *$

  - $A \rightarrow * \| f$

$\Rightarrow$ Common binary tree required
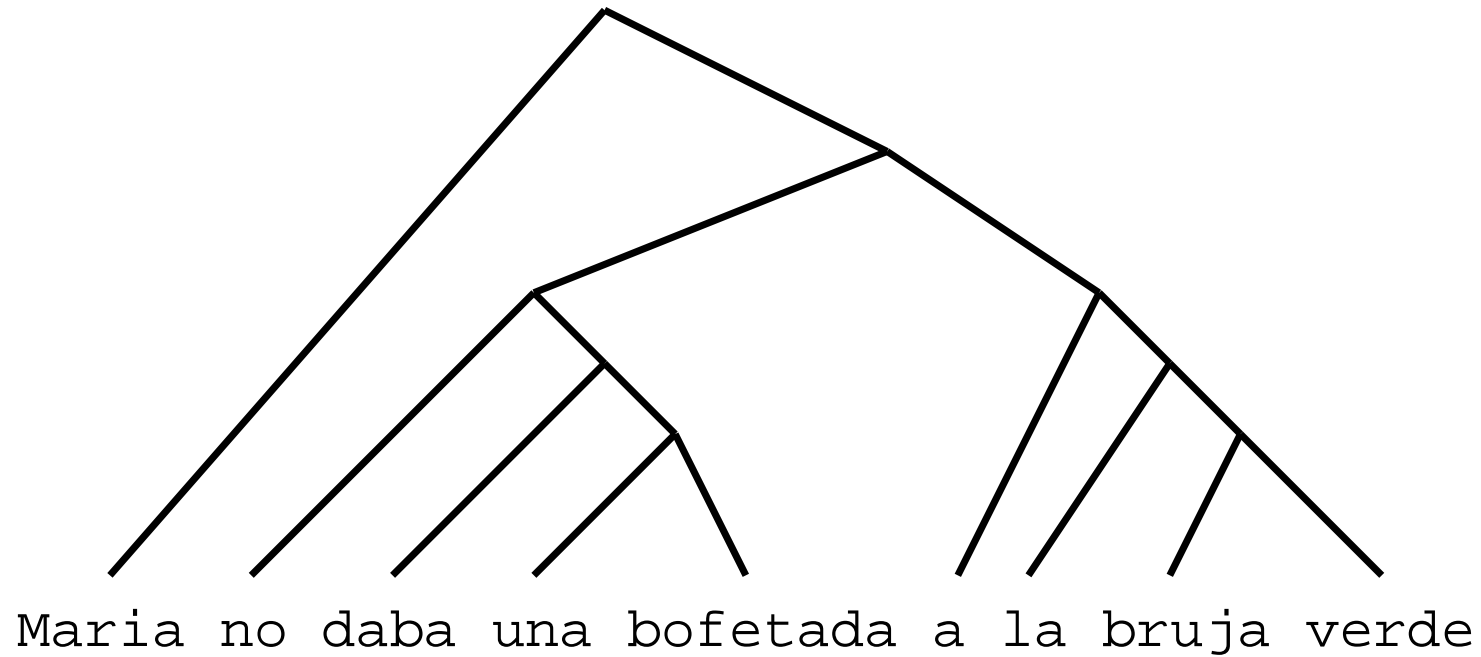
  - limits the complexity of reorderings

Syntax Trees

Mary did not slap the green witch

- English binary tree

Syntax Trees (2)

Maria no daba una bofetada a la bruja verde

- Spanish binary tree

Syntax Trees (3)


```
Mary    did not   slap     *        *        *   the   green witch
Maria    *   no   daba    una    bofetada   a   la    verde bruja
```
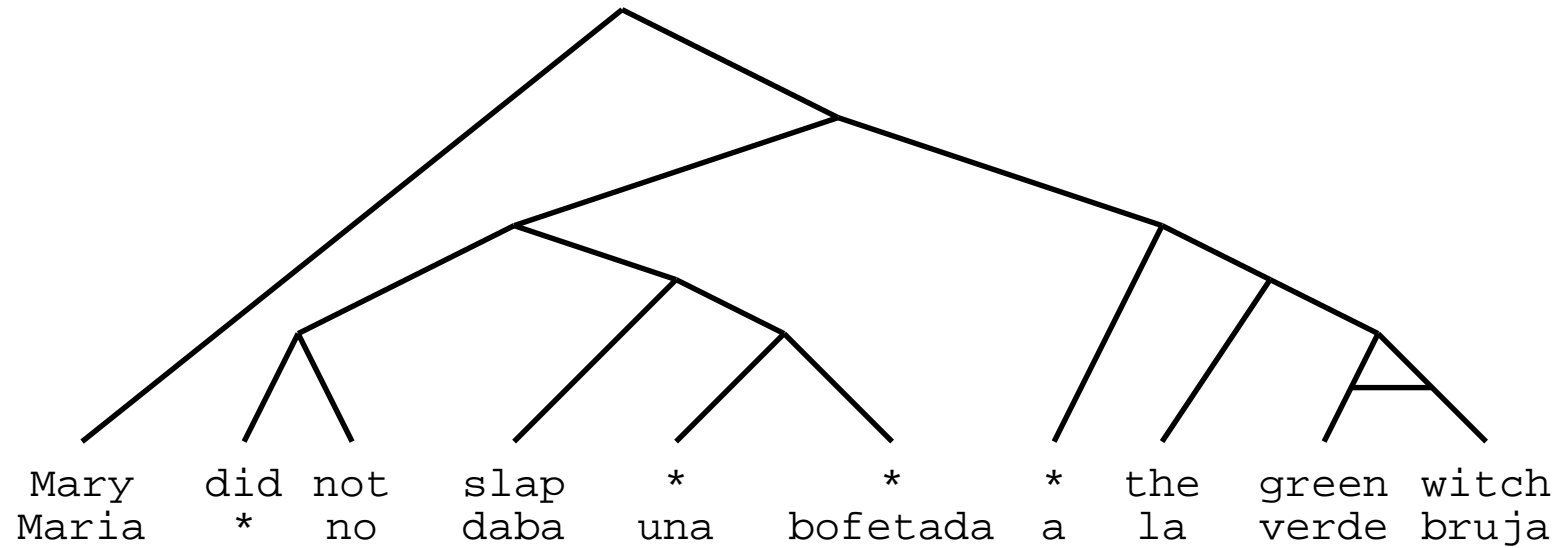
- Combined tree with reordering of Spanish

Inversion Transduction Grammars

- Decoding by parsing (as before)

- Variations

    – may use real syntax on either side or both

   – may use multi-word units at leaf nodes

Chiang: Hierarchical Phrase Model

- Chiang [ACL 2005, CL 2007]

  – context free bi-grammar

  – one non-terminal symbol

  – right hand side of rule may include non-terminals and terminals

- Now one of the dominant approaches to SMT

- Basis of the Cambridge Engineering's highly competitive SMT systems
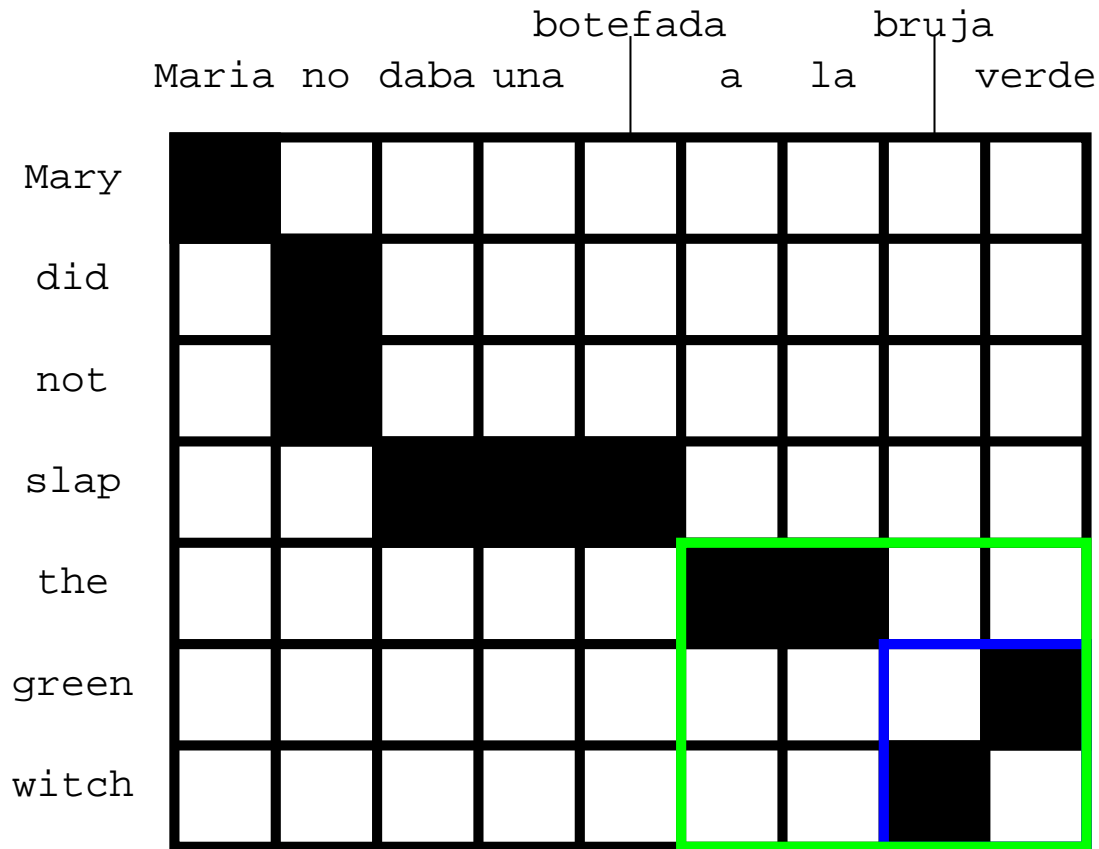
Types of Rules

- **Word translation**

  – X → maison ‖ house

- **Phrasal translation**

  – X → daba una bofetada ‖ slap

- **Mixed non-terminal / terminal**

  – X → X bleue ‖ blue X

  – X → ne X pas ‖ not X

  – X → X1 X2 ‖ X2 of X1

- **Technical rules**

  – S → S X ‖ S X

  – S → X ‖ X

Learning Hierarchical Rules
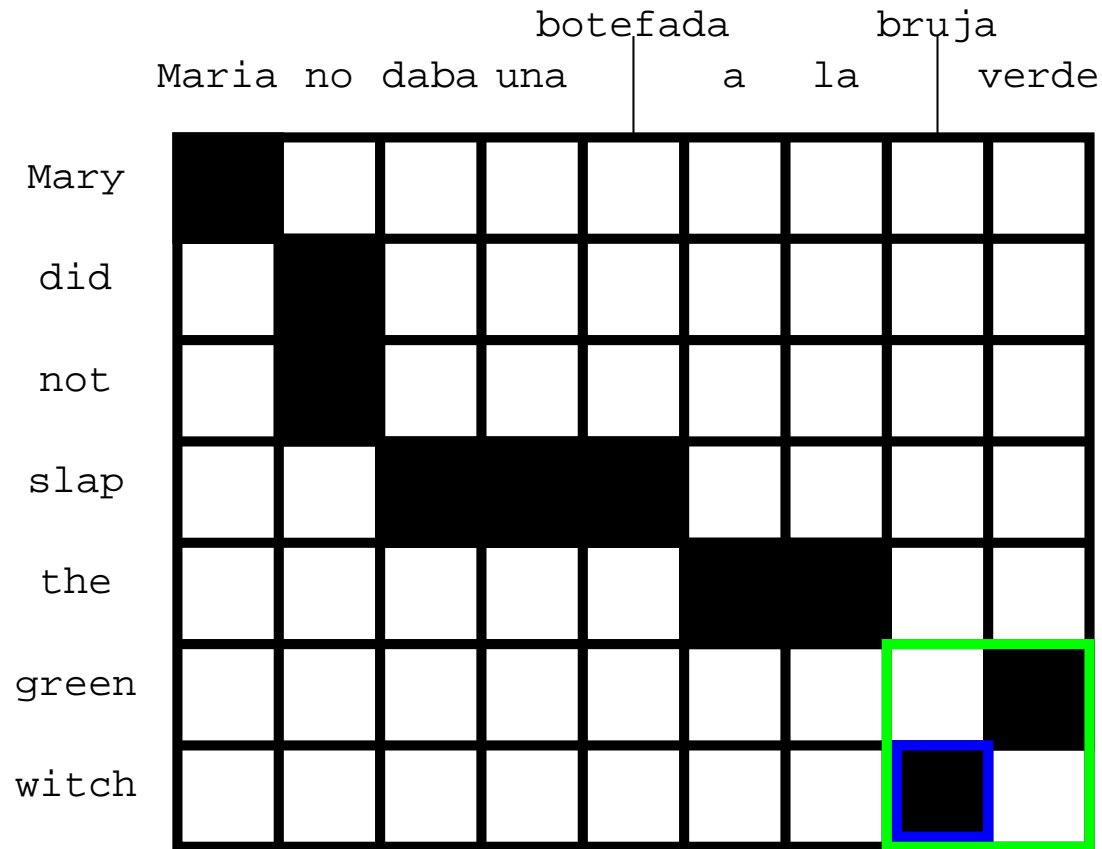
$$X \rightarrow X \text{ verde} \parallel \text{green } X$$

Learning Hierarchical Rules

$$X \rightarrow \text{a la } X \parallel \text{the } X$$

Details

- **Too many rules**

  $\rightarrow$ filtering of rules necessary

- **Efficient parse decoding possible**

  – hypothesis stack for each span of foreign words

  – only one non-terminal $\rightarrow$ hypotheses comparable

  – length limit for spans that do not start at beginning

Improved Translations

- we must also this criticism should be taken seriously .

→ we must also take this criticism seriously .

- i am with him that it is necessary , the institutional balance by means of a political revaluation of both the commission and the council to maintain .

→ i agree with him in this , that it is necessary to maintain the institutional balance by means of a political revaluation of both the commission and the council .

- thirdly , we believe that the principle of differentiation of negotiations note .

→ thirdly , we maintain the principle of differentiation of negotiations .

- perhaps it would be a constructive dialog between the government and opposition parties , social representative a positive impetus in the right direction .

→ perhaps a constructive dialog between government and opposition parties and social representative could give a positive impetus in the right direction .