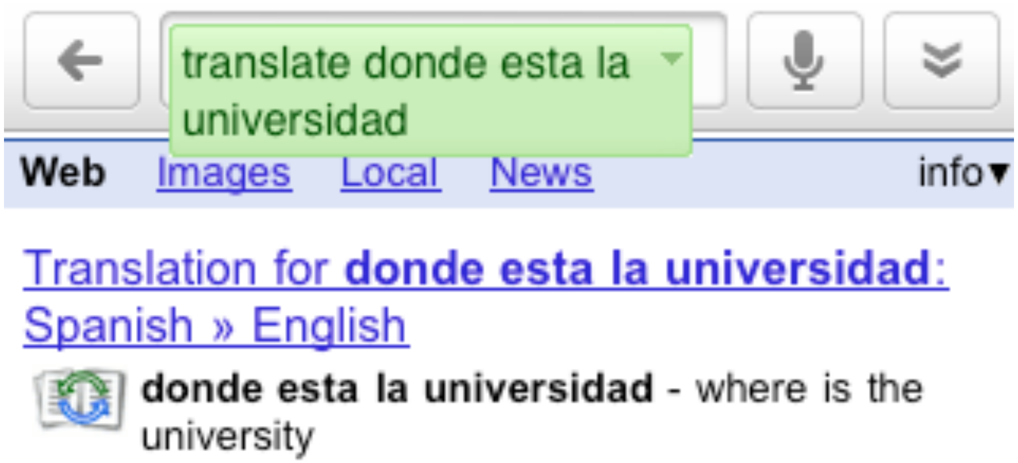
Large-Context Models for Large-Scale Machine Translation

John DeNero

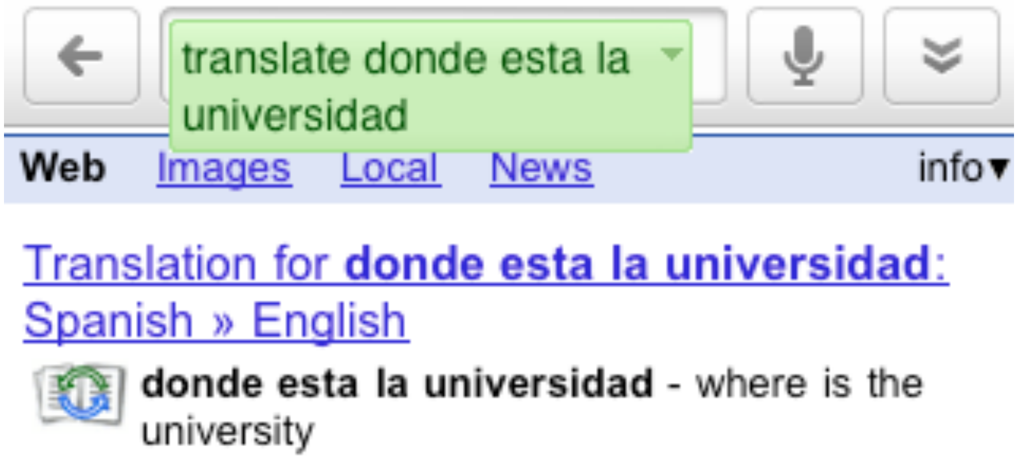
Dissertation Talk

Statistical Language Systems are Working

Statistical Language Systems are Working



Statistical Language Systems are Working

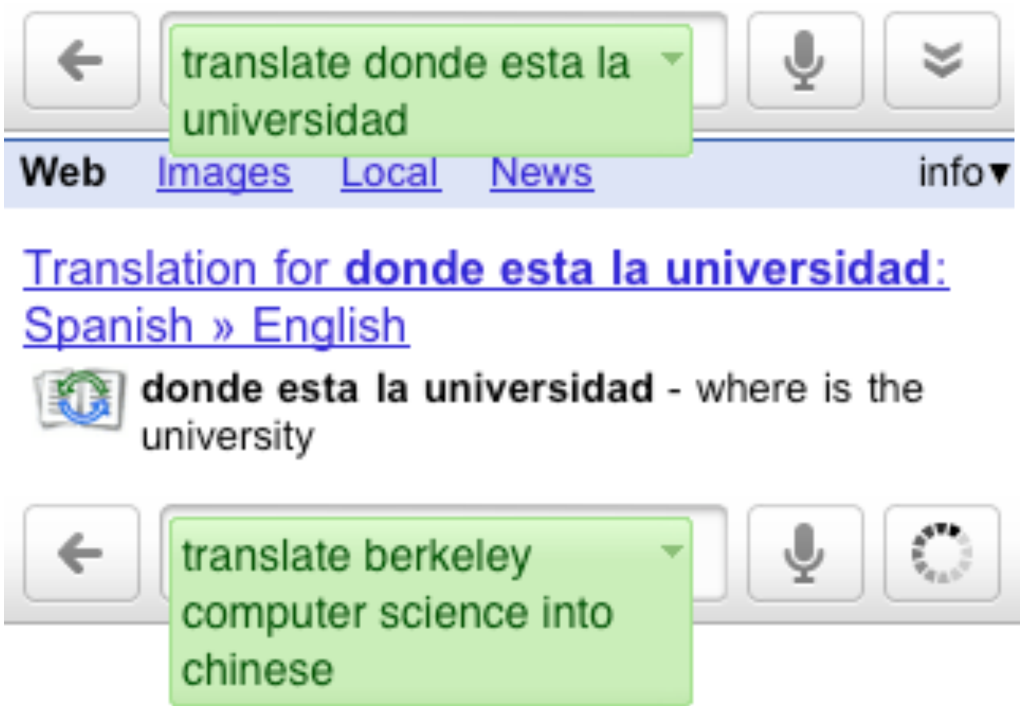


How?

Google spent $5.6 billion on infrastructure in the last 3 years [1]

1 Google.com annual report of capital expenditure, "the majority of which was related to IT infrastructure investments."

Statistical Language Systems are Working

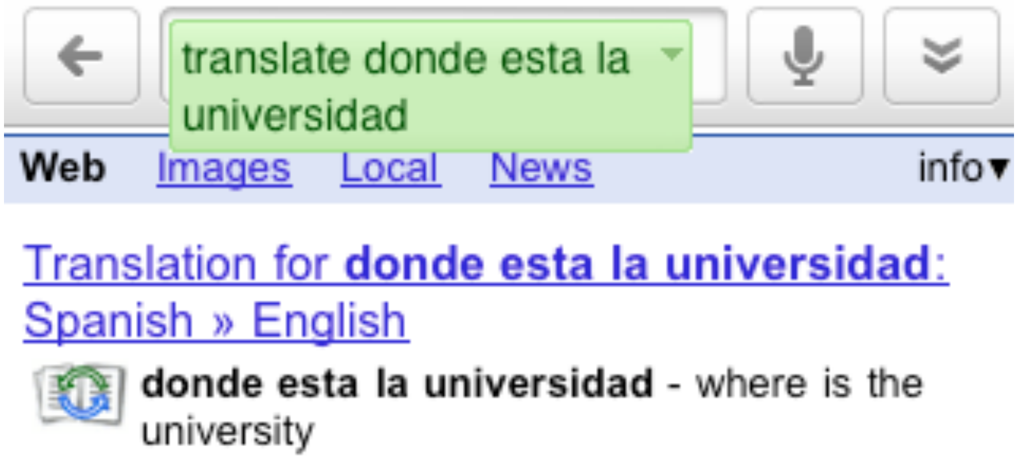


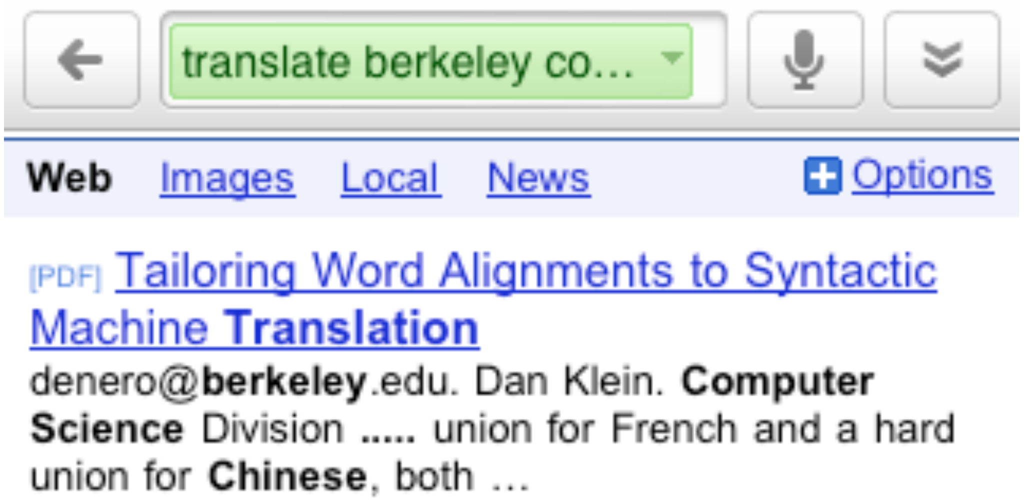
How?

Google spent $5.6 billion on infrastructure in the last 3 years [1]

1 Google.com annual report of capital expenditure, "the majority of which was related to IT infrastructure investments."

Statistical Language Systems are Working



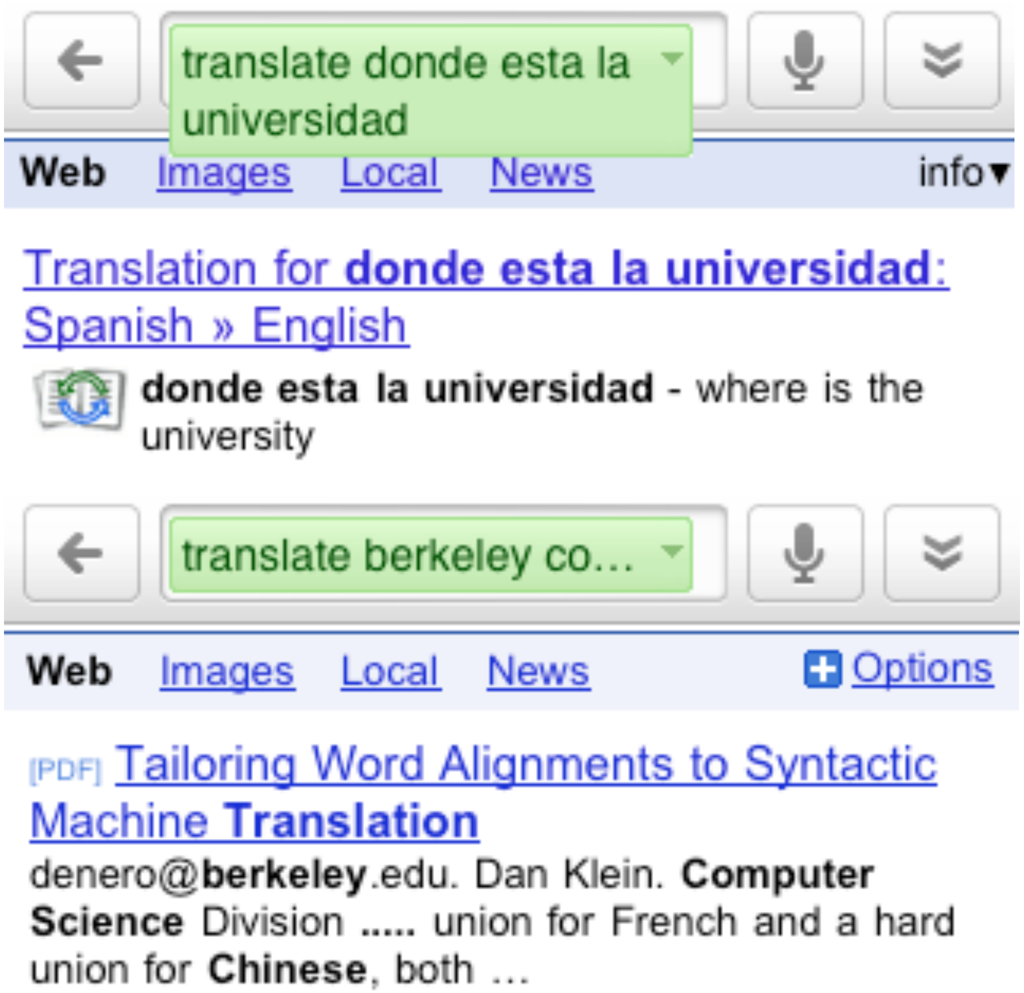


How?

Google spent $5.6 billion on infrastructure in the last 3 years [1]

1 Google.com annual report of capital expenditure, "the majority of which was related to IT infrastructure investments."

Statistical Language Systems are Working



How?

Google spent $5.6 billion on infrastructure in the last 3 years [1]

> " People use [Google Translate] hundreds of millions of times a week." [2]

1 Google.com annual report of capital expenditure, "the majority of which was related to IT infrastructure investments."

2 "Google's Computing Power Refines Translation Tool," New York Times, 9 March 2010, Technology Section.

The Many Use(r)s of Machine Translation

Assimilation	Dissemination	Communication

The Many Use(r)s of Machine Translation

Assimilation

- Document translation
- Broadcast monitoring
- Intelligence gathering

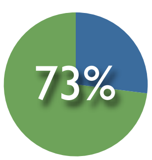
Dissemination

Communication

The Many Use(r)s of Machine Translation

Assimilation

- Document translation
- Broadcast monitoring
- Intelligence gathering



Most Internet users can't read the English Web

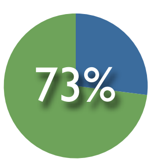
Dissemination

Communication

The Many Use(r)s of Machine Translation

Assimilation

- Document translation
- Broadcast monitoring
- Intelligence gathering



Most Internet users can't read the English Web

Dissemination

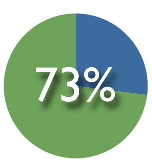
tripadvisor®
get the truth. then go.™

Communication

The Many Use(r)s of Machine Translation

Assimilation

- Document translation
- Broadcast monitoring
- Intelligence gathering



Most Internet users can't read the English Web

Dissemination

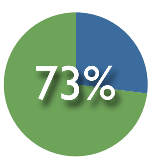
machine translated

Communication

The Many Use(r)s of Machine Translation

Assimilation

- Document translation
- Broadcast monitoring
- Intelligence gathering



Most Internet users can't read the English Web

Dissemination

get the truth. then go.™

machine translated

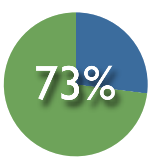
Communication

The Many Use(r)s of Machine Translation

Assimilation

USA

- Document translation
- Broadcast monitoring
- Intelligence gathering



Most Internet users can't read the English Web

Dissemination

tripadvisor®
get the truth. then go.™
machine translated

twinslator BETA

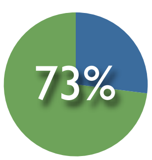
John I'm in Berkeley

Communication

The Many Use(r)s of Machine Translation

Assimilation

- Document translation
- Broadcast monitoring
- Intelligence gathering



Most Internet users can't read the English Web

Dissemination

machine translated

John I'm in Berkeley

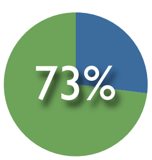
Juan Estoy en Berkeley

Communication

The Many Use(r)s of Machine Translation

Assimilation

- Document translation
- Broadcast monitoring
- Intelligence gathering



Most Internet users can't read the English Web

Dissemination

machine translated

John I'm in Berkeley

Juan Estoy en Berkeley

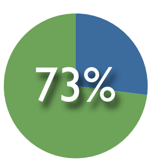
Communication

- Emergency room triage
- Military deployments
- Multilingual education
- 9-1-1 Response
- Commerce with tourists

The Many Use(r)s of Machine Translation

Assimilation

- Document translation
- Broadcast monitoring
- Intelligence gathering



Most Internet users can't read the English Web

Dissemination

tripadvisor® get the truth. then go.™

machine translated

John I'm in Berkeley

Juan Estoy en Berkeley

Communication

- Emergency room triage
- Military deployments
- Multilingual education
- 9-1-1 Response
- Commerce with tourists

John

婆婆

[Grandma-in-law]

Data-Driven Machine Translation

Target language corpus gives examples of well-formed sentences

I will get to it later	See you later	He will do it

Parallel corpus gives translation examples

I will do it gladly Yo lo haré de muy buen grado	You will see later Después lo veras

Machine translation system:

Data-Driven Machine Translation

Target language corpus gives examples of well-formed sentences

I will get to it later

See you later

He will do it

Parallel corpus gives translation examples

I will do it gladly

Yo lo haré de muy buen grado

You will see later

Después lo veras

Machine translation system:

Model of translation

Data-Driven Machine Translation

Target language corpus gives examples of well-formed sentences

I will get to it later

See you later

He will do it

Parallel corpus gives translation examples

I will do it gladly

Yo lo haré de muy buen grado

You will see later

Después lo veras

Machine translation system:

Source language

Yo lo haré después

NOVEL SENTENCE

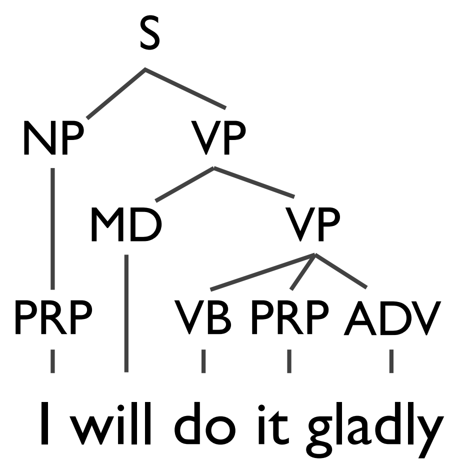
Model of translation

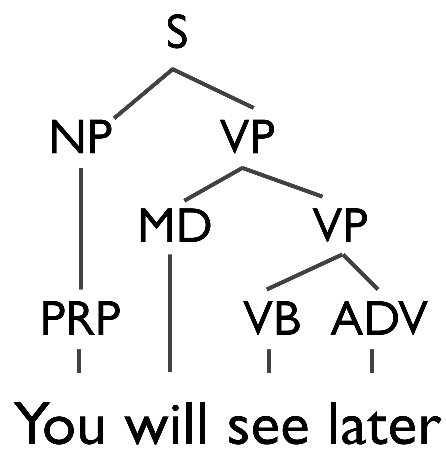
Target language

I will do it later

Stitching Together Fragments

Parallel corpus gives translation examples

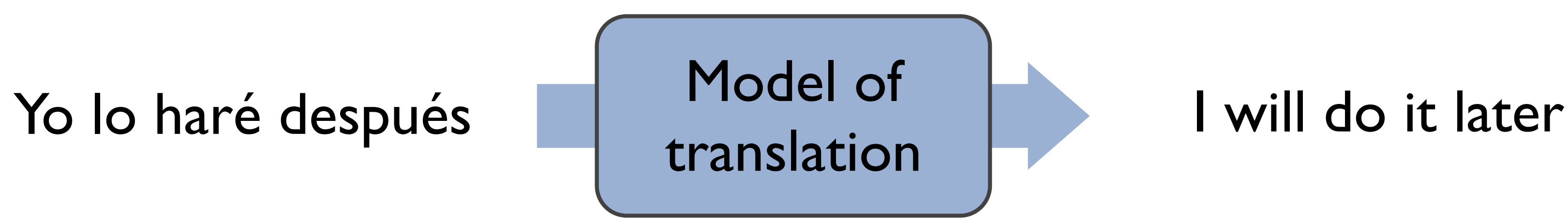




Yo lo haré de muy buen grado

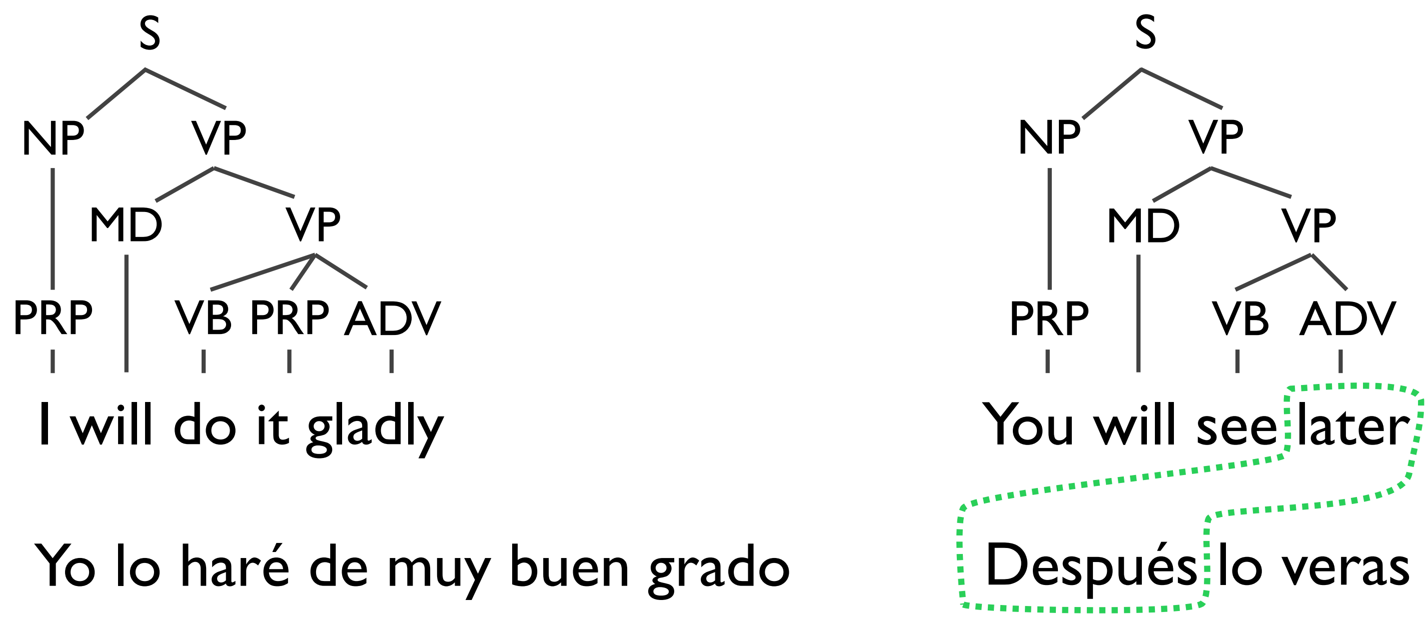
Después lo veras

Machine translation system:

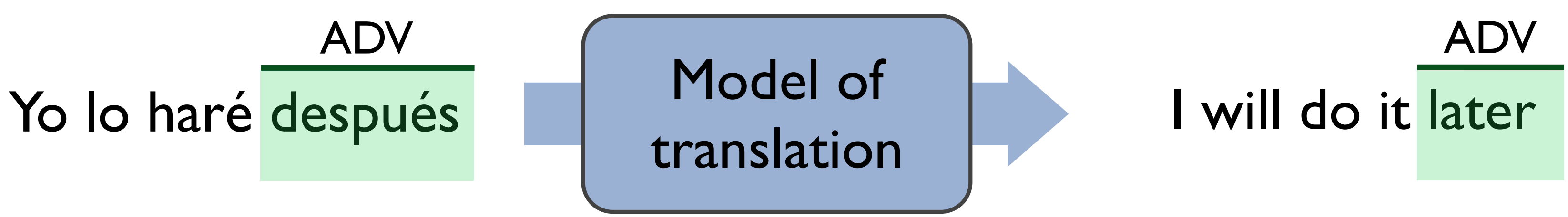


Stitching Together Fragments

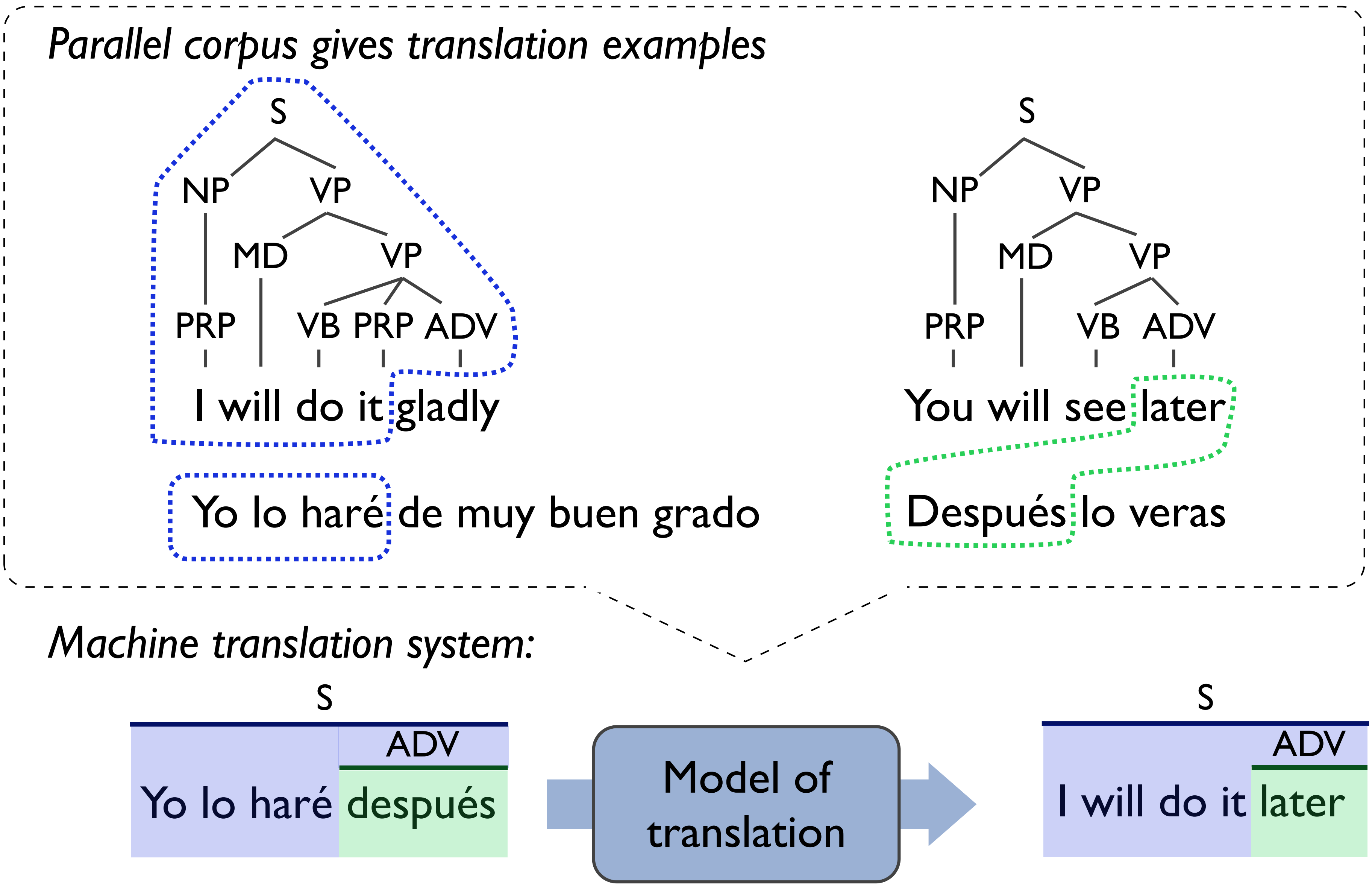
Parallel corpus gives translation examples



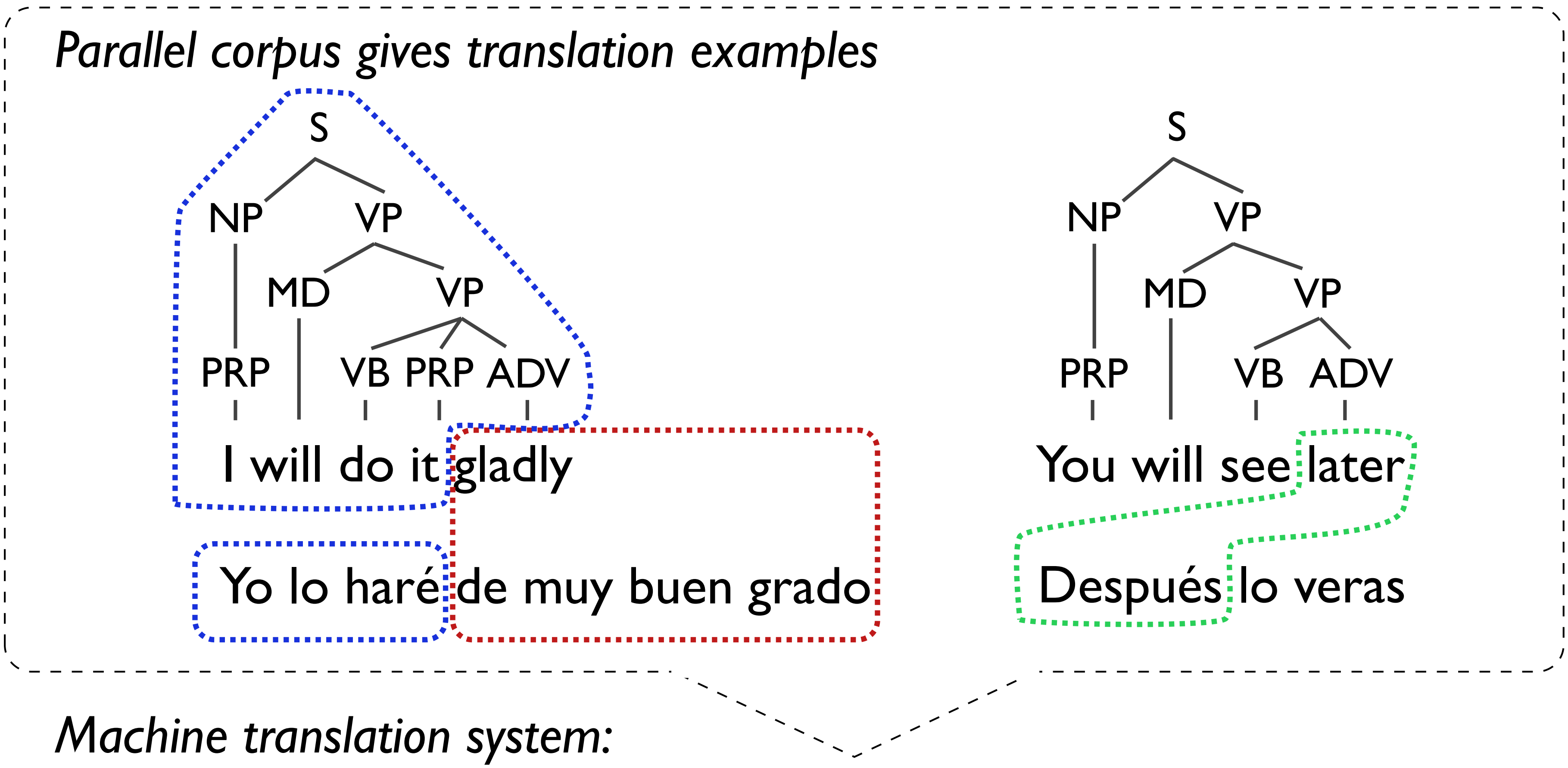
Machine translation system:



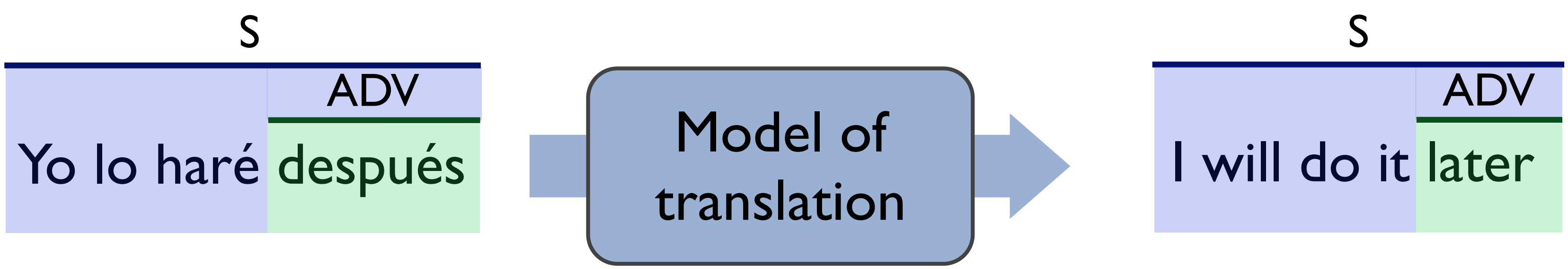
Stitching Together Fragments



Stitching Together Fragments



Machine translation system:



An Example Syntax-Based Translation

Arabic source sentence:

ورفض الباز الادلاء باى تصريحات فور وصوله الى المقاطعة

An Example Syntax-Based Translation

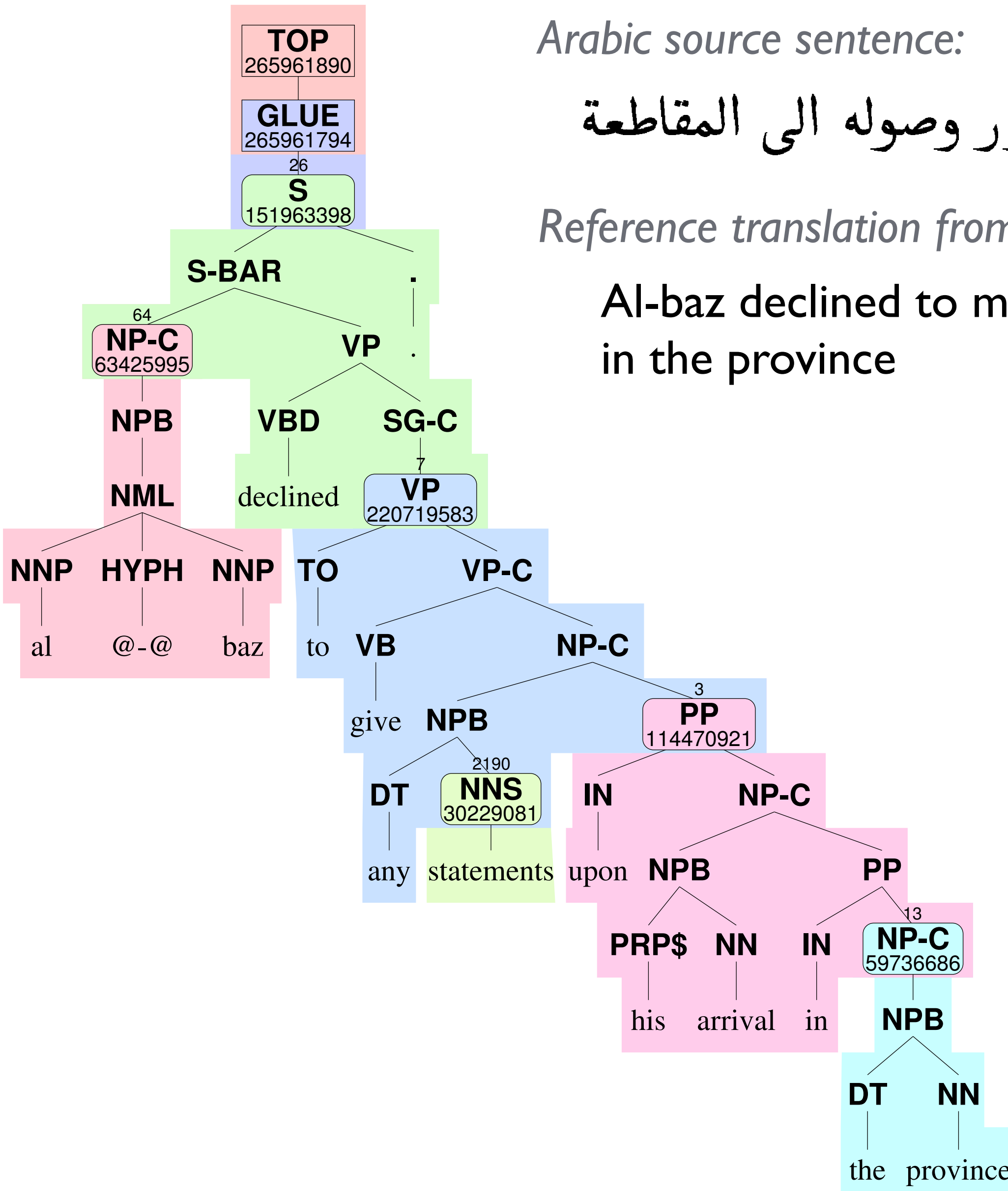
Arabic source sentence:

ورفض الباز الادلاء باى تصريحات فور وصوله الى المقاطعة

Reference translation from a human translator:

Al-baz declined to make any statements upon his arrival in the province

An Example Syntax-Based Translation



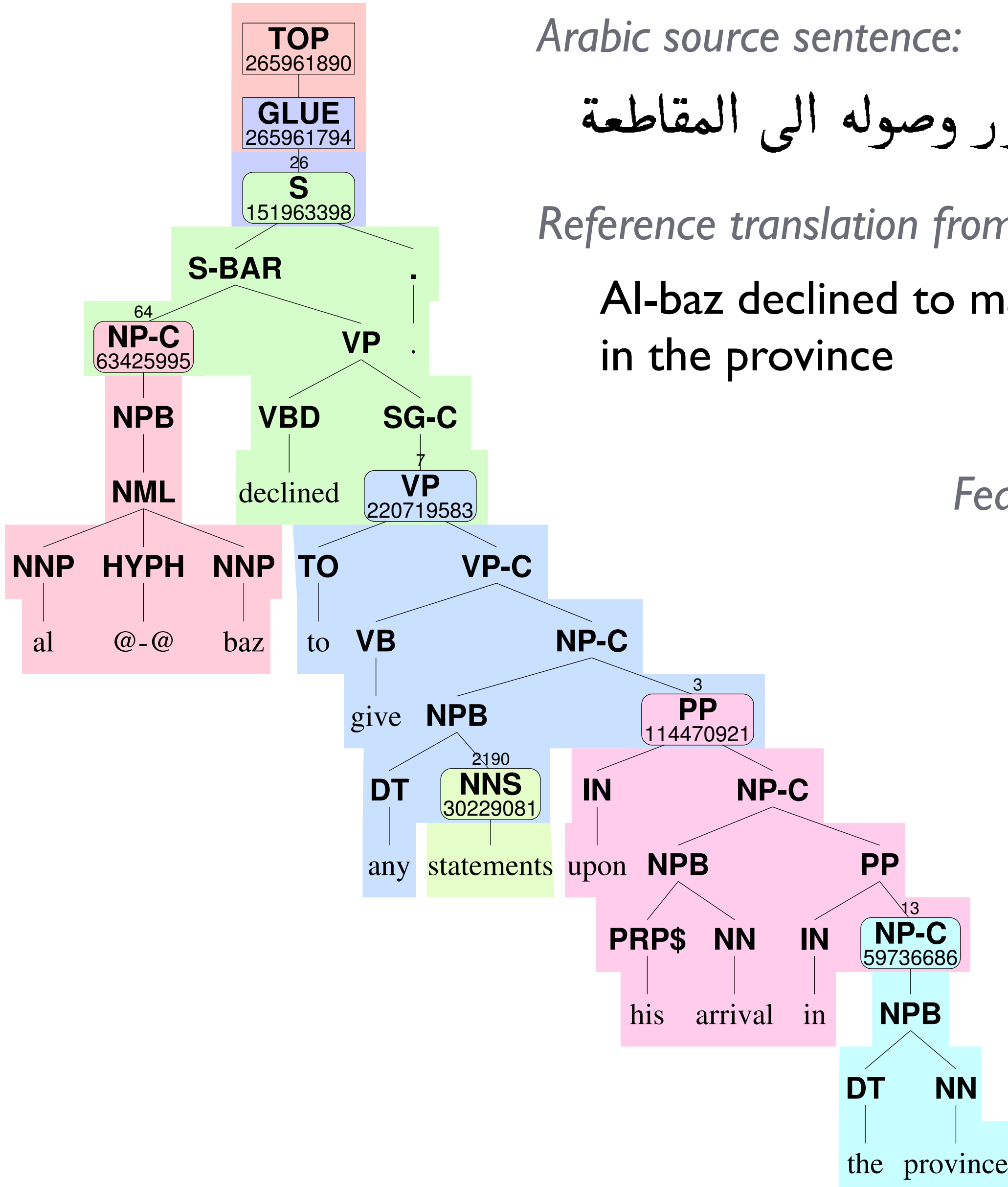
Arabic source sentence:

ورفض الباز الادلاء باى تصريحات فور وصوله الى المقاطعة

Reference translation from a human translator:

Al-baz declined to make any statements upon his arrival in the province

An Example Syntax-Based Translation



Arabic source sentence:

ورفض الباز الادلاء باى تصريحات فور وصوله الى المقاطعة

Reference translation from a human translator:

Al-baz declined to make any statements upon his arrival in the province

Features:

feature	weight	value	product
derivation-size	0.41	8	3.30
glue-rule	3.89	2	7.78
green	-0.08	0	0
gt_prob	0.40	36.18	14.43
identity	-9.97	0	0
is_lexicalized	-0.65	6	-3.91
lex_pef	1.02	5.47	5.60
lex_pfe	0.31	4.44	1.39
lm1	1	22.76	22.76
lm1-unk	30.08	0	0
lm2	0.74	26.66	19.79
lm2-unk	-39.18	0	0
missingWord	-1.29	0	0
model1inv	1.02	10.60	10.81
model1nrm	1.35	11.29	15.22
nonmonotone	4.17	0	0
olive	1.95	0	0
psm1n	0.50	24.65	12.30
text-length	-3.87	15	-58.05
trivial_cond_prob	0.41	3.34	1.38
unk-rule	19.28	0	0
reported **totalcost**	52.82	$\vec{v} \cdot \vec{w}$	52.82

The Steps in a Modern Translation System

Learn a model

Apply the model

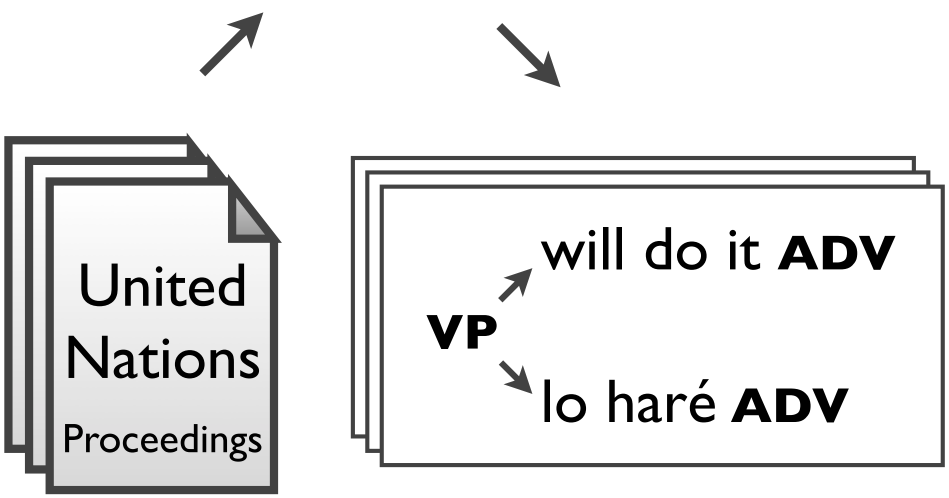
Choose a translation

The Steps in a Modern Translation System

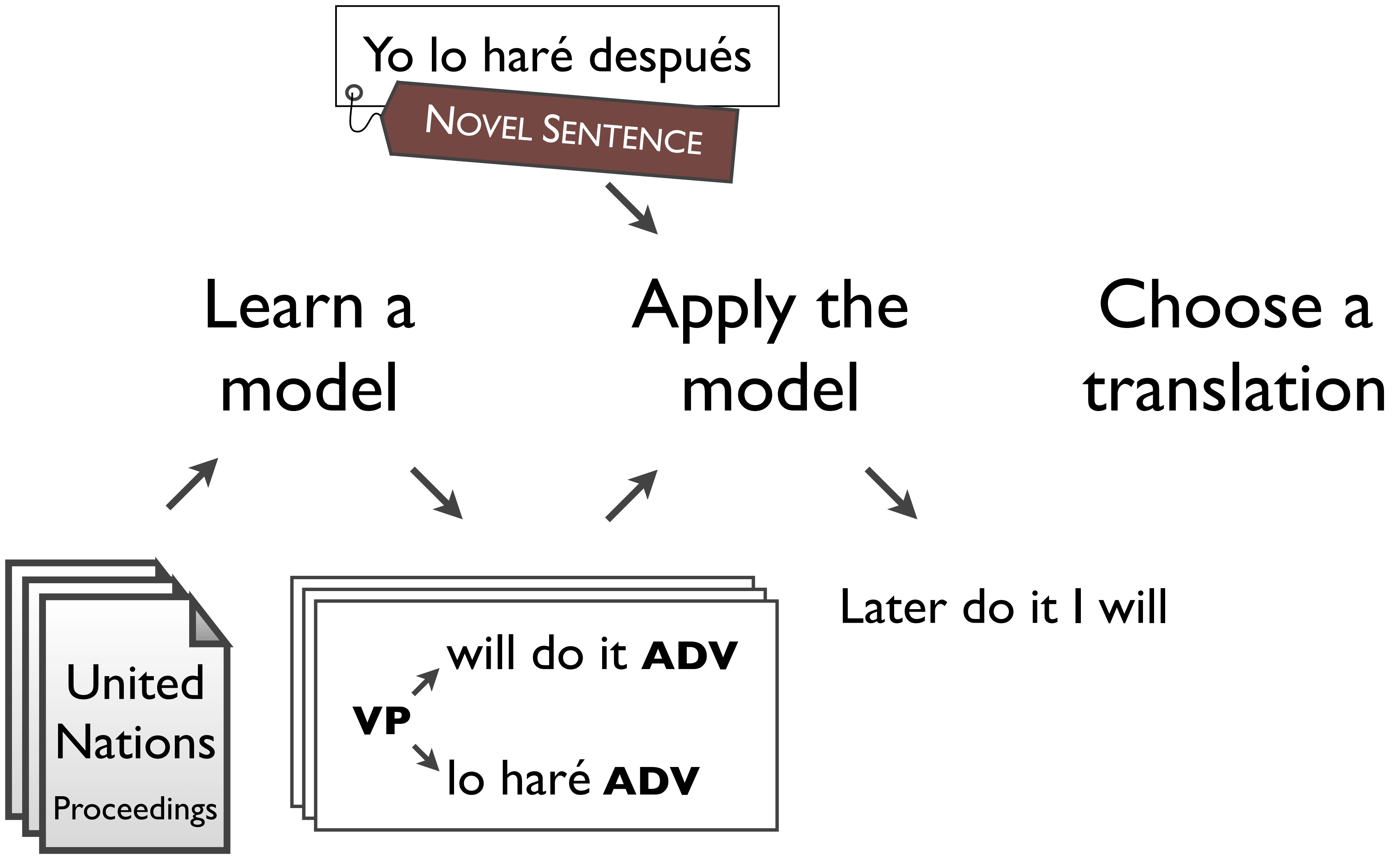
Learn a model

Apply the model

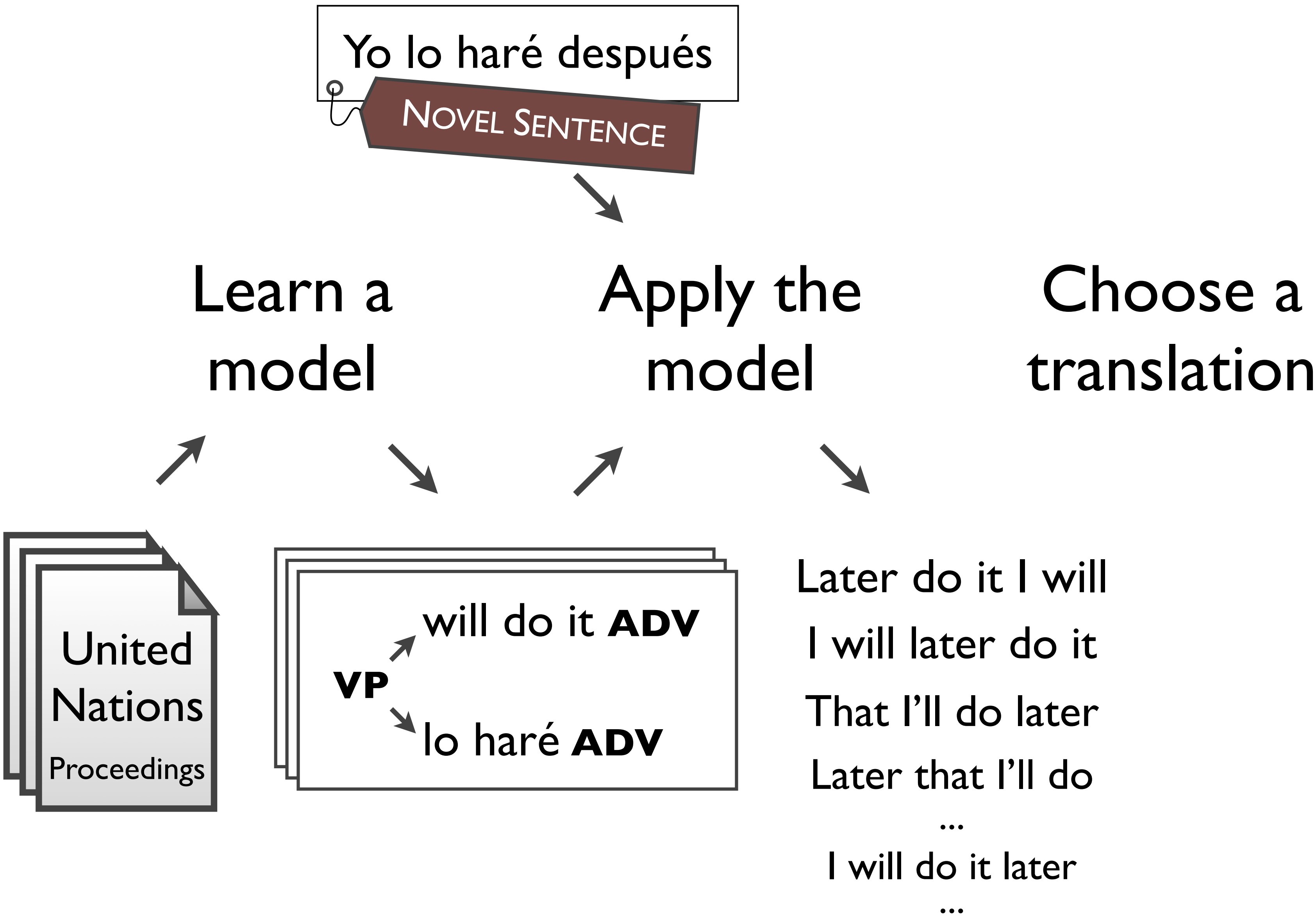
Choose a translation



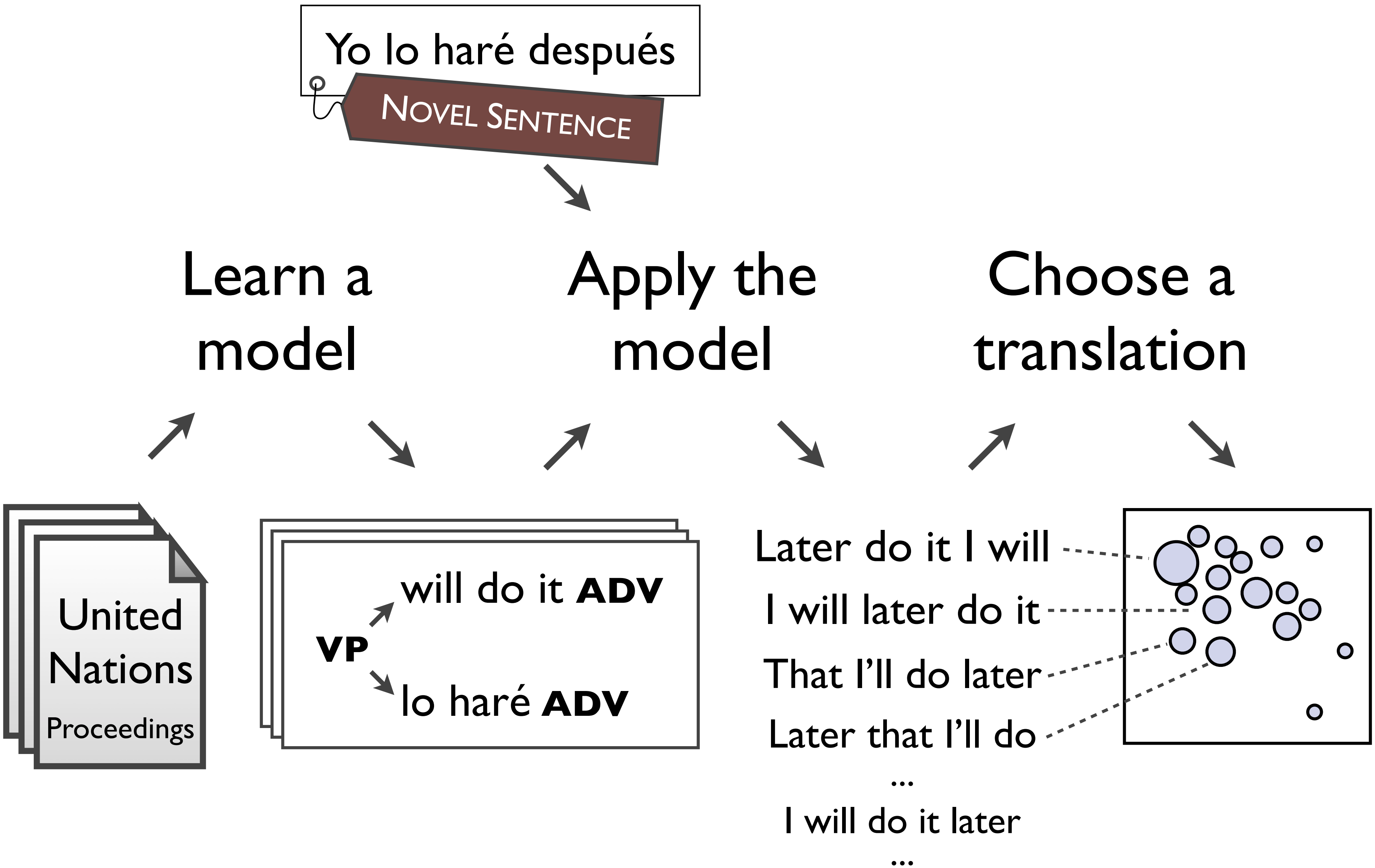
The Steps in a Modern Translation System



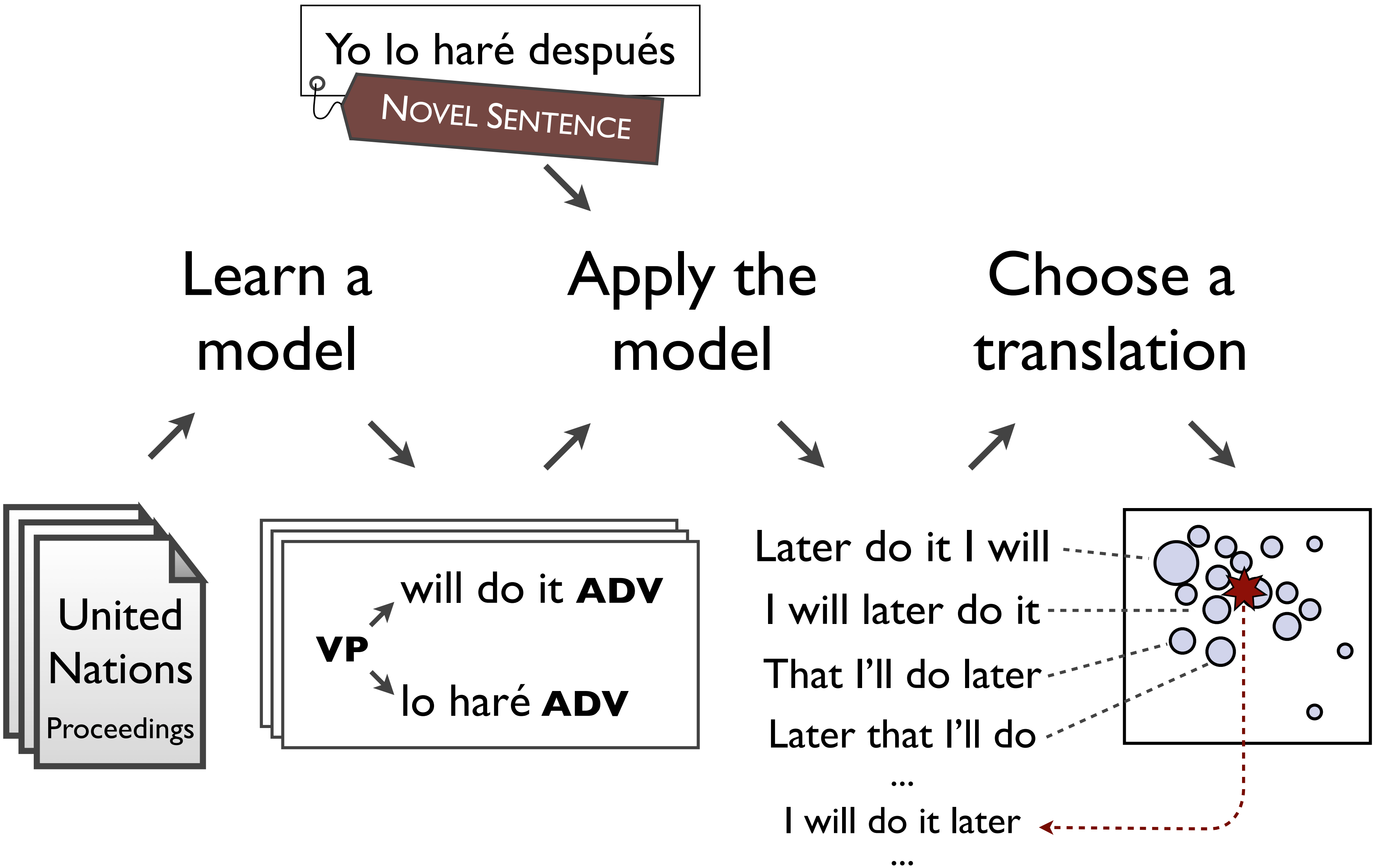
The Steps in a Modern Translation System



The Steps in a Modern Translation System



The Steps in a Modern Translation System



The Steps in a Modern Translation System

Learn a model

Apply the model

Choose a translation

The Alignment Problem in Translation

Thank you , I will do it gladly .

Gracias
,
lo
haré
de
muy
buen
grado
.

The Alignment Problem in Translation

Thank you

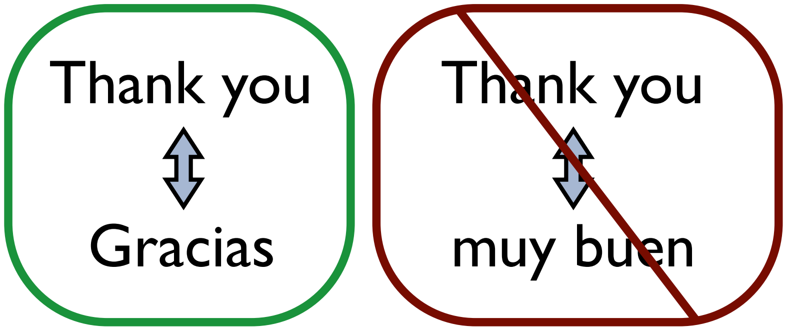
⇕

Gracias

Thank you , I will do it gladly .

Gracias
,
lo
haré
de
muy
buen
grado
.

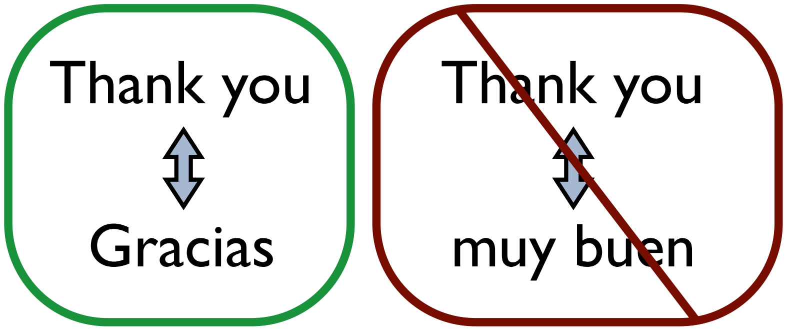
The Alignment Problem in Translation

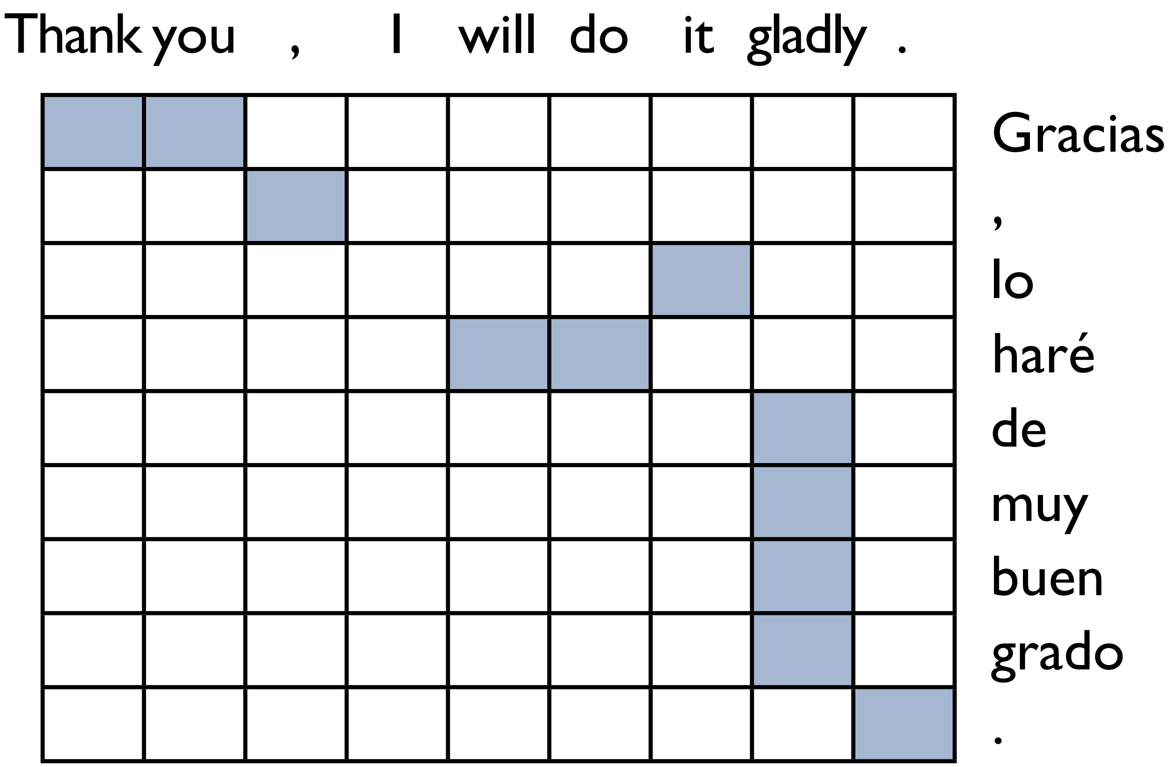


Thank you , I will do it gladly .

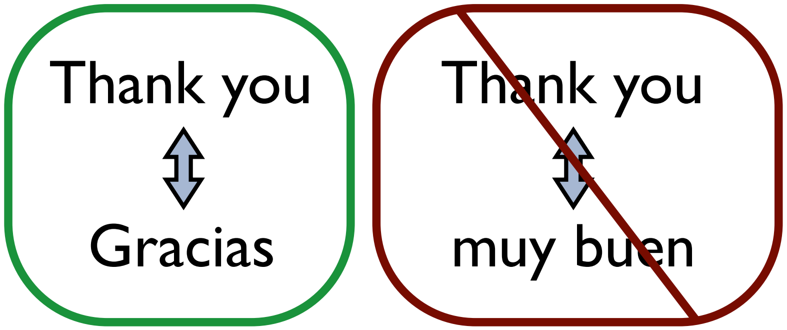
Gracias
,
lo
haré
de
muy
buen
grado
.

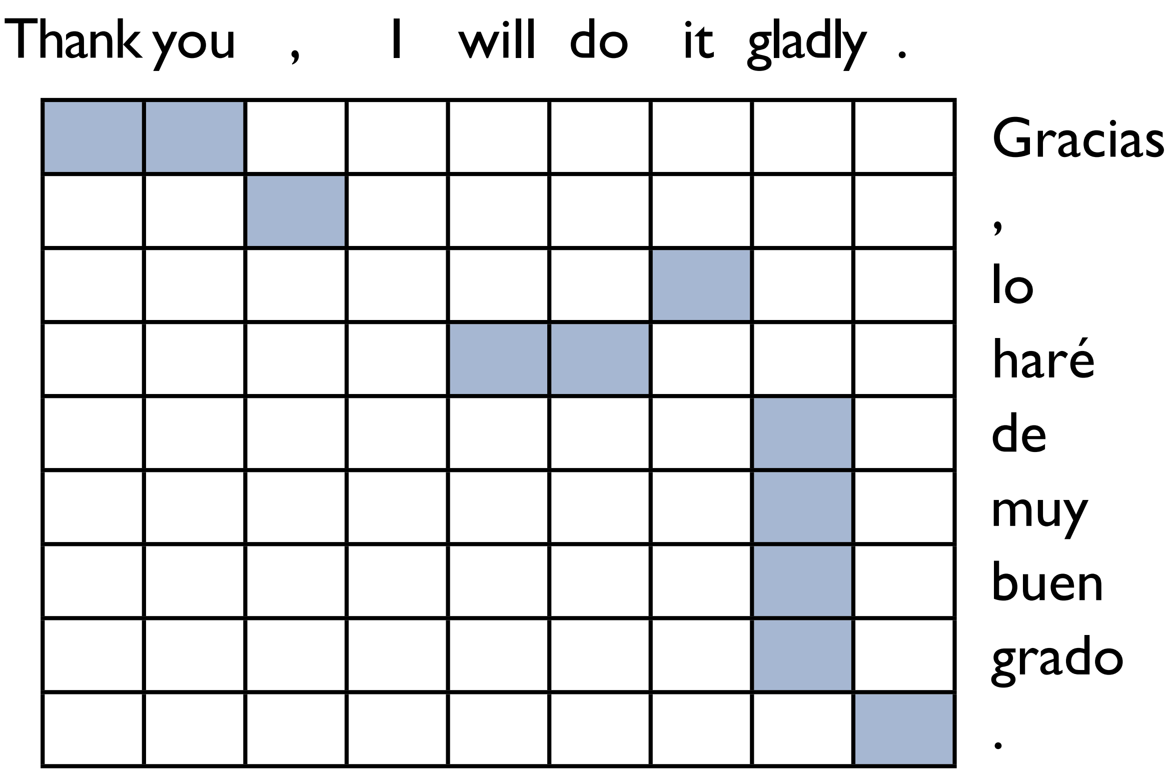
The Alignment Problem in Translation





The Alignment Problem in Translation

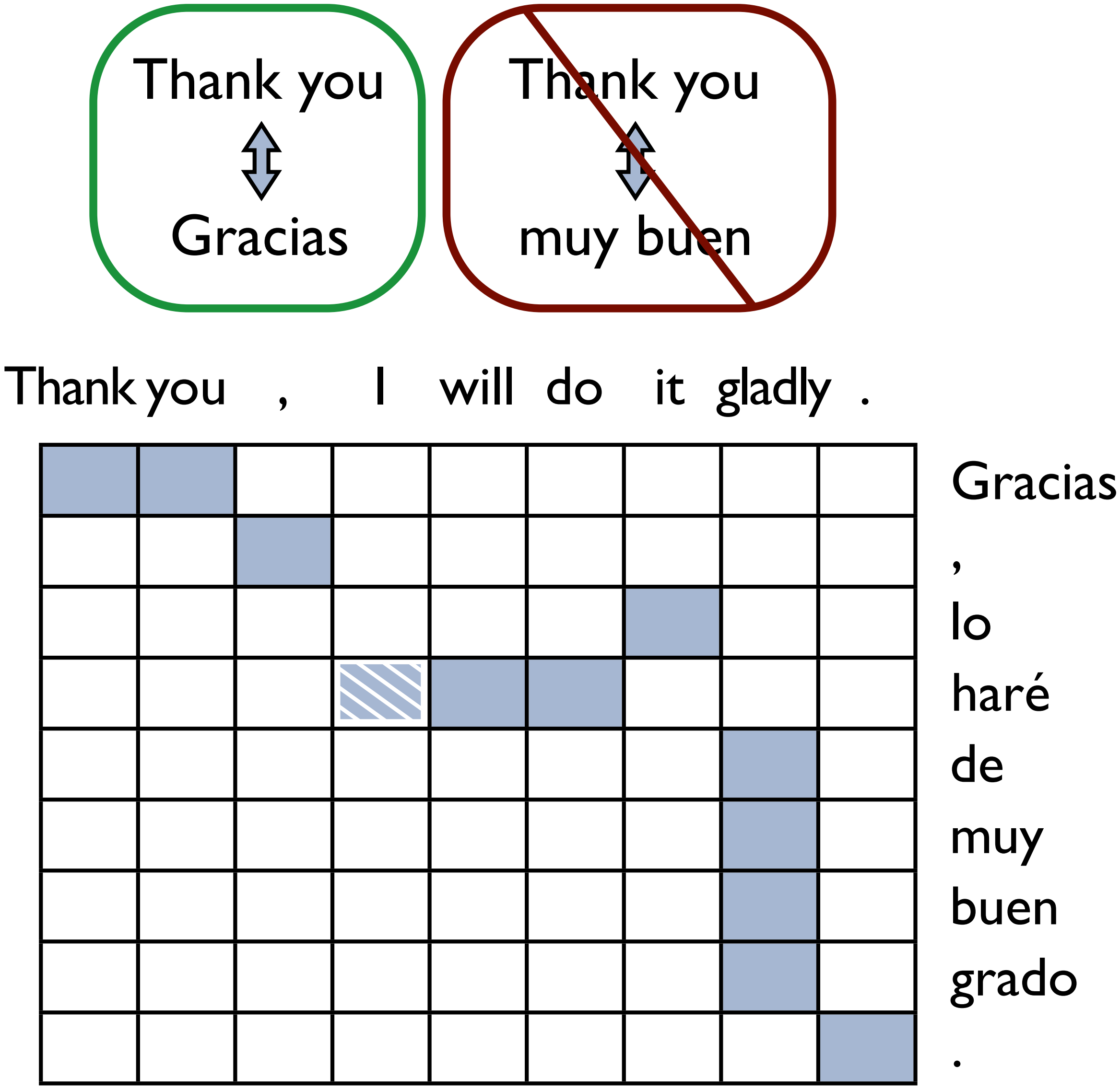




About the task:

- A lot can be inferred from lexical statistics
- Correct alignments are not one-to-one
- Some cases are tricky, even for people

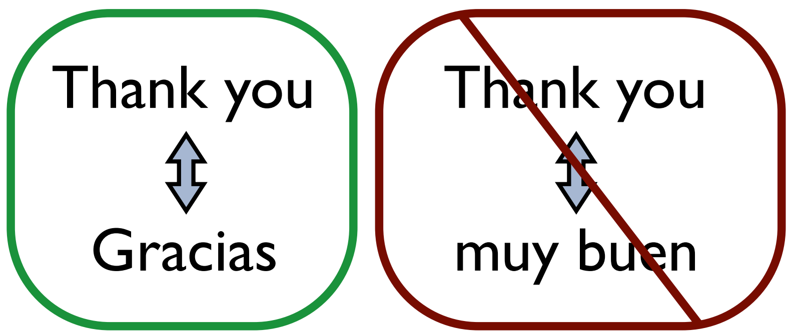
The Alignment Problem in Translation

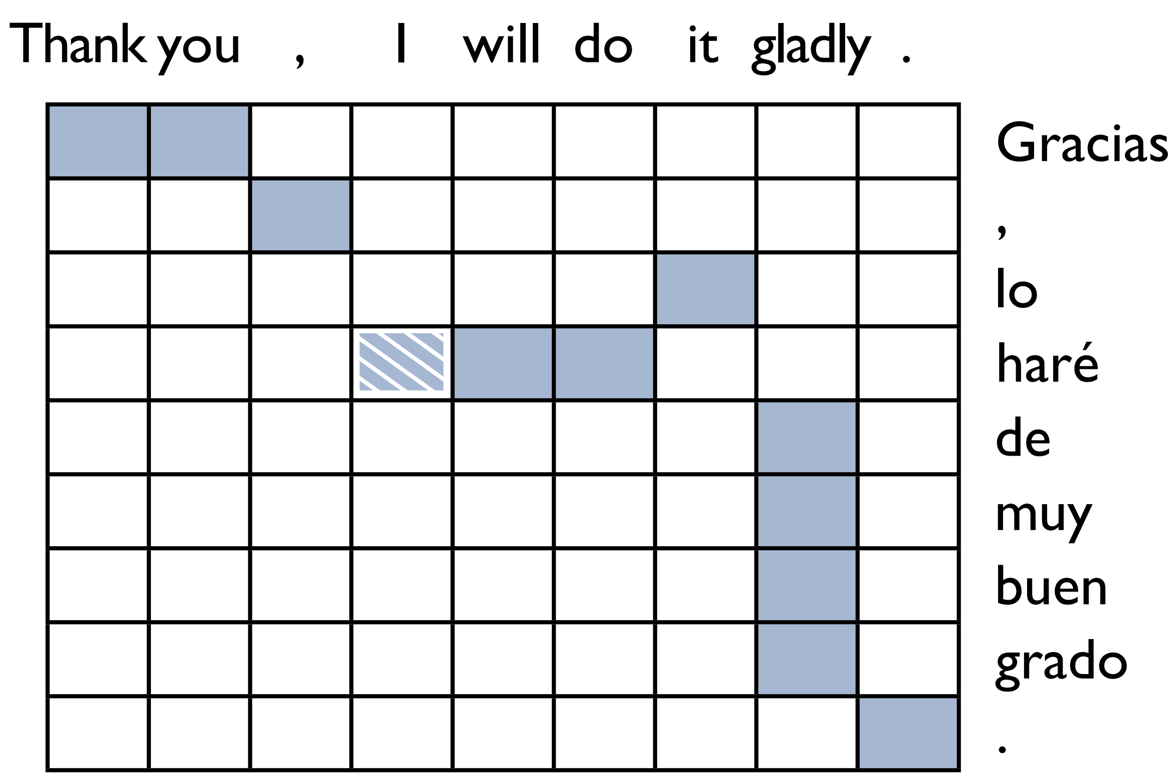


About the task:

- A lot can be inferred from lexical statistics
- Correct alignments are not one-to-one
- Some cases are tricky, even for people

The Alignment Problem in Translation





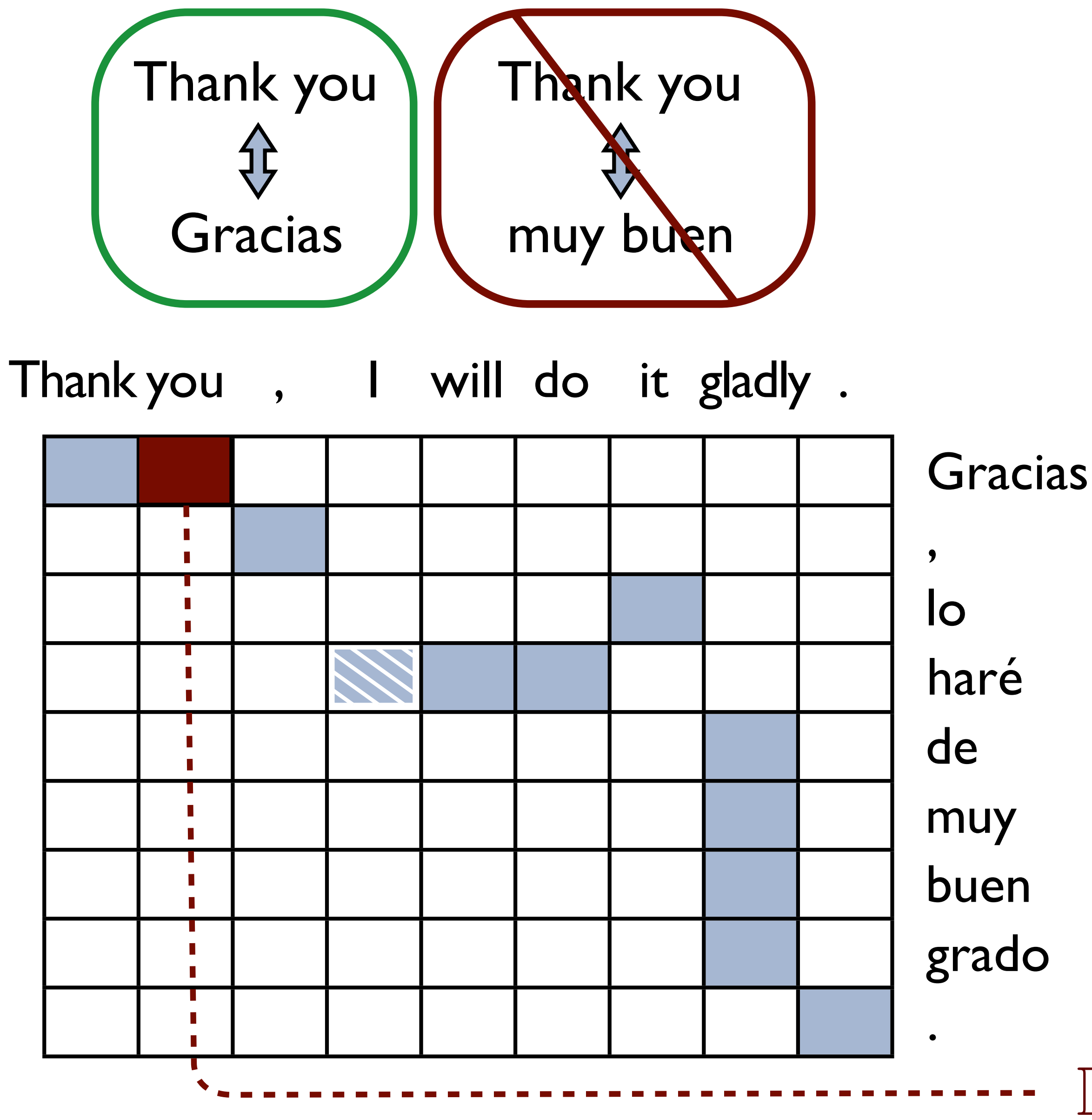
About the task:

- A lot can be inferred from lexical statistics
- Correct alignments are not one-to-one
- Some cases are tricky, even for people

About solutions:

- Word-to-word links
- Learning driven by conditional word distributions

The Alignment Problem in Translation



$\mathbb{P}(\text{gracias}|\text{you})$

About the task:

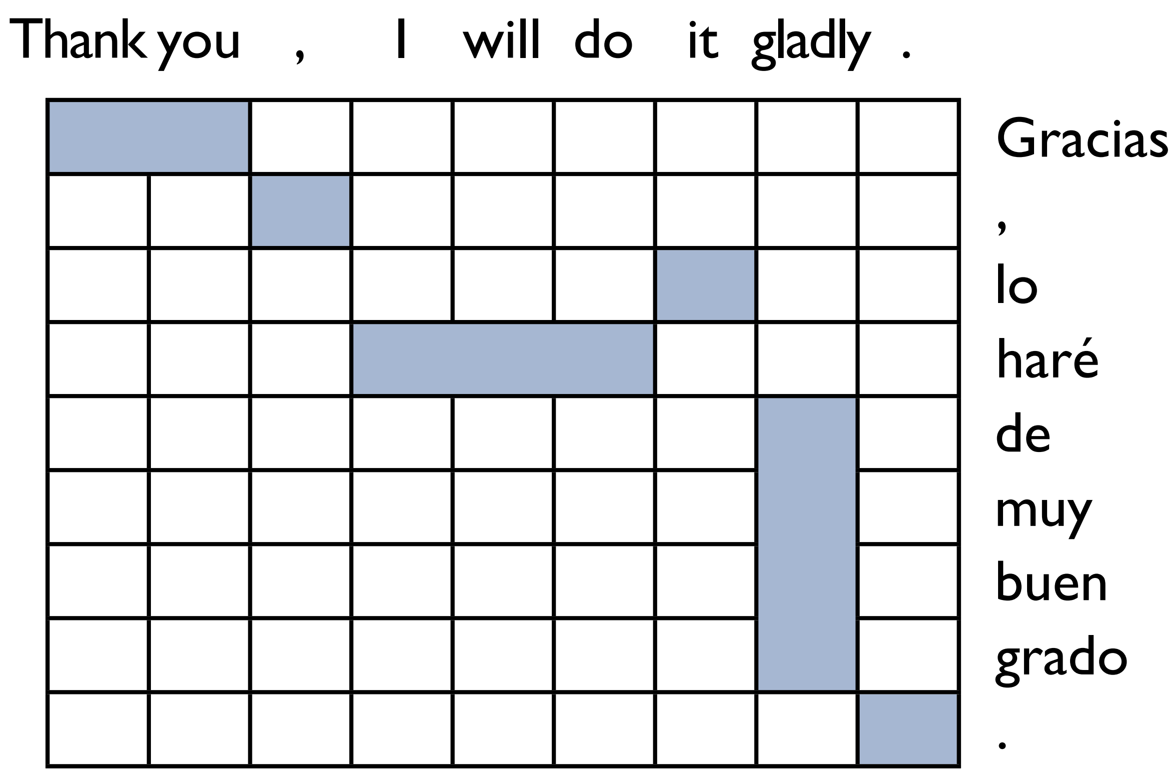
- A lot can be inferred from lexical statistics
- Correct alignments are not one-to-one
- Some cases are tricky, even for people

About solutions:

- Word-to-word links
- Learning driven by conditional word distributions

Large-Context Alignment Challenges

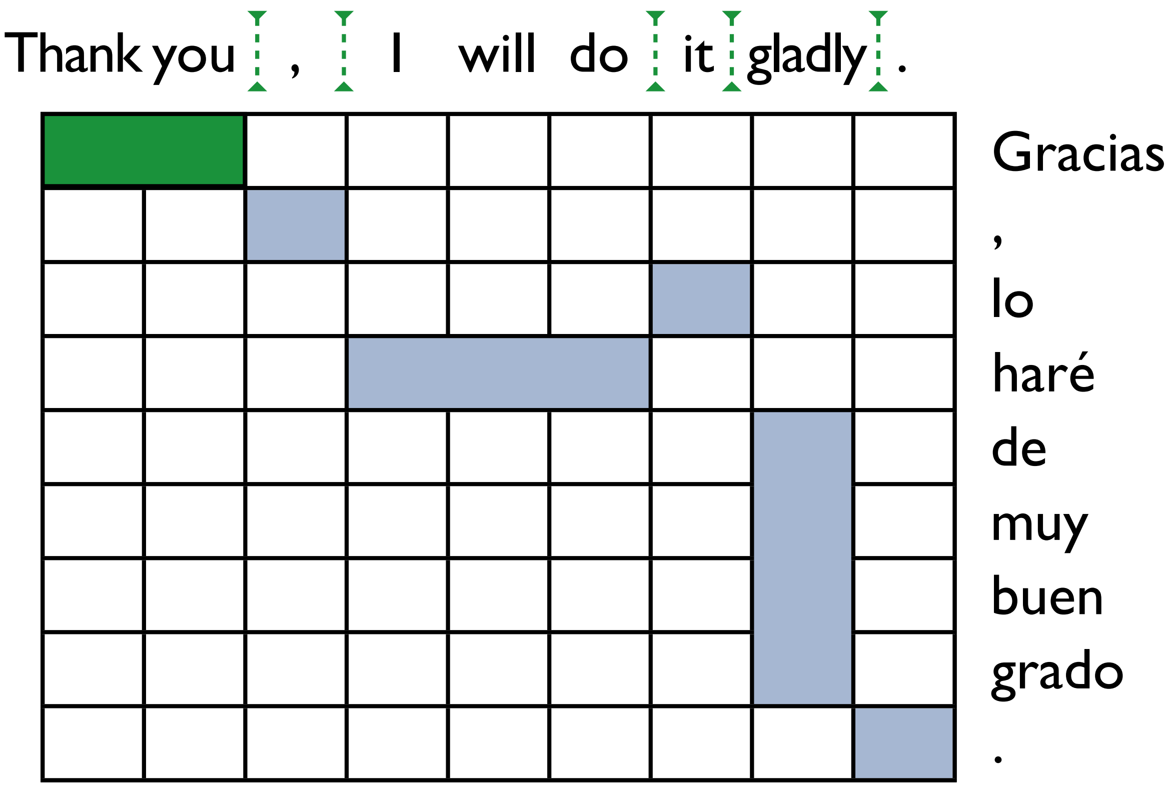
Goal: Model multi-word structures during alignment



Large-Context Alignment Challenges

Goal: Model multi-word structures during alignment

$\cancel{\mathbb{P}(\text{gracias} | \text{you})}$ $\mathbb{P}(\text{gracias}, \text{Thank you})$



Challenge $\boxed{1}$

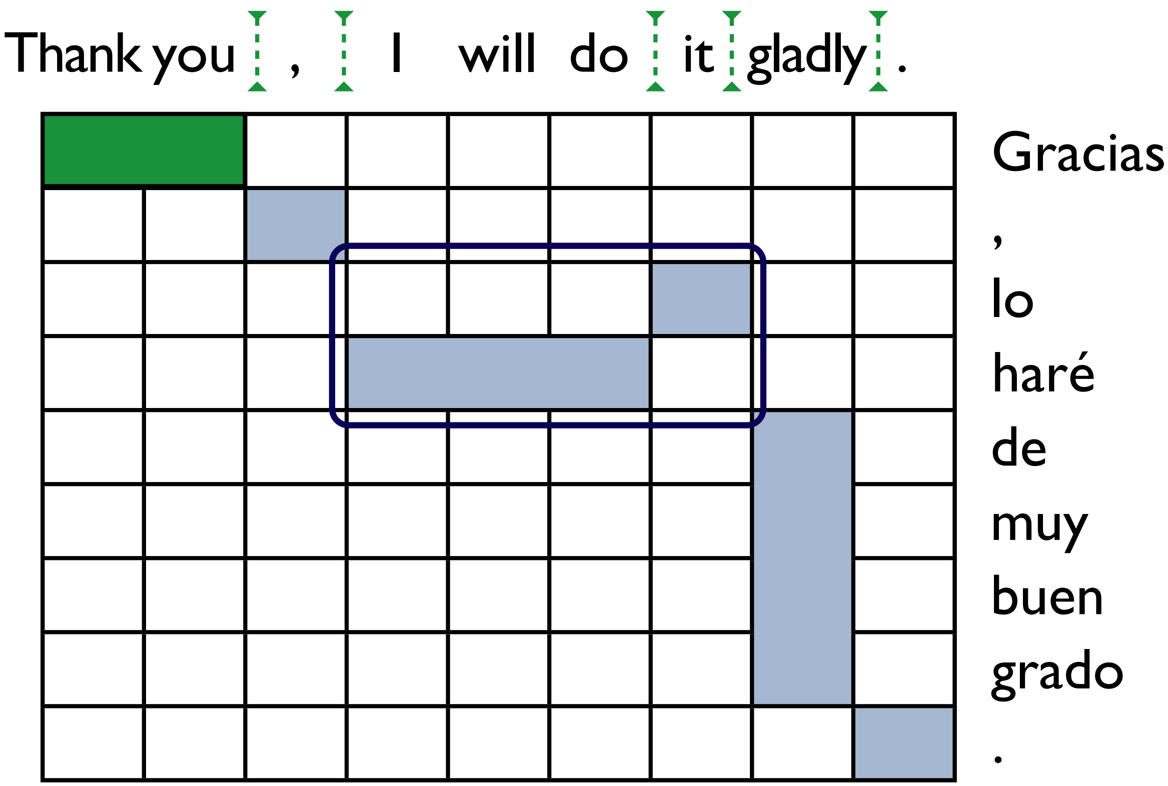
- Jointly infer phrase boundaries and alignments
- Boundaries depend on both languages

Large-Context Alignment Challenges

Goal: Model multi-word structures during alignment

$\cancel{\mathbb{P}(\text{gracias}|\text{you})}$ $\mathbb{P}(\text{gracias, Thank you})$

$\phi(\text{lo haré, I will do it})$



Challenge 1

- Jointly infer phrase boundaries and alignments
- Boundaries depend on both languages

Challenge 2

- Capture context
- Compose phrases

Modeling Phrasal Correspondence

Paradigm: Train a generative model that explains observed translations via latent structure

Modeling Phrasal Correspondence

Paradigm: Train a generative model that explains observed translations via latent structure

Process: Phrase pairs are generated independently

Modeling Phrasal Correspondence

Paradigm: Train a generative model that explains observed translations via latent structure

Process: Phrase pairs are generated independently

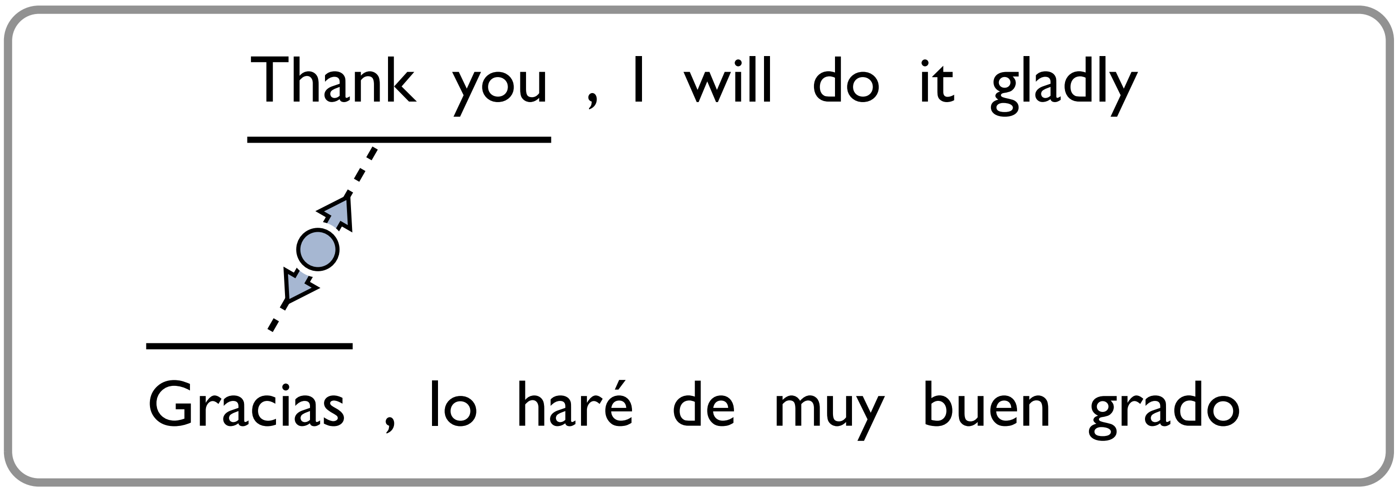
Thank you , I will do it gladly

Gracias , lo haré de muy buen grado

Modeling Phrasal Correspondence

Paradigm: Train a generative model that explains observed translations via latent structure

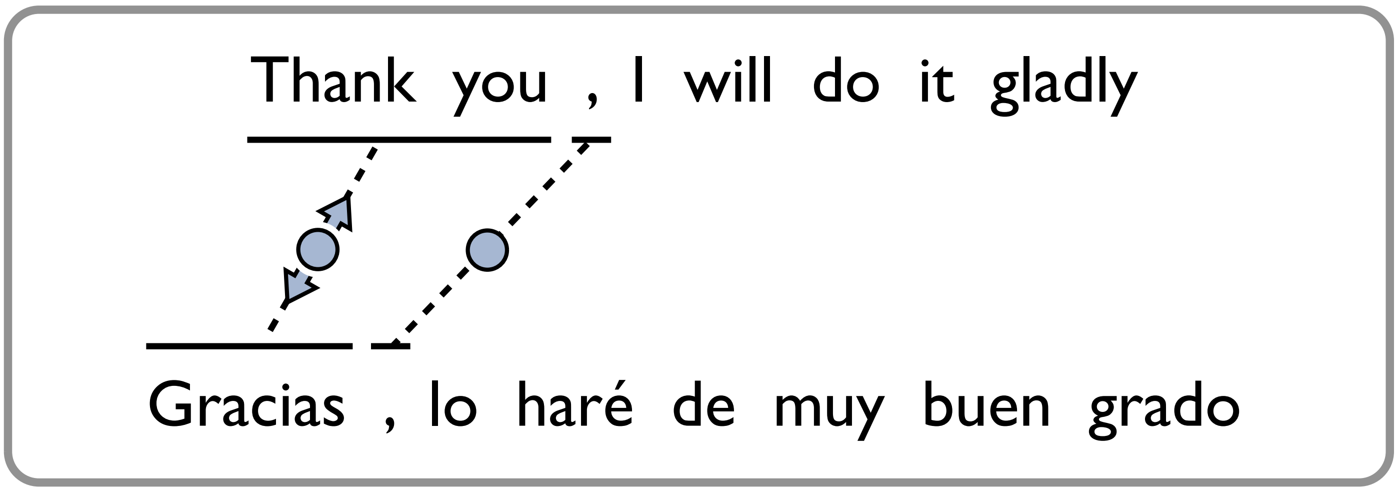
Process: Phrase pairs are generated independently



Modeling Phrasal Correspondence

Paradigm: Train a generative model that explains observed translations via latent structure

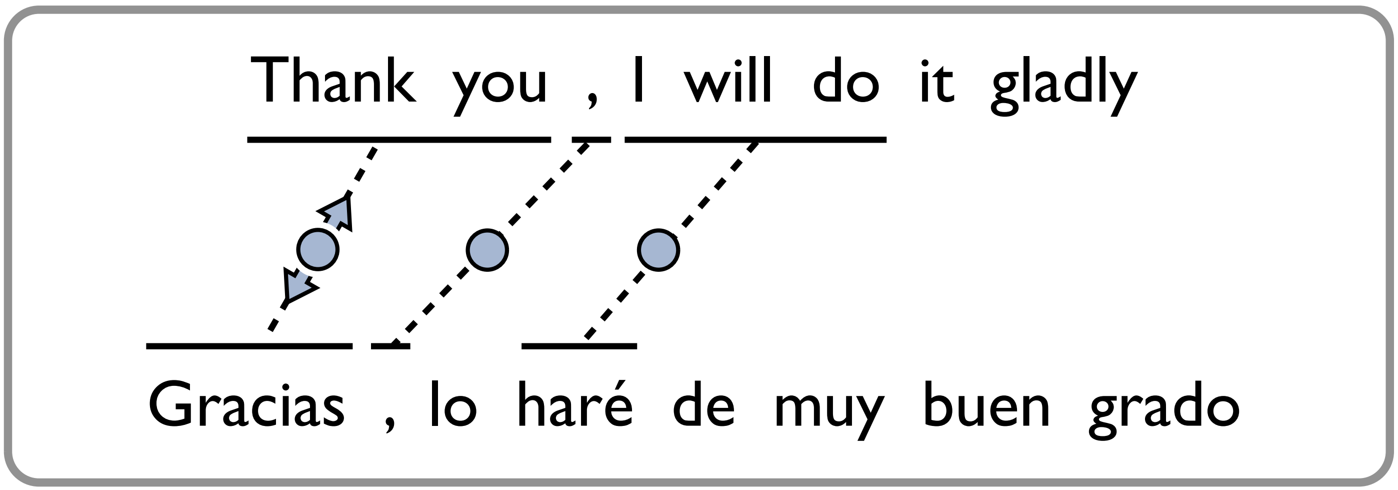
Process: Phrase pairs are generated independently



Modeling Phrasal Correspondence

Paradigm: Train a generative model that explains observed translations via latent structure

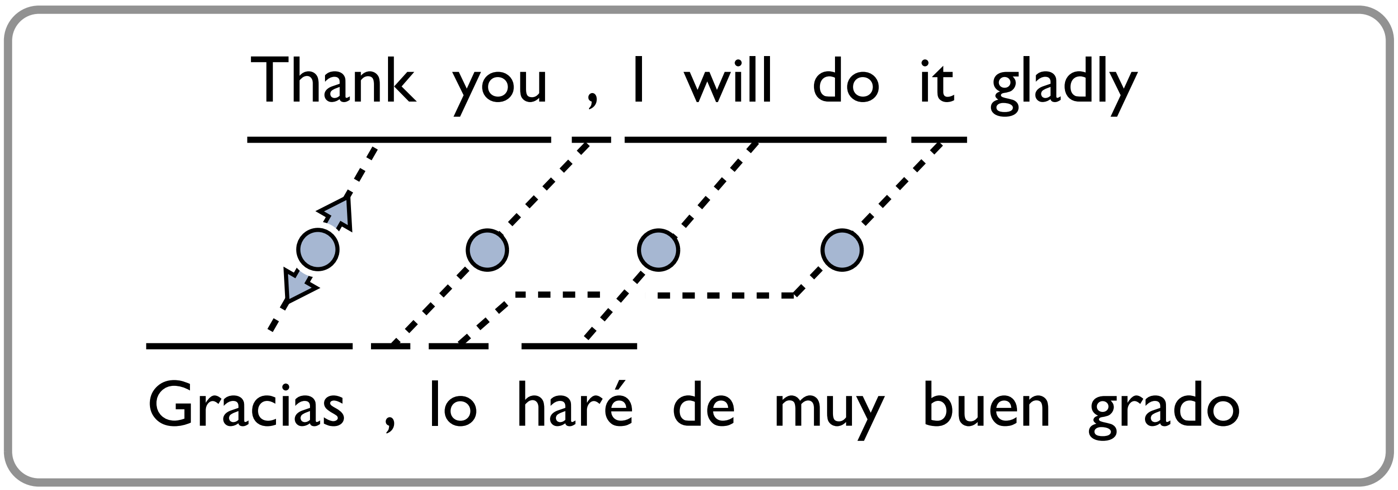
Process: Phrase pairs are generated independently



Modeling Phrasal Correspondence

Paradigm: Train a generative model that explains observed translations via latent structure

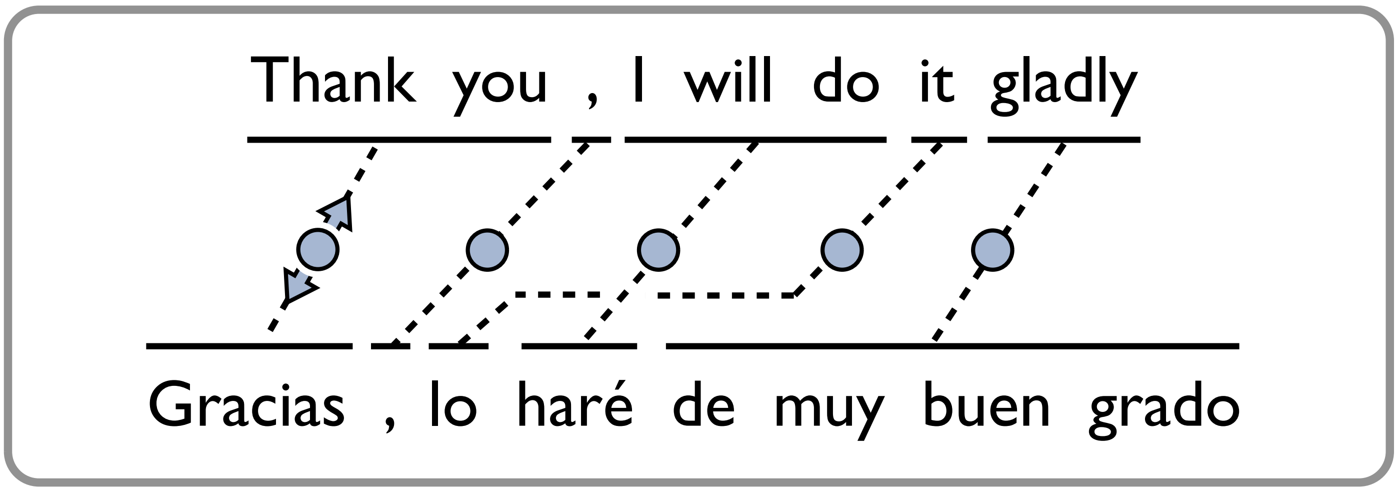
Process: Phrase pairs are generated independently



Modeling Phrasal Correspondence

Paradigm: Train a generative model that explains observed translations via latent structure

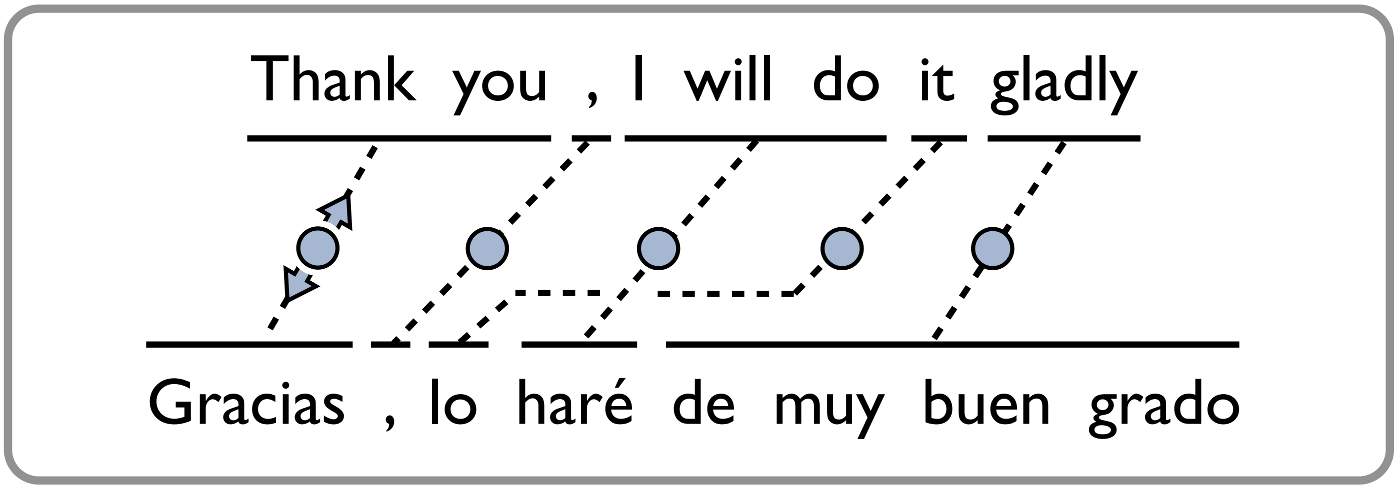
Process: Phrase pairs are generated independently



Modeling Phrasal Correspondence

Paradigm: Train a generative model that explains observed translations via latent structure

Process: Phrase pairs are generated independently

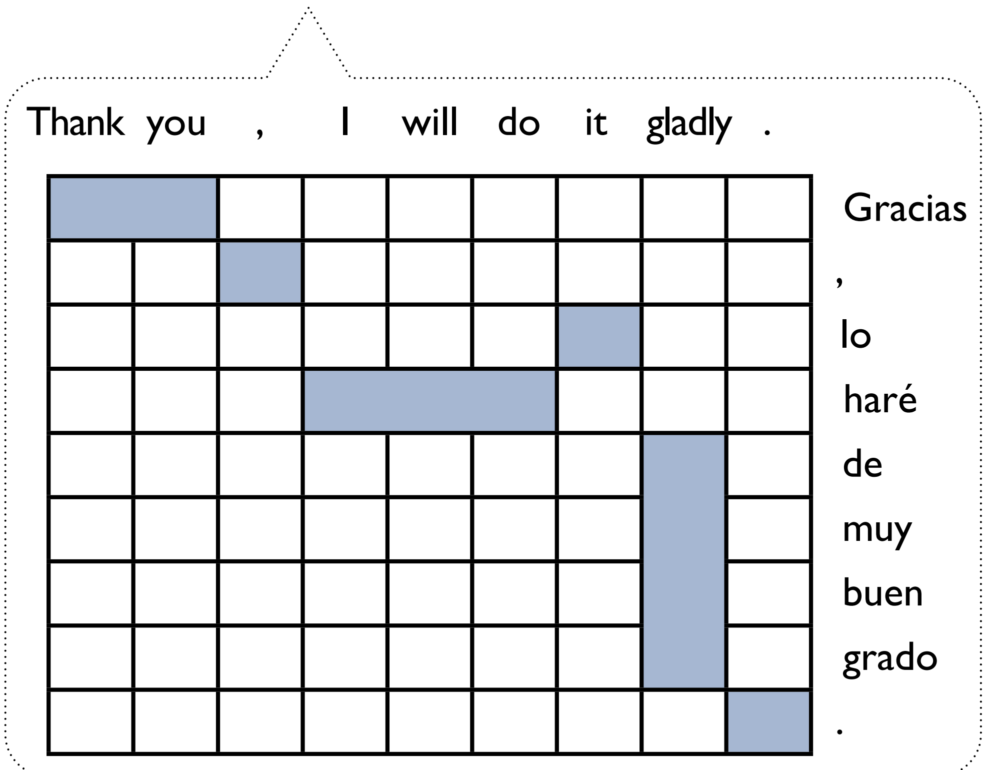


Optimization: Explain all translations with shared parameters

Modeling Phrasal Correspondence

We learn θ, a multinomial distribution over phrase pairs

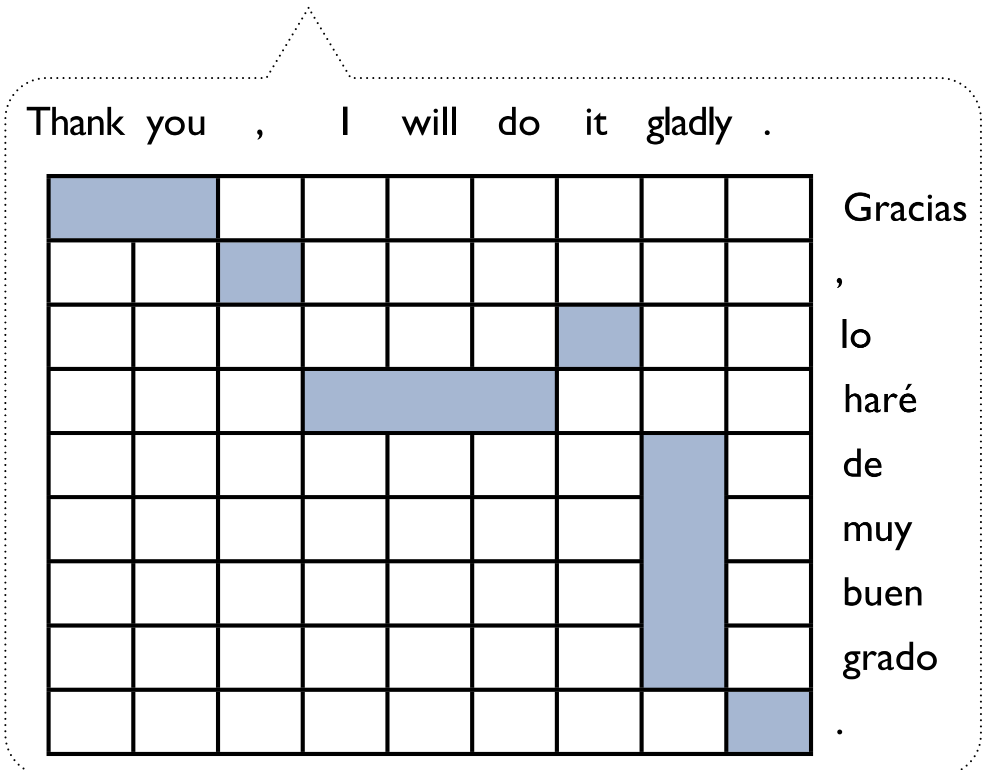
$$\mathbb{P}(A = a) = \theta(\text{Thank you, Gracias}) \;\cdot\; \theta(\text{I will do, haré}) \;\cdot\; \theta(\text{it, lo}) \cdots$$



Modeling Phrasal Correspondence

We learn θ, a multinomial distribution over phrase pairs

$$\mathbb{P}(A = a) = \theta(\text{Thank you, Gracias}) \;\cdot\; \theta(\text{I will do, haré}) \;\cdot\; \theta(\text{it, lo}) \cdots$$



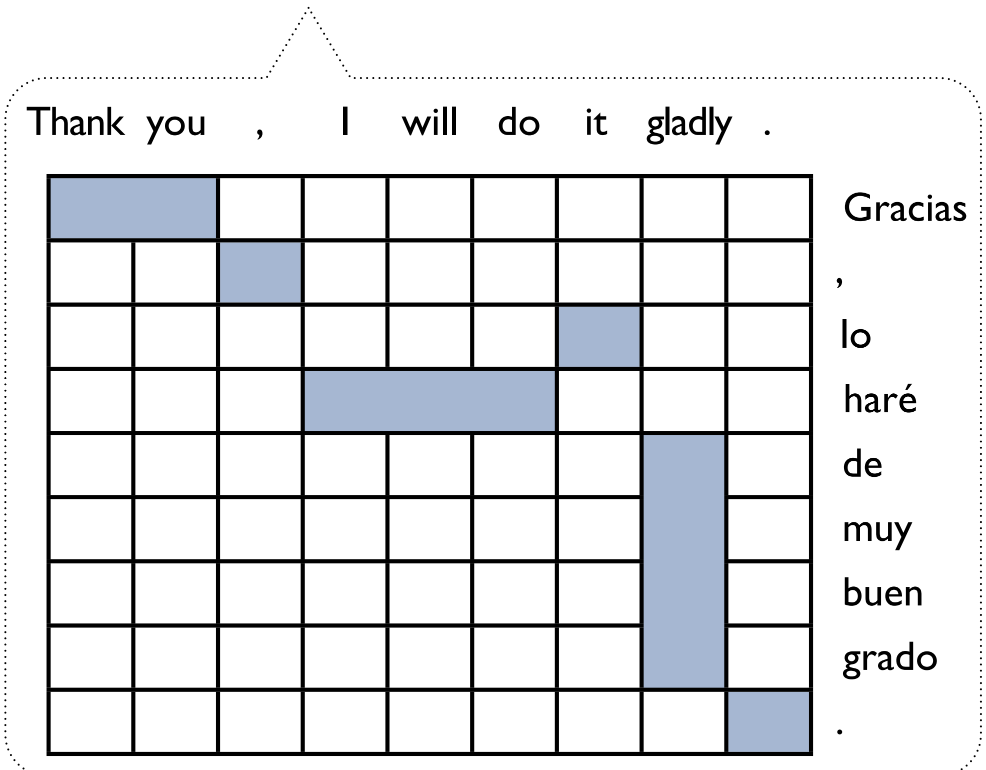
$$P(A = a) = \prod_{(e,s)\in a} \theta(e, s) \; ^{*}$$

* Terms omitted: Phrase pair count and phrase permutation

Modeling Phrasal Correspondence

We learn θ, a multinomial distribution over phrase pairs

$$\mathbb{P}(A = a) = \theta(\text{Thank you, Gracias}) \cdot \theta(\text{I will do, haré}) \cdot \theta(\text{it, lo}) \cdots$$



$$P(A = a) = \prod_{(e,s) \in a} \theta(e, s) \; ^{*}$$

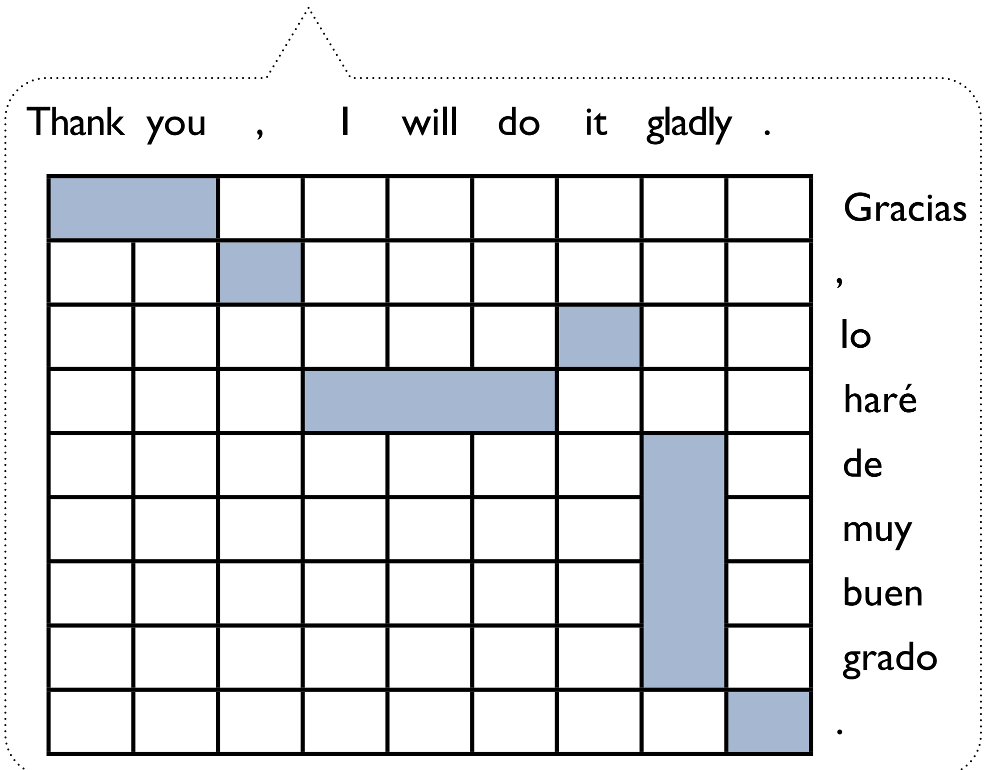
$$\mathcal{L}(\theta) = \prod_{d \in D} \left[\sum_{a \in \mathcal{A}(d)} P(A = a) \right]$$

* Terms omitted: Phrase pair count and phrase permutation

Modeling Phrasal Correspondence

We learn θ, a multinomial distribution over phrase pairs

$$\mathbb{P}(A = a) = \theta(\text{Thank you, Gracias}) \cdot \theta(\text{I will do, haré}) \cdot \theta(\text{it, lo}) \cdots$$



$$P(A = a) = \prod_{(e,s) \in a} \theta(e, s)^{\,*}$$

For each sentence pair:

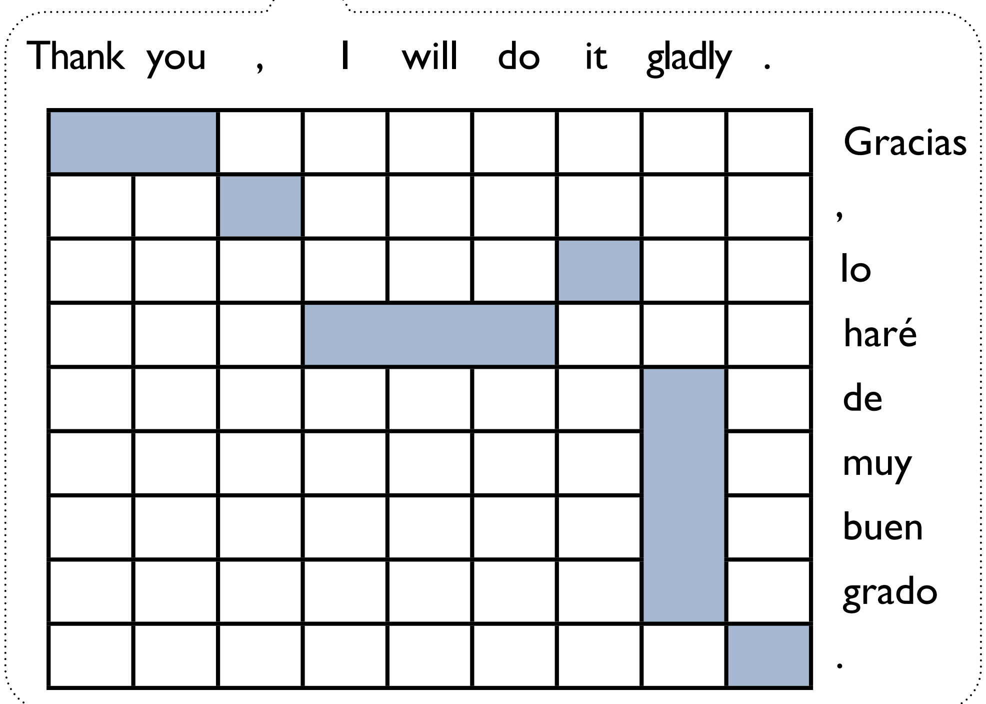
$$\mathcal{L}(\theta) = \prod_{d \in D} \left[\sum_{a \in \mathcal{A}(d)} P(A = a) \right]$$

* Terms omitted: Phrase pair count and phrase permutation

Modeling Phrasal Correspondence

We learn θ, a multinomial distribution over phrase pairs

$$\mathbb{P}(A = a) = \theta(\text{Thank you, Gracias}) \cdot \theta(\text{I will do, haré}) \cdot \theta(\text{it, lo}) \cdots$$



$$P(A = a) = \prod_{(e,s) \in a} \theta(e, s)^{\,*}$$

For each sentence pair:

For each alignment:

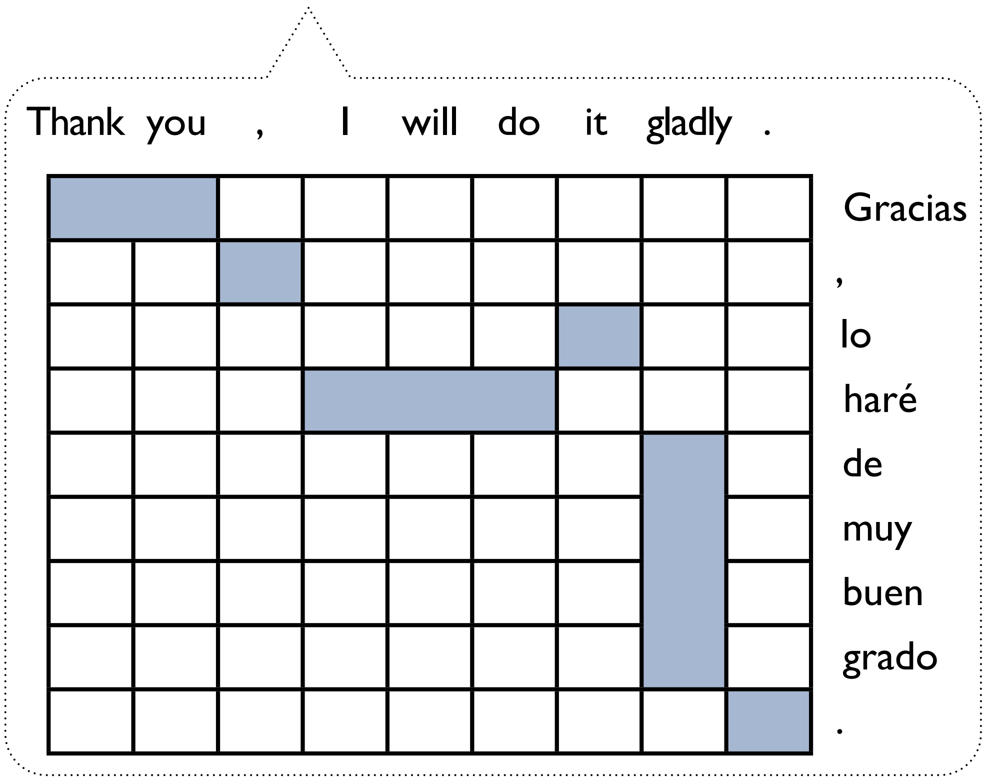
$$\mathcal{L}(\theta) = \prod_{d \in D} \left[\sum_{a \in \mathcal{A}(d)} P(A = a) \right]$$

* Terms omitted: Phrase pair count and phrase permutation

Modeling Phrasal Correspondence

We learn θ, a multinomial distribution over phrase pairs

$$\mathbb{P}(A = a) = \theta(\text{Thank you, Gracias}) \cdot \theta(\text{I will do, haré}) \cdot \theta(\text{it, lo}) \cdots$$



$$P(A = a) = \prod_{(e,s)\in a} \theta(e, s)^{\,*}$$

For each sentence pair:

For each alignment:

$$\mathcal{L}(\theta) = \prod_{d \in D} \left[\sum_{a \in \mathcal{A}(d)} P(A = a) \right]$$

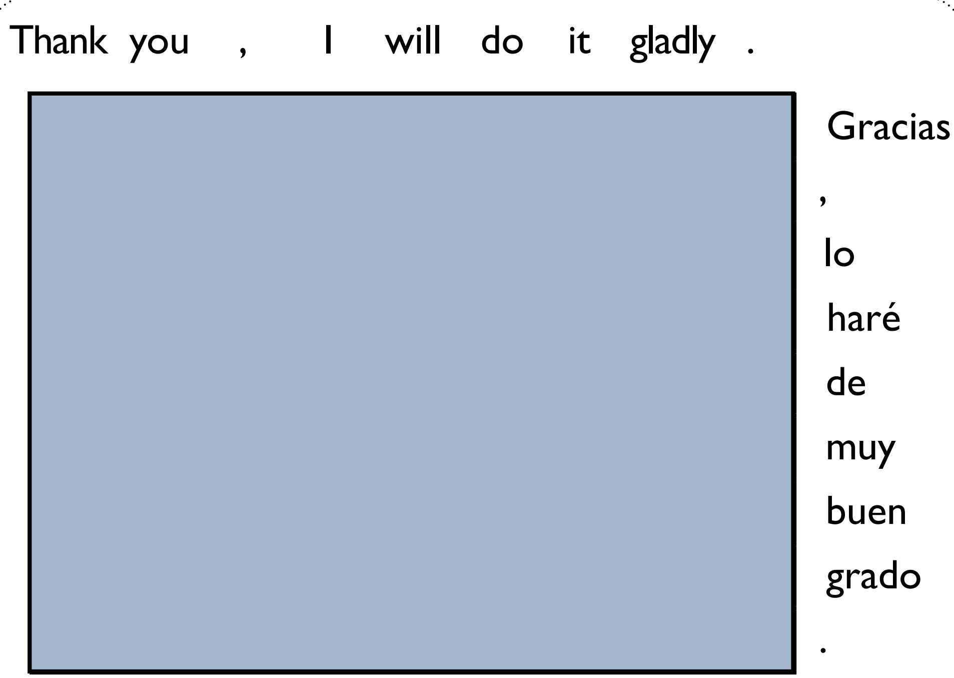
Maximizing likelihood gives a degenerate solution: huge phrases!

* Terms omitted: Phrase pair count and phrase permutation

Modeling Phrasal Correspondence

We learn θ, a multinomial distribution over phrase pairs

$$\mathbb{P}(A = a) = \theta(\text{Thank you, Gracias}) \cdot \theta(\text{I will do, haré}) \cdot \theta(\text{it, lo}) \cdots$$



$$P(A = a) = \prod_{(e,s)\in a} \theta(e, s)^{*}$$

For each sentence pair:

For each alignment:

$$\mathcal{L}(\theta) = \prod_{d \in D} \left[\sum_{a \in \mathcal{A}(d)} P(A = a) \right]$$

Maximizing likelihood gives a degenerate solution: huge phrases!

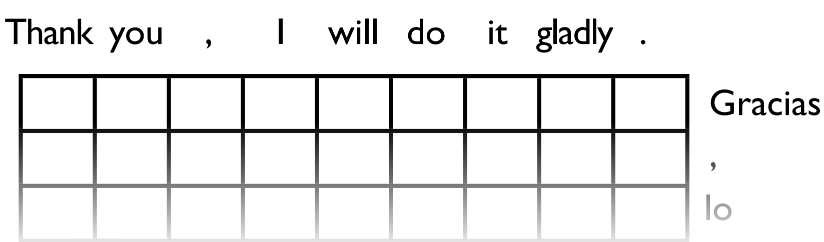
* Terms omitted: Phrase pair count and phrase permutation

Guiding Phrasal Correspondence Models

$$\theta \sim \mathrm{DP}(\theta_0, \alpha)$$

Base distribution: θ_0 Prefers short phrases

Dirichlet process: $\mathrm{DP}(\,\cdot\,, \alpha)$ Non-parametric cache model



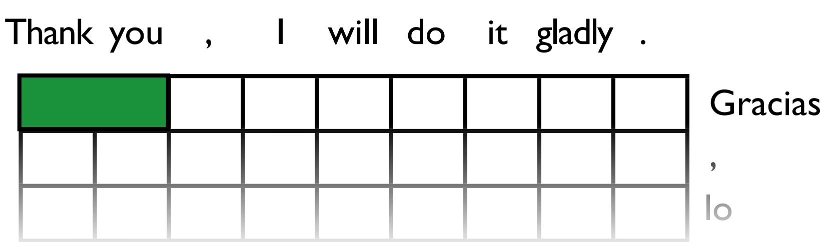
Phrase Pair Cache (c):

English-Spanish phrase pair	Count
...	
(Thank you, Gracias)	||||
(Thanks, Gracias)	|||
(Thank you, Muchas gracias)	||
...	

Guiding Phrasal Correspondence Models

$$\theta \sim \mathrm{DP}(\theta_0, \alpha)$$

Base distribution:	θ_0	Prefers short phrases
Dirichlet process:	$\mathrm{DP}(\ \cdot\ , \alpha)$	Non-parametric cache model



Phrase Pair Cache (c):

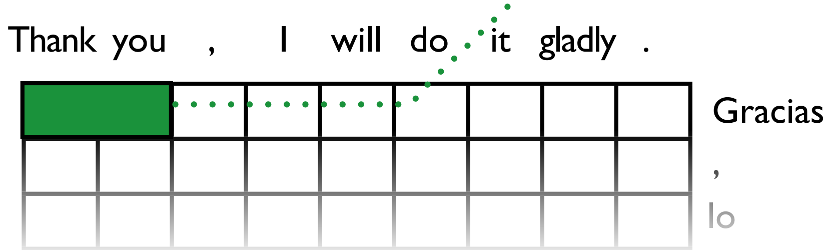
English-Spanish phrase pair	Count
...	
(Thank you, Gracias)	||||
(Thanks, Gracias)	|||
(Thank you, Muchas gracias)	||
...	

Guiding Phrasal Correspondence Models

$$\theta \sim \mathrm{DP}(\theta_0, \alpha)$$

Base distribution: $\boxed{\theta_0}$ Prefers short phrases

Dirichlet process: $\mathrm{DP}(\,\cdot\,, \alpha)$ Non-parametric cache model



Phrase Pair Cache (c):

<u>English-Spanish phrase pair</u> <u>Count</u>

...

(Thank you, Gracias) | | | |

(Thanks, Gracias) | | |

(Thank you, Muchas gracias) | |

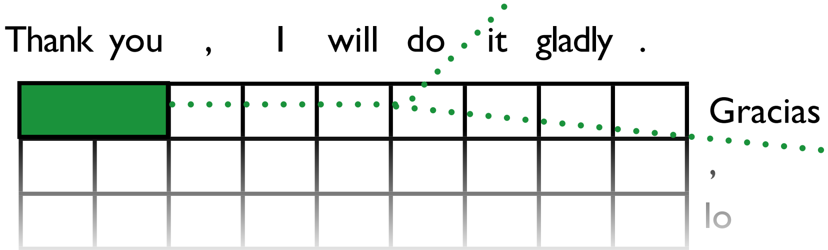
...

Guiding Phrasal Correspondence Models

$$\theta \sim \mathrm{DP}(\theta_0, \alpha)$$

Base distribution: $\boxed{\theta_0}$ Prefers short phrases

Dirichlet process: $\mathrm{DP}(\;\cdot\;, \alpha)$ Non-parametric cache model



Phrase Pair Cache (c):

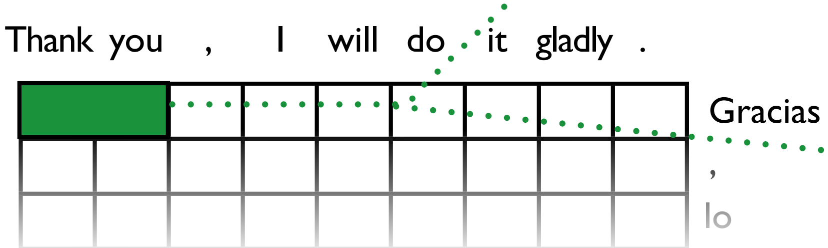
<table>
<tr><th>English-Spanish phrase pair</th><th>Count</th></tr>
<tr><td>...</td><td></td></tr>
<tr><td>(Thank you, Gracias)</td><td>| | | |</td></tr>
<tr><td>(Thanks, Gracias)</td><td>| | |</td></tr>
<tr><td>(Thank you, Muchas gracias)</td><td>| |</td></tr>
<tr><td>...</td><td></td></tr>
</table>

Guiding Phrasal Correspondence Models

$$\theta \sim \mathrm{DP}(\theta_0, \alpha)$$

Base distribution: $\boxed{\theta_0}$ Prefers short phrases

Dirichlet process: $\mathrm{DP}(\ \cdot\ , \alpha)$ Non-parametric cache model



Phrase Pair Cache (c):

<u>English-Spanish phrase pair</u> <u>Count</u>

...

(Thank you, Gracias) ||||

(Thanks, Gracias) |||

(Thank you, Muchas gracias) ||

...

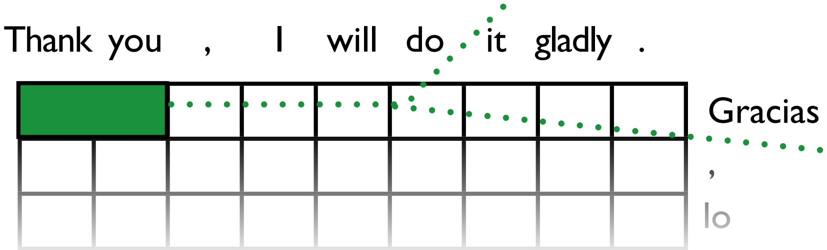
$$\mathbb{P}(z|c) = \frac{c(z) + \alpha \cdot \theta_0(z)}{|c| + \alpha}$$

Guiding Phrasal Correspondence Models

$$\theta \sim \mathrm{DP}(\theta_0, \alpha)$$

Base distribution: $\boxed{\theta_0}$ Prefers short phrases

Dirichlet process: $\mathrm{DP}(\ \cdot\ , \alpha)$ Non-parametric cache model



Phrase Pair Cache (c):

<table>
<tr><th>English-Spanish phrase pair</th><th>Count</th></tr>
<tr><td>...</td><td></td></tr>
<tr><td>(Thank you, Gracias)</td><td>||||</td></tr>
<tr><td>(Thanks, Gracias)</td><td>|||</td></tr>
<tr><td>(Thank you, Muchas gracias)</td><td>||</td></tr>
<tr><td>...</td><td></td></tr>
</table>

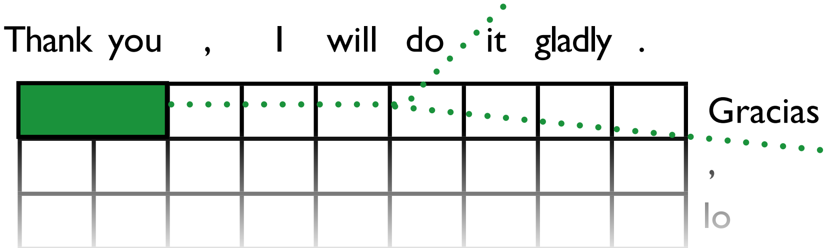
$$\mathbb{P}(z|c) = \frac{c(z) + \alpha \cdot \theta_0(z)}{|c| + \alpha}$$

Guiding Phrasal Correspondence Models

$$\theta \sim \mathrm{DP}(\theta_0, \alpha)$$

Base distribution: $\boxed{\theta_0}$ Prefers short phrases

Dirichlet process: $\mathrm{DP}(\ \cdot\ , \alpha)$ Non-parametric cache model



$$\mathbb{P}(z|c) = \frac{\boxed{c(z)} + \boxed{\alpha \cdot \theta_0(z)}}{\boxed{|c|} + \boxed{\alpha}}$$

Phrase Pair Cache (c):

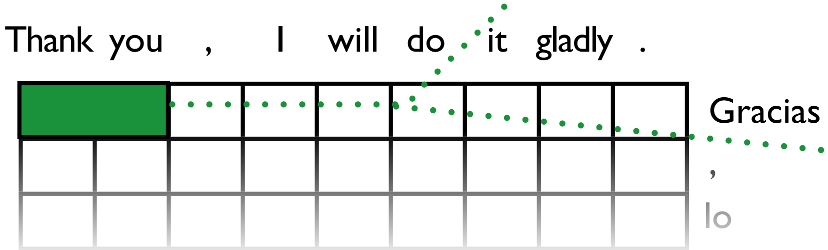
English-Spanish phrase pair	Count
...	
(Thank you, Gracias)	$\boxed{\text{||||}}$
(Thanks, Gracias)	|||
(Thank you, Muchas gracias)	||
...	

Guiding Phrasal Correspondence Models

$$\theta \sim \mathrm{DP}(\theta_0, \alpha)$$

Base distribution: $\boxed{\theta_0}$ Prefers short phrases

Dirichlet process: $\mathrm{DP}(\ \cdot\ , \alpha)$ Non-parametric cache model



Phrase Pair Cache (c):

English-Spanish phrase pair	Count
...	
(Thank you, Gracias)	$\boxed{\vert\vert\vert\vert}$
(Thanks, Gracias)	$\vert\vert\vert$
(Thank you, Muchas gracias)	$\vert\vert$
...	

$$\mathbb{P}(z|c) = \frac{\boxed{c(z)} + \boxed{\alpha \cdot \theta_0(z)}}{\boxed{|c|} + \boxed{\alpha}}$$

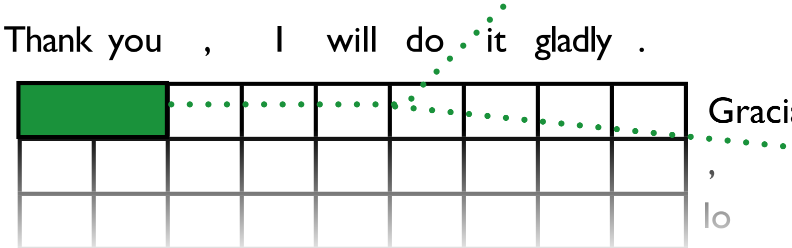
grow *fixed*

Guiding Phrasal Correspondence Models

$$\theta \sim \mathrm{DP}(\theta_0, \alpha)$$

Base distribution: $\boxed{\theta_0}$ Prefers short phrases

Dirichlet process: $\mathrm{DP}(\ \cdot\ , \alpha)$ Non-parametric cache model



Phrase Pair Cache (c):

English-Spanish phrase pair	Count
...	
(Thank you, Gracias)	||||
(Thanks, Gracias)	|||
(Thank you, Muchas gracias)	||
...	

$$\mathbb{P}(z|c) = \frac{c(z) + \alpha \cdot \theta_0(z)}{|c| + \alpha}$$

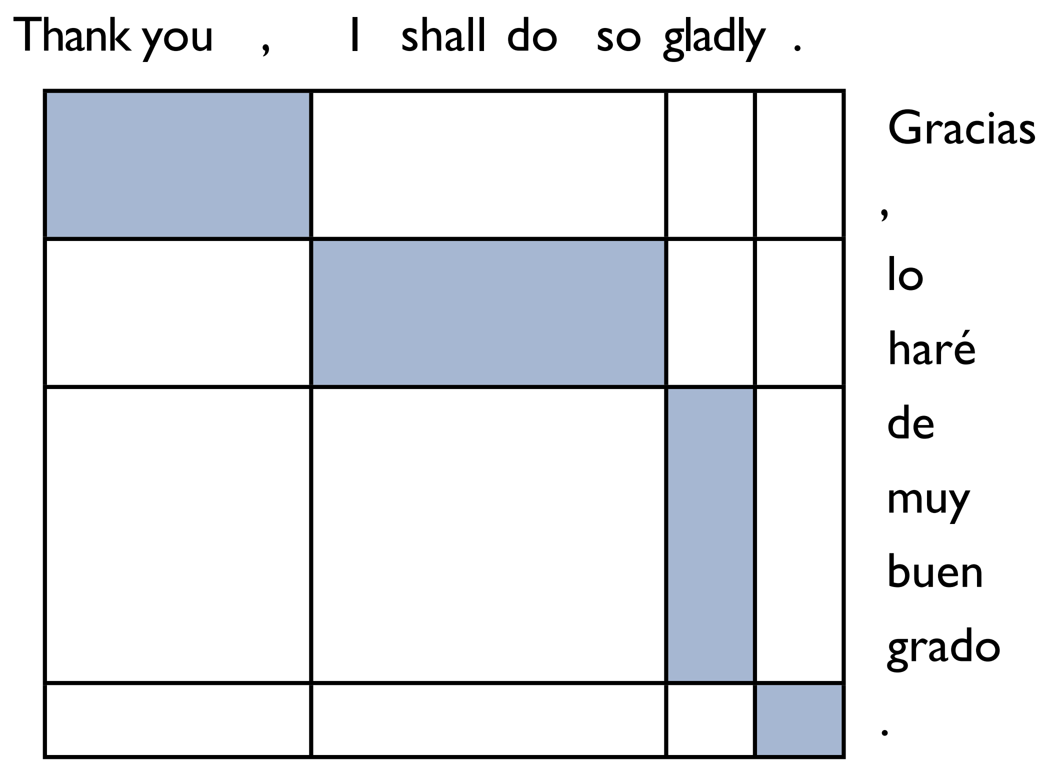
grow *fixed*

Iterative realignment of all the data by sampling → Consistent, efficient estimation

What Happens in Practice

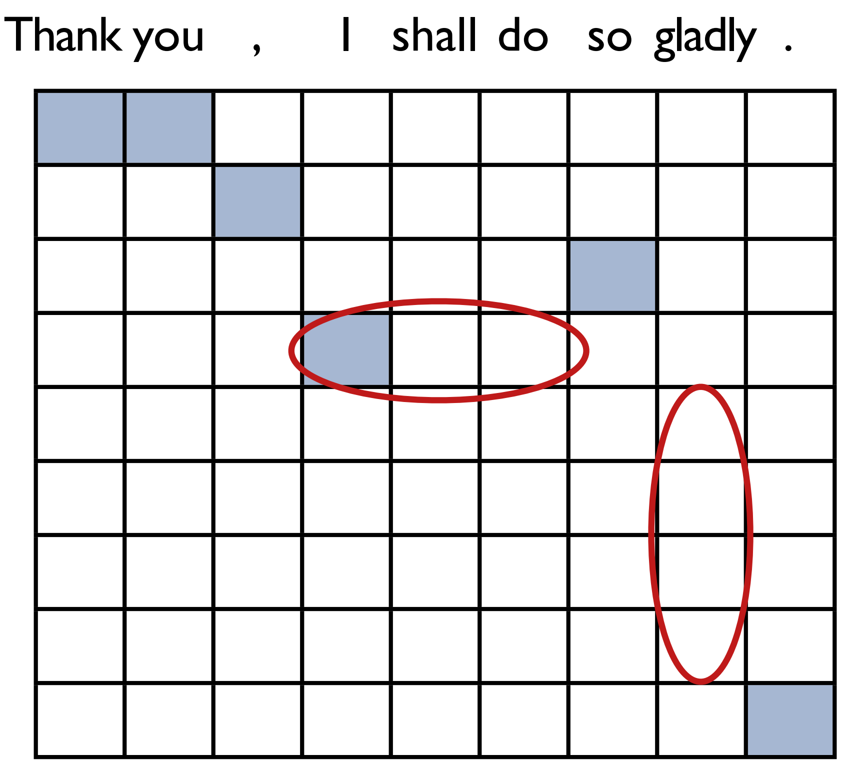
A state-of-the-art
word-level alignment

A sampled phrase alignment
from our system

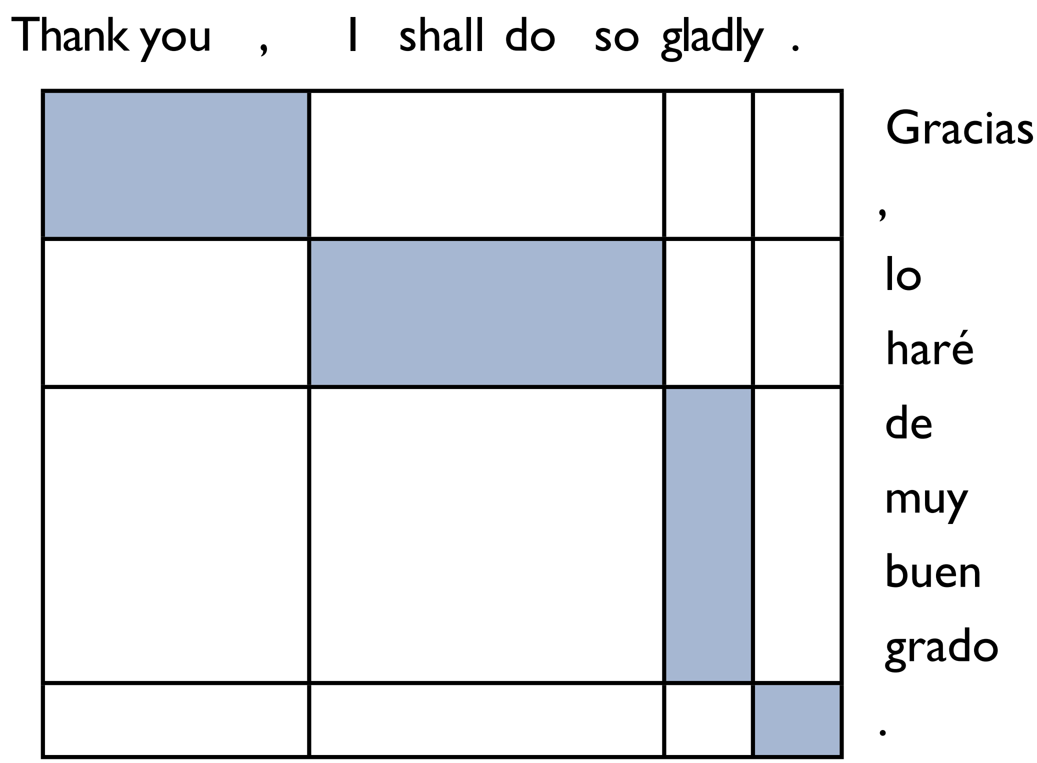


What Happens in Practice

A state-of-the-art word-level alignment



A sampled phrase alignment from our system



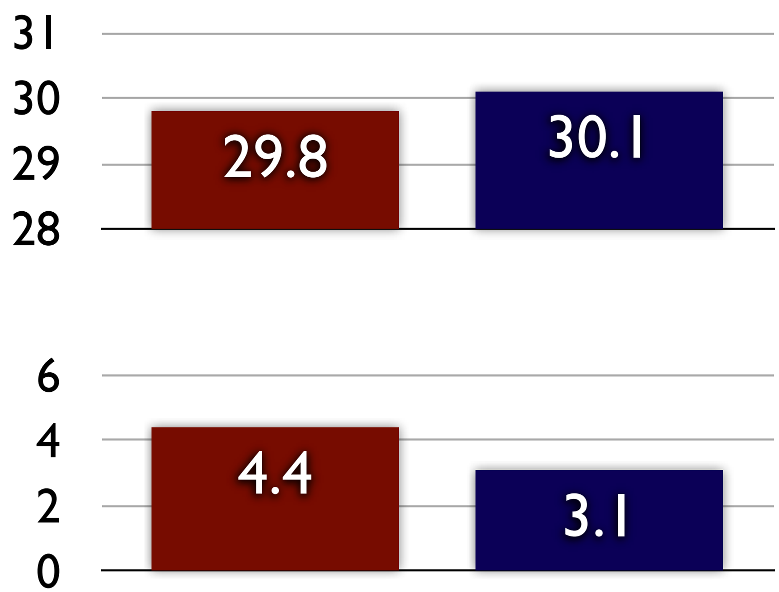
Performance Results

Translation performance in a phrase-based system (Moses) for Spanish-to-English parliamentary proceedings (Europarl)

- Word-level baseline
- Phrase-level model [DeNero et al. EMNLP '08]*

Translation quality (BLEU)

Millions of phrase pairs



Higher is better

Lower is cheaper

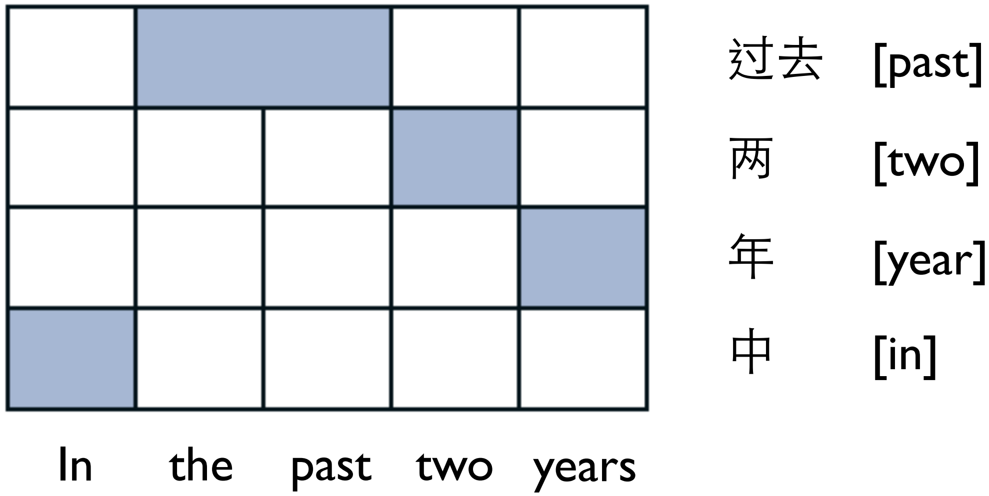
* John DeNero, Alex Bouchard-Côté, and Dan Klein. *Sampling Alignment Structure under a Bayesian Translation Model*, EMNLP 2008.

Subsequent Work

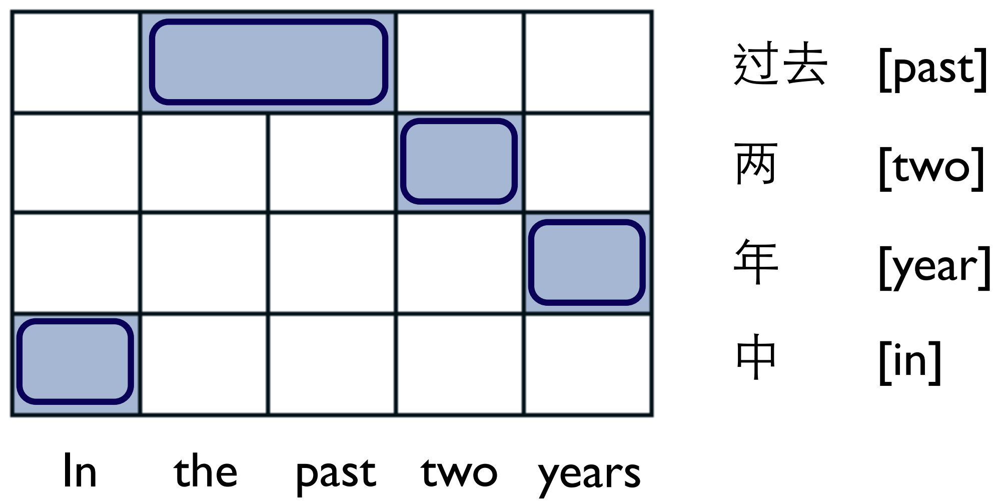
We described a non-parametric Bayesian prior and a consistent sampling procedure (EMNLP 2008)

- Phil Blunsom, Trevor Cohn, Chris Dyer, and Miles Osborne. *A Gibbs sampler for phrasal synchronous grammar induction*, ACL 2009.
- Matt Post and Daniel Gildea. *Bayesian Learning of a Tree Substitution Grammars*, ACL 2009.
- Trevor Cohn and Phil Blunsom. *A Bayesian Model of Syntax-Directed Tree to String Grammar Induction*, EMNLP 2009.
- Ding Liu and Daniel Gildea. *Bayesian Learning of Phrasal Tree-to-String Templates*, EMNLP 2009.
- Abhishek Arun, Chris Dyer, Barry Haddow, Phil Blunsom, Adam Lopez, and Philipp Koehn. *Monte Carlo inference and maximization for phrase-based translation*, CoNLL 2009.
- Phil Blunsom and Trevor Cohn. *Inducing Synchronous Grammars with Slice Sampling*, NAACL 2010.

A Model of Composed Phrases

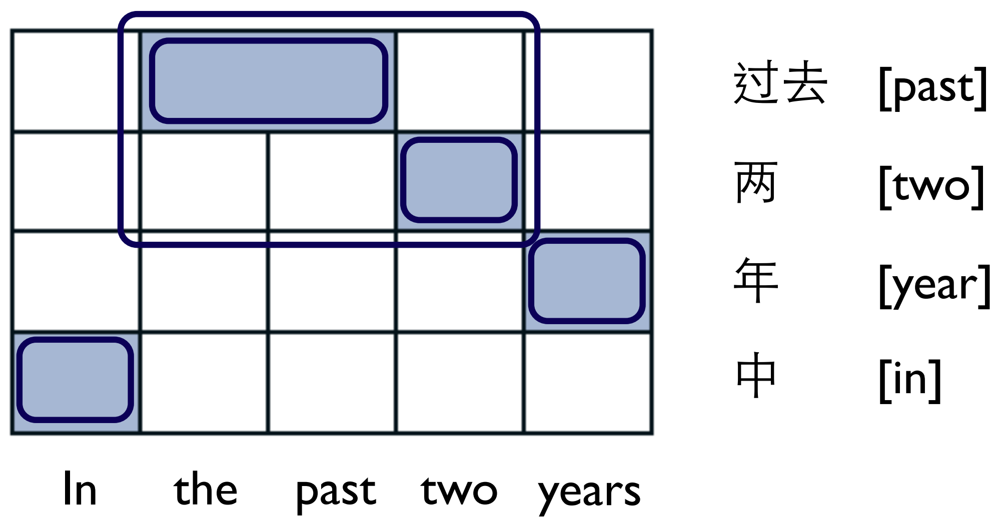


A Model of Composed Phrases



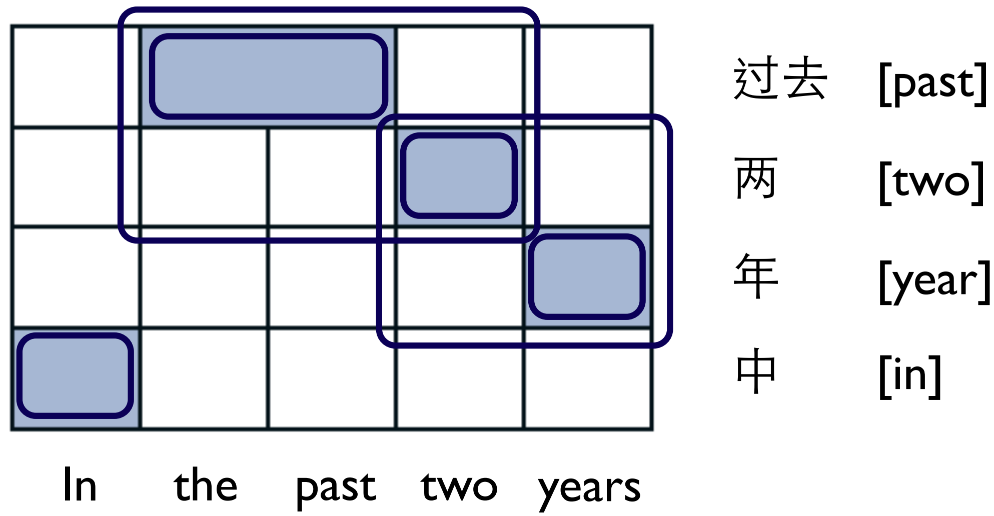
2

A Model of Composed Phrases

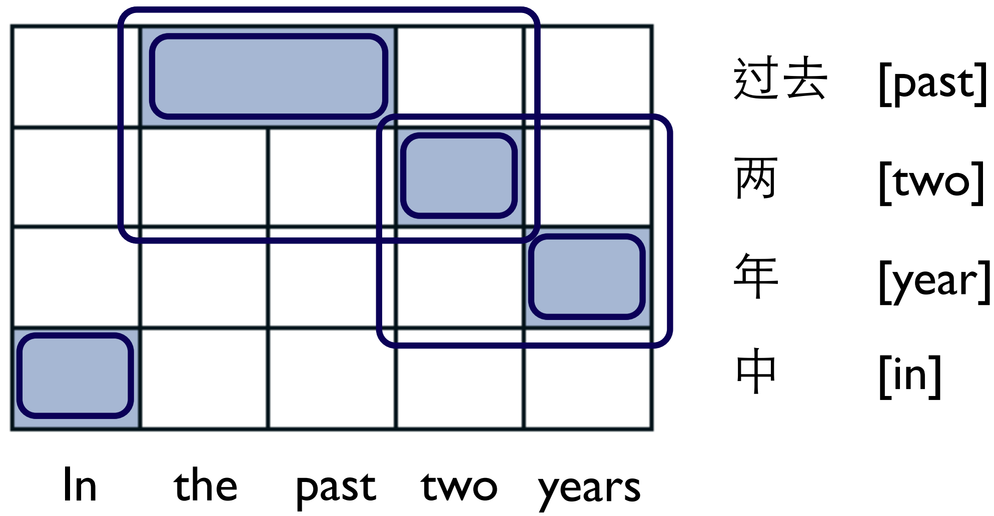


A Model of Composed Phrases

2

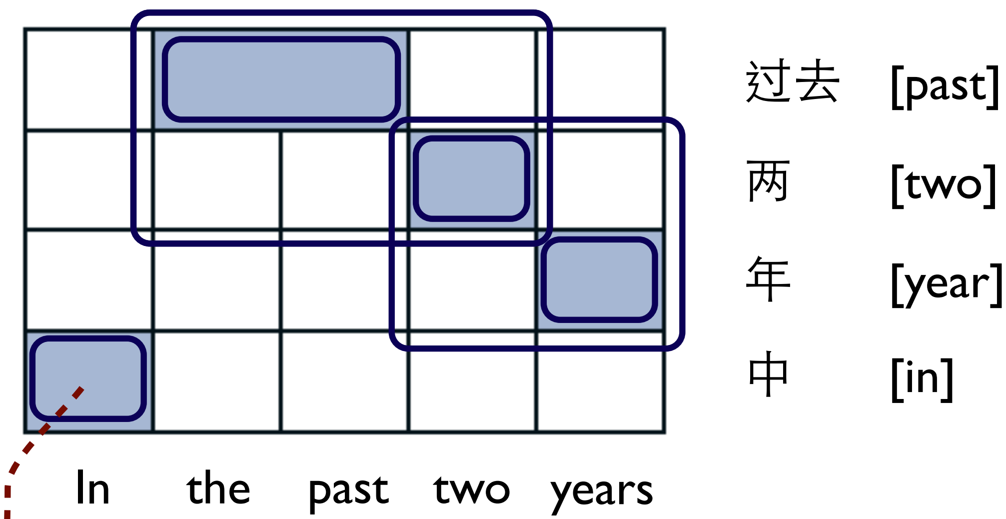


A Model of Composed Phrases



A model can predict the whole analysis above, including minimal links & composed phrase pairs .

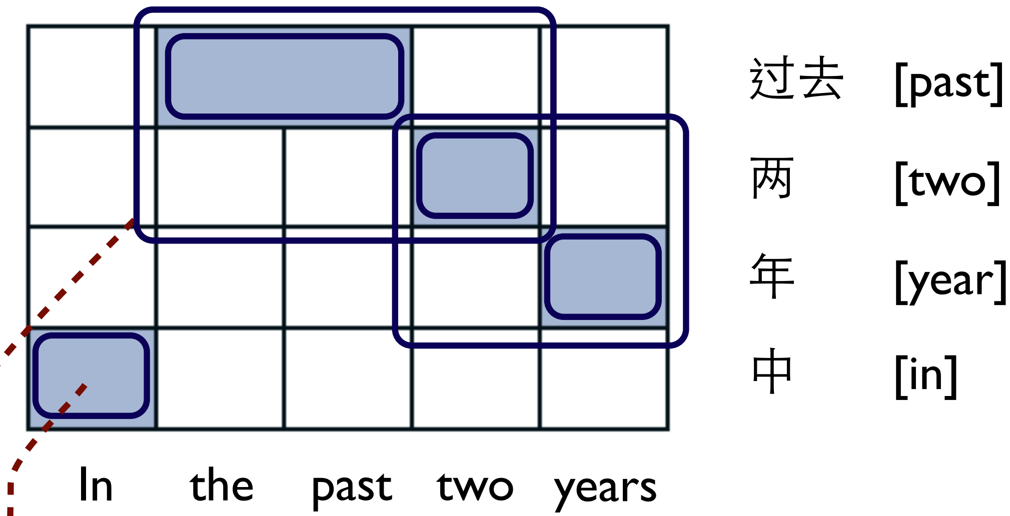
A Model of Composed Phrases



Features on word links: < P(in|中) = 0.8 , InDictionary = 1.0, ... >

A model can predict the whole analysis above, including minimal links & composed phrase pairs .

A Model of Composed Phrases



Features on

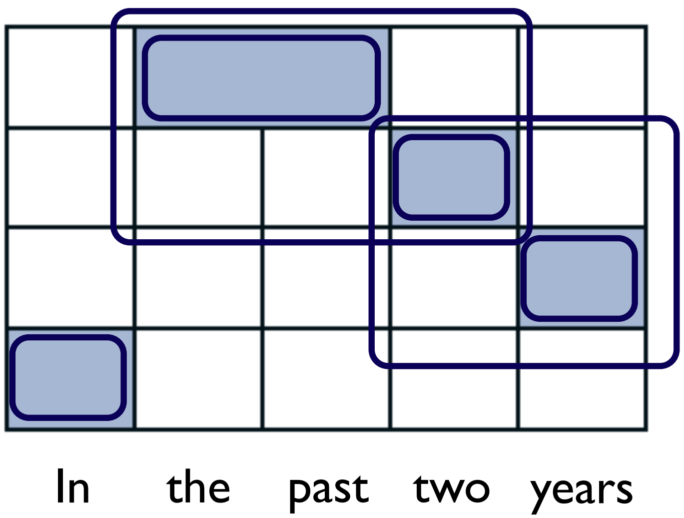
word links: < P(in|中) = 0.8 , InDictionary = 1.0, ... >

phrase pairs: < Count(*the past two*, 过去 两) = 7, Size(3,2) = 1, ... >

A model can predict the whole analysis above, including minimal links & composed phrase pairs .

Learning from Supervised Data

Guess: Model Prediction



过去 [past]

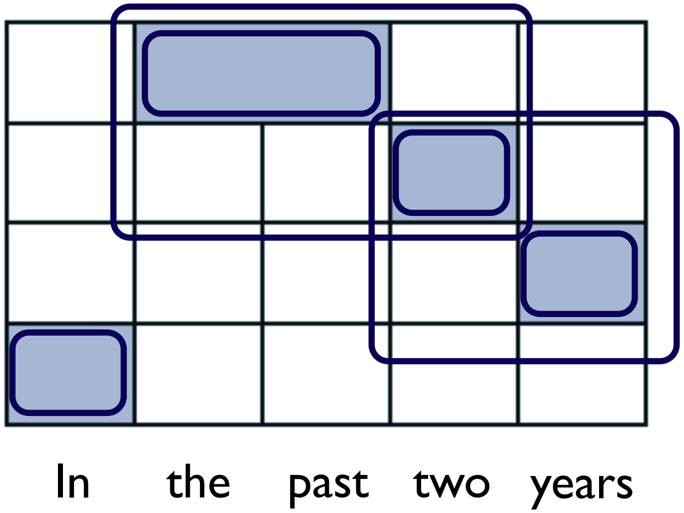
两 [two]

年 [year]

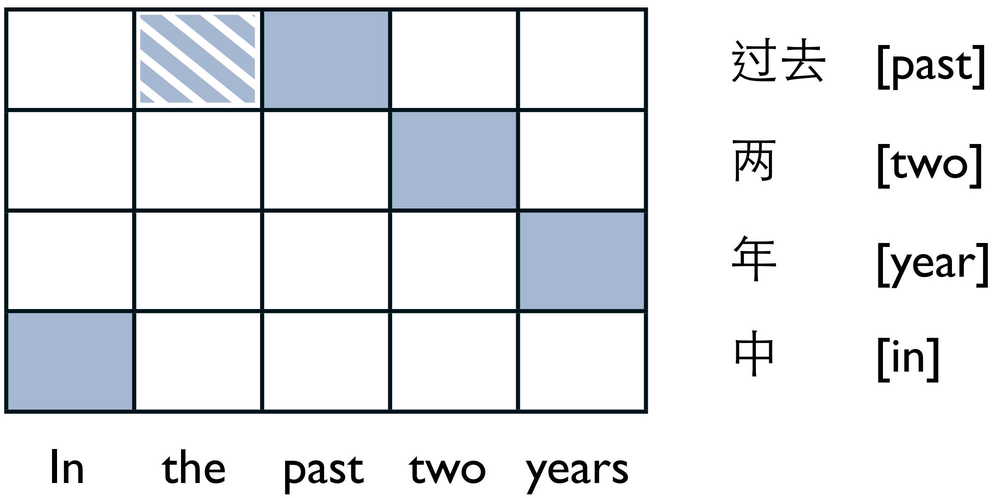
中 [in]

Learning from Supervised Data

Guess: Model Prediction

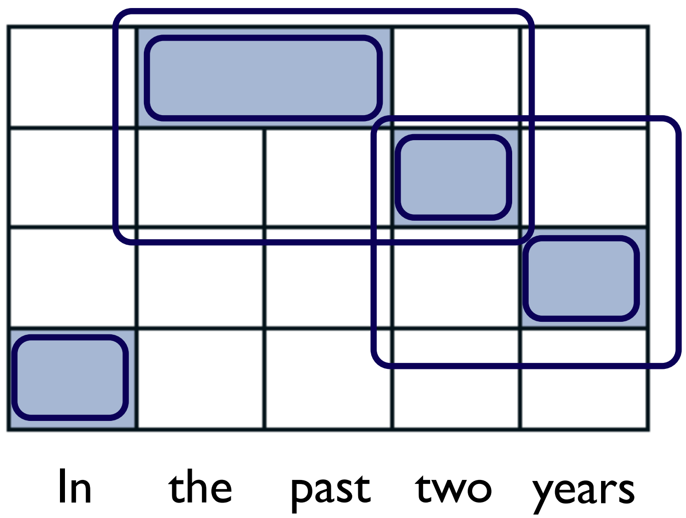


Gold: Human Annotation

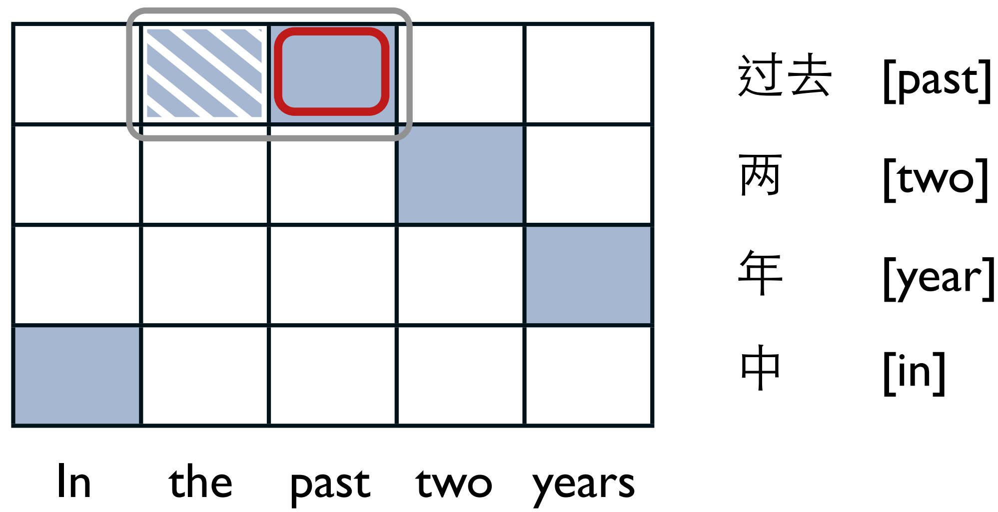


Learning from Supervised Data

Guess: Model Prediction



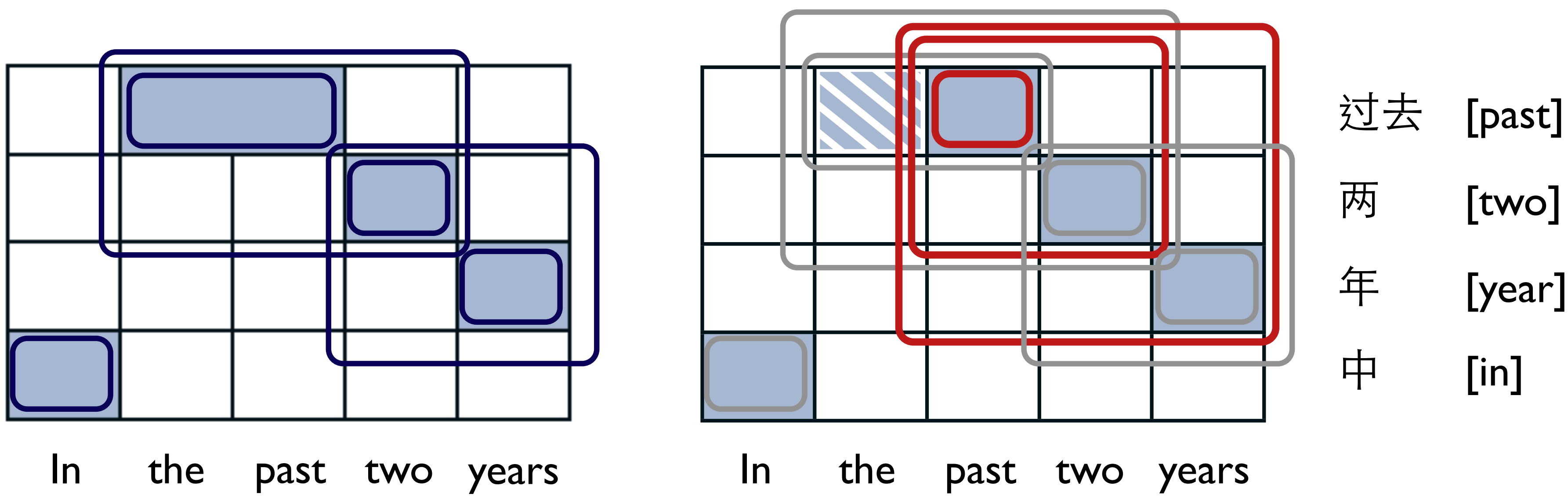
Gold: Human Annotation



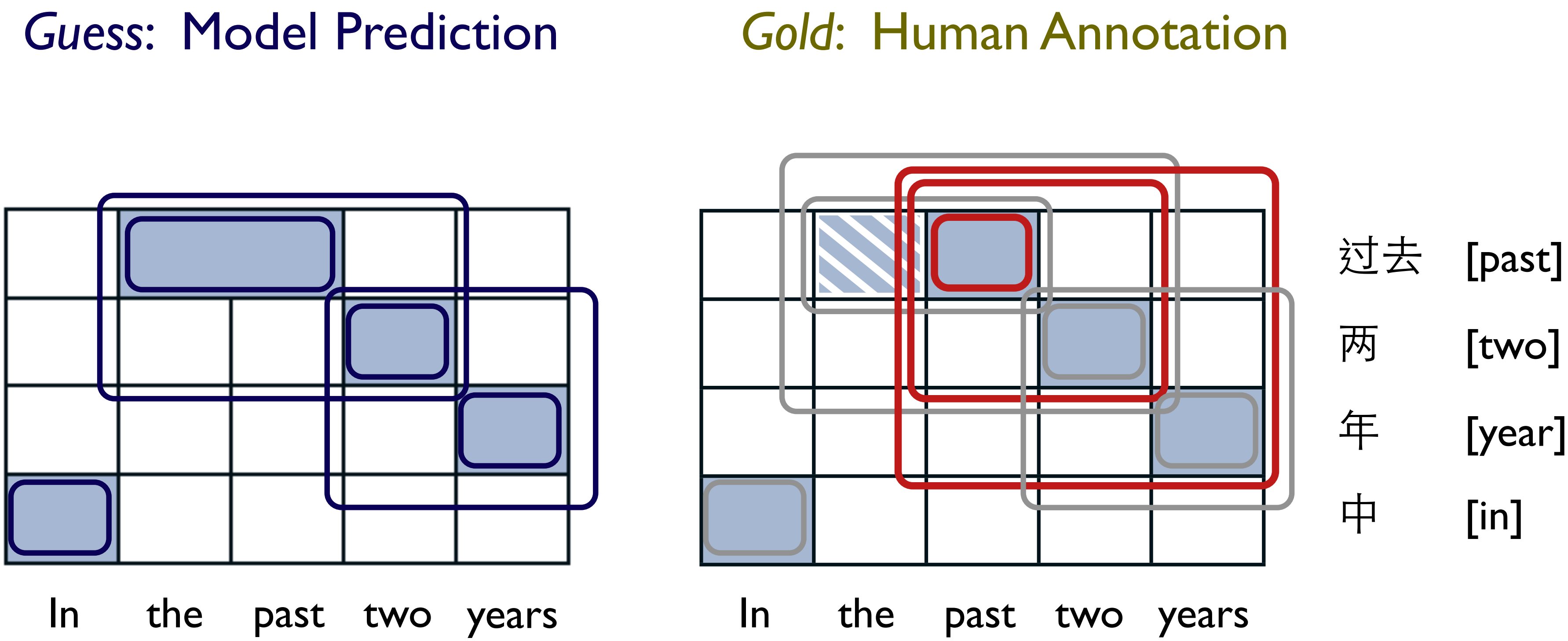
Learning from Supervised Data

Guess: Model Prediction

Gold: Human Annotation



Learning from Supervised Data

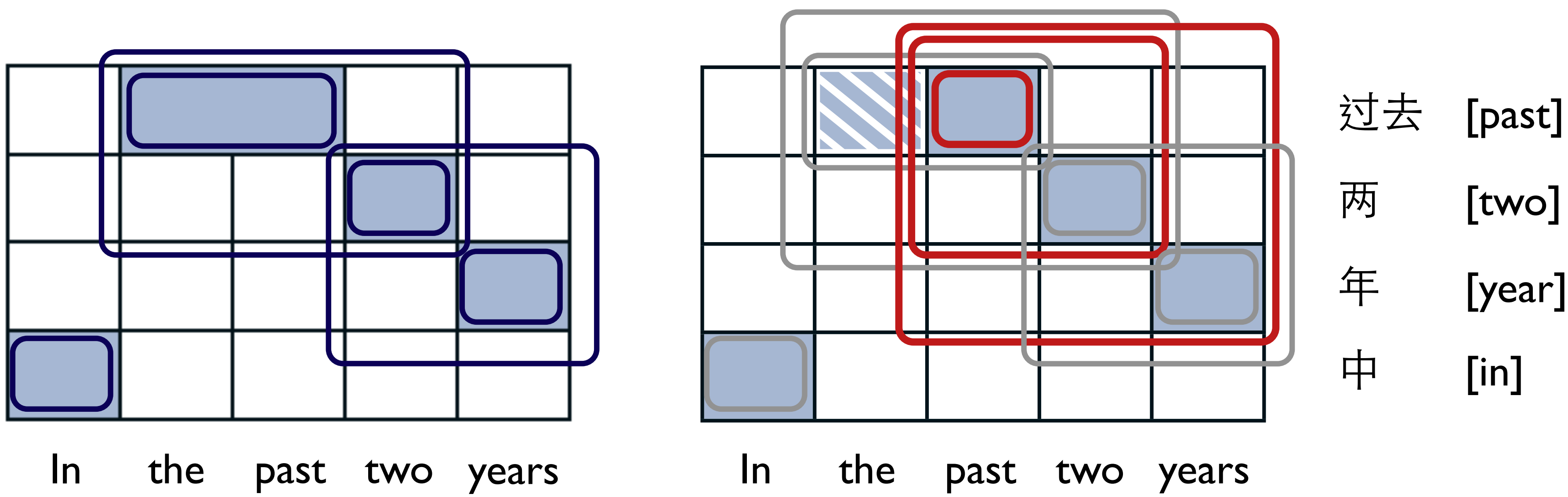


Loss function: Number of differing rounded rectangles

Learning from Supervised Data

Guess: Model Prediction

Gold: Human Annotation



Loss function: Number of differing rounded rectangles

Online learning (MIRA) adjusts model parameters to prefer the *gold* over the *guess* by a margin of the loss

Finding the Optimal Correspondence

$$\underset{y \in \mathrm{ITG}(x)}{\arg\max} \;\; \theta \cdot \left[\; \phi_{word}(x, y) \; + \; \phi_{phrase}(x, y) \;\right]$$

					过去 [past]
					两 [two]
					年 [year]
					中 [in]
In	the	past	two	years	

Finding the Optimal Correspondence

$$\arg\max_{y \in \mathrm{ITG}(x)} \; \theta \cdot \left[\; \phi_{word}(x,y) \; + \; \phi_{phrase}(x,y) \;\right]$$

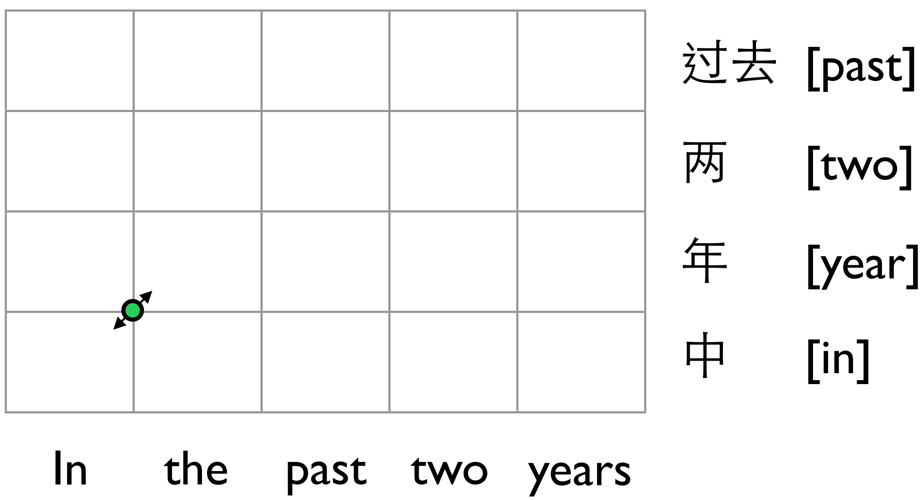
Hierarchical decomposition

	In	the	past	two	years
过去 [past]					
两 [two]					
年 [year]					
中 [in]					

Finding the Optimal Correspondence

$$\arg\max_{y \in \mathrm{ITG}(x)} \; \theta \cdot \left[\, \phi_{word}(x, y) \;+\; \phi_{phrase}(x, y) \,\right]$$

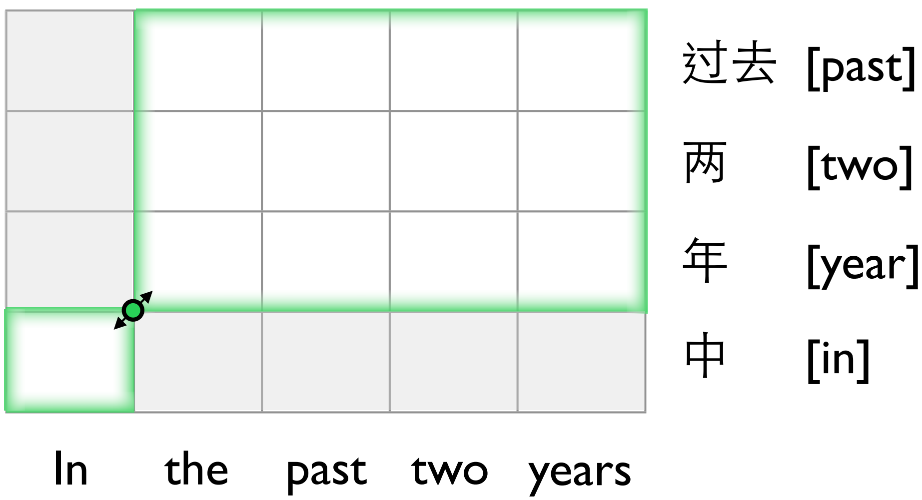
Hierarchical decomposition



Finding the Optimal Correspondence

$$\arg\max_{y \in \mathrm{ITG}(x)} \; \theta \cdot \left[\, \phi_{word}(x, y) \; + \; \phi_{phrase}(x, y) \,\right]$$

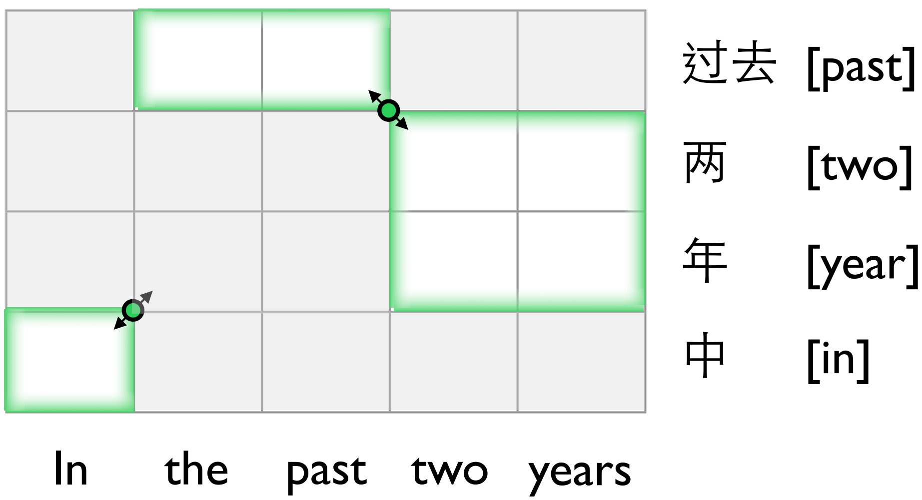
Hierarchical decomposition



Finding the Optimal Correspondence

$$\arg\max_{y \in \mathrm{ITG}(x)} \; \theta \cdot [\; \phi_{word}(x, y) \; + \; \phi_{phrase}(x, y) \;]$$

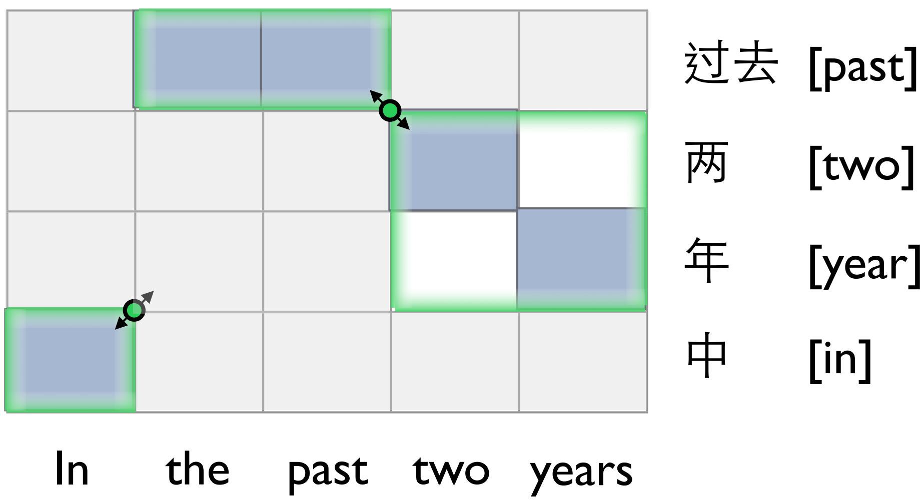
Hierarchical decomposition



Finding the Optimal Correspondence

$$\arg\max_{y \in \mathrm{ITG}(x)} \ \theta \cdot \left[\ \phi_{word}(x,y) \ + \ \phi_{phrase}(x,y)\ \right]$$

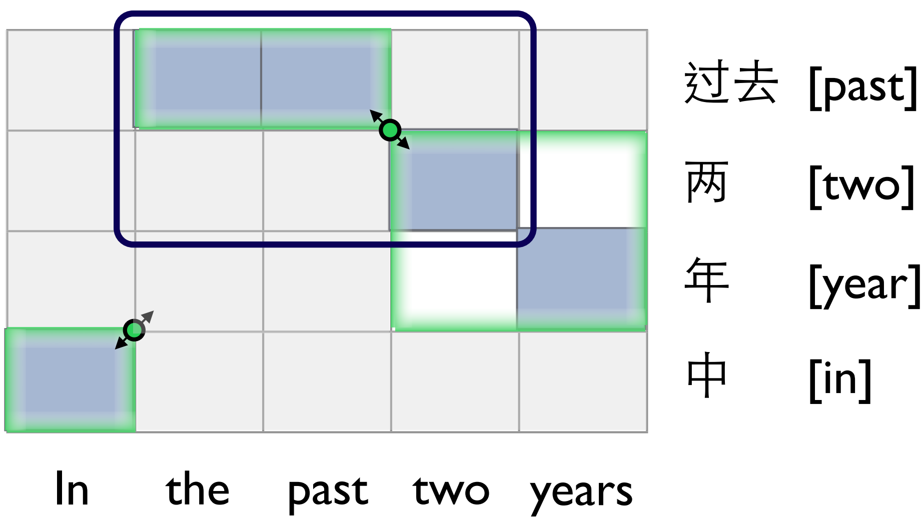
Hierarchical decomposition



Finding the Optimal Correspondence

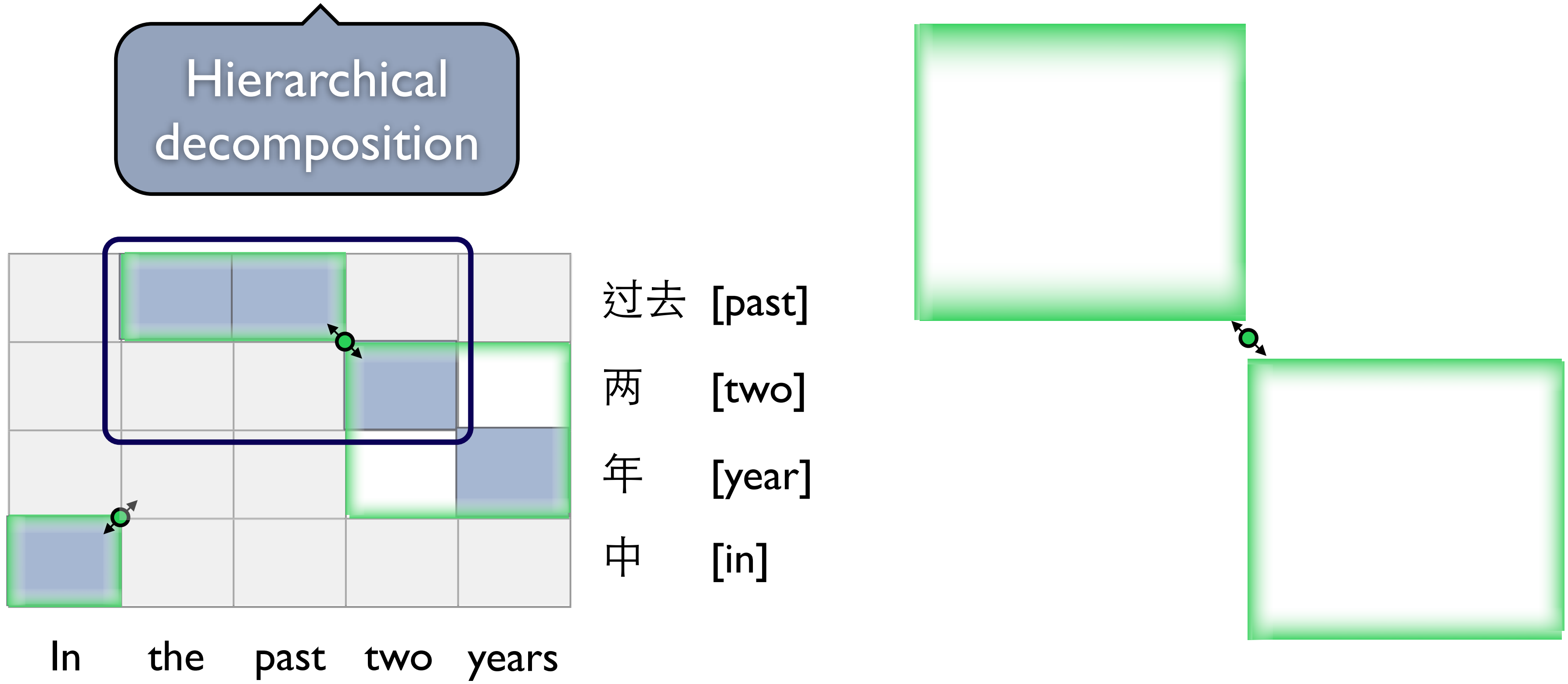
$$\arg\max_{y \in \mathrm{ITG}(x)} \theta \cdot [\ \phi_{word}(x,y) + \phi_{phrase}(x,y)\ ]$$

Hierarchical decomposition



Finding the Optimal Correspondence

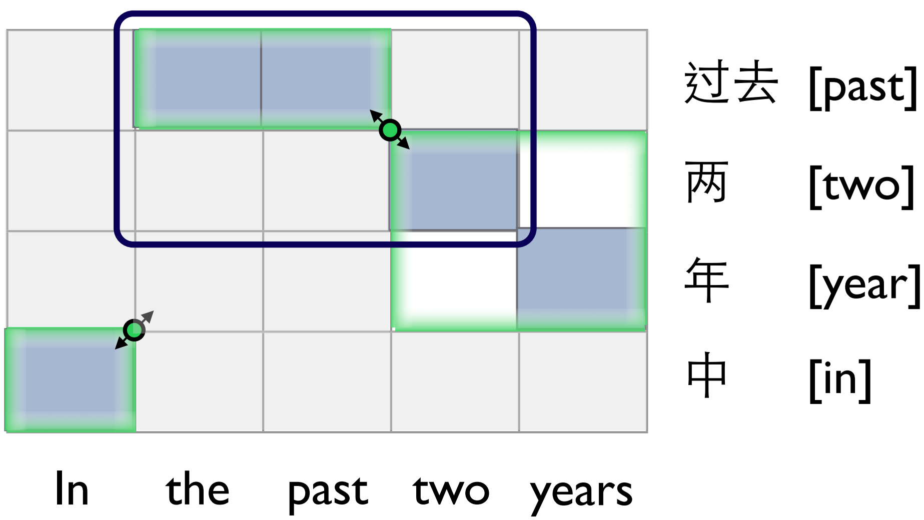
$$\arg\max_{y \in \mathrm{ITG}(x)} \; \theta \cdot \left[\, \phi_{word}(x,y) \;+\; \phi_{phrase}(x,y) \,\right]$$

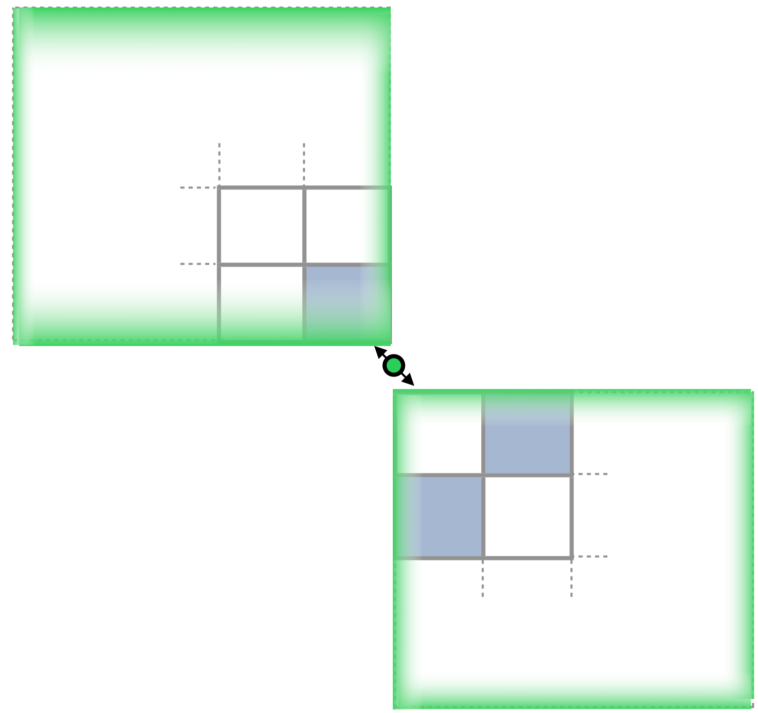


Finding the Optimal Correspondence

$$\arg\max_{y \in \mathrm{ITG}(x)} \; \theta \cdot [\; \phi_{word}(x,y) \; + \; \phi_{phrase}(x,y) \;]$$

Hierarchical decomposition

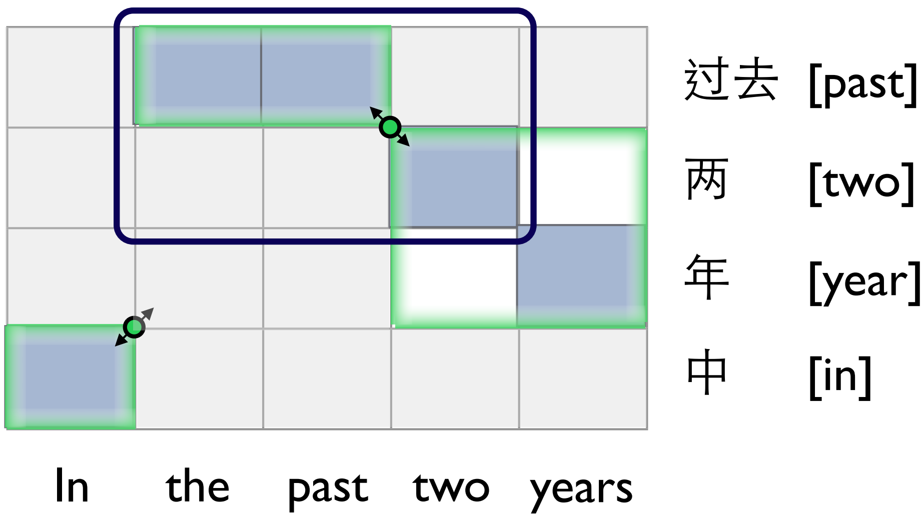


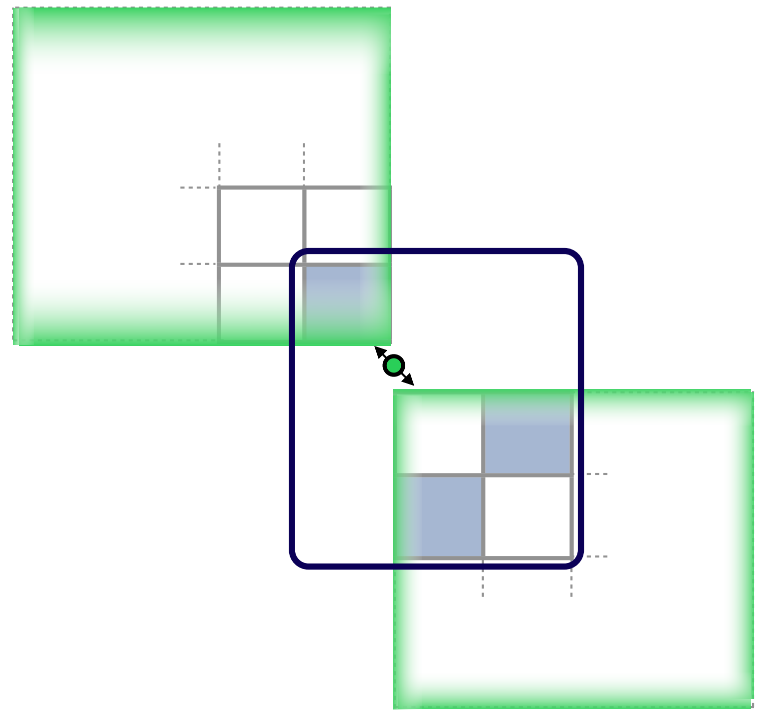


Finding the Optimal Correspondence

$$\arg\max_{y \in \mathrm{ITG}(x)} \; \theta \cdot [\; \phi_{word}(x, y) \; + \; \phi_{phrase}(x, y) \;]$$

Hierarchical decomposition

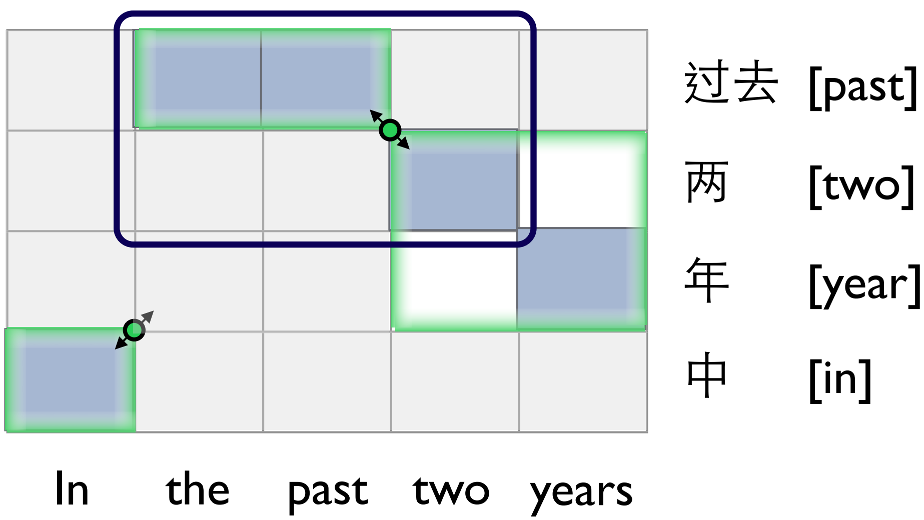


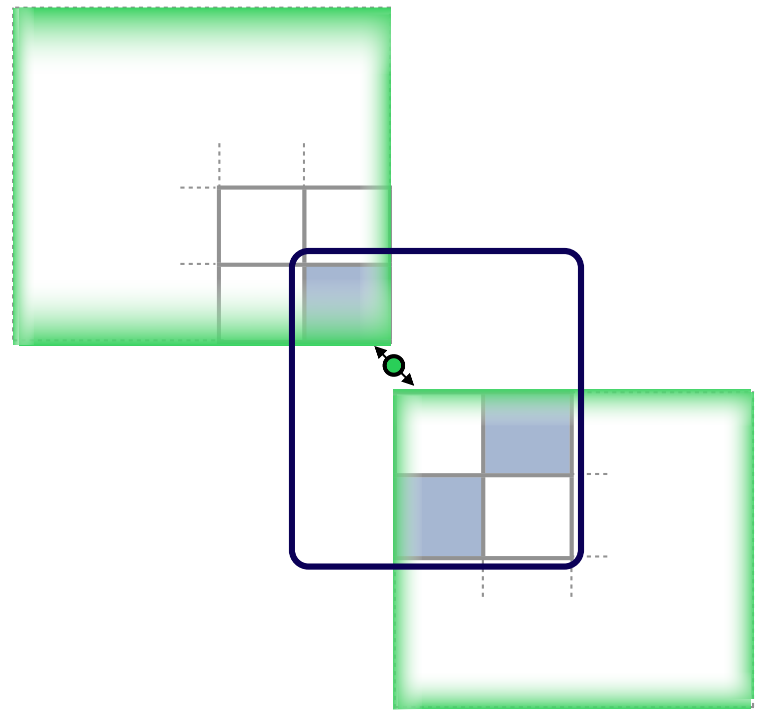


Finding the Optimal Correspondence

$$\arg\max_{y \in \mathrm{ITG}(x)} \ \theta \cdot [\ \phi_{word}(x,y) \ + \ \phi_{phrase}(x,y)\ ]$$

Hierarchical decomposition





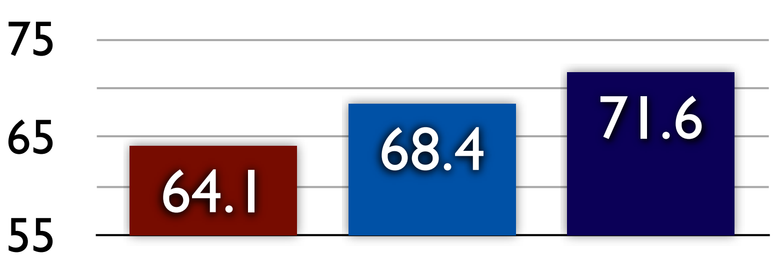
ITG parser with a state space that tracks peripheral alignments for each region

Experimental Results

- Unsupervised word model baseline
- Supervised word model [Haghighi, Blitzer, DeNero, and Klein. ACL '09]*
- Composed Phrase Pair Model [DeNero and Klein. In submission]**

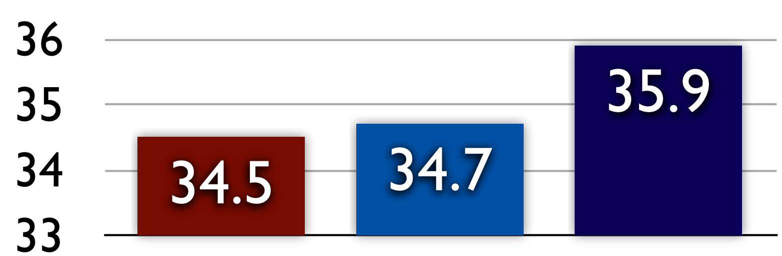
Alignment quality relative to human-annotated data

Phrase Pair F1



Translation quality for Chinese-to-English

BLEU



* Aria Haghighi, John Blitzer, John DeNero, and Dan Klein. *Better Word Alignments with Supervised ITG Models*, ACL 2009.

** John DeNero and Dan Klein. *Supervised Modeling of Extraction Sets for Machine Translation*, in submission.

The Steps in a Modern Translation System

Learn a model

Apply the model

Choose a translation

The Steps in a Modern Translation System

Learn a model	Apply the model	Choose a translation

- Large data sets provide statistics for larger structures

The Steps in a Modern Translation System

Learn a model	Apply the model	Choose a translation

- Large data sets provide statistics for larger structures
- Non-parametric models scale with the data

The Steps in a Modern Translation System

Learn a model | Apply the model | Choose a translation

- Large data sets provide statistics for larger structures
- Non-parametric models scale with the data
- The more context we incorporate, the better we do

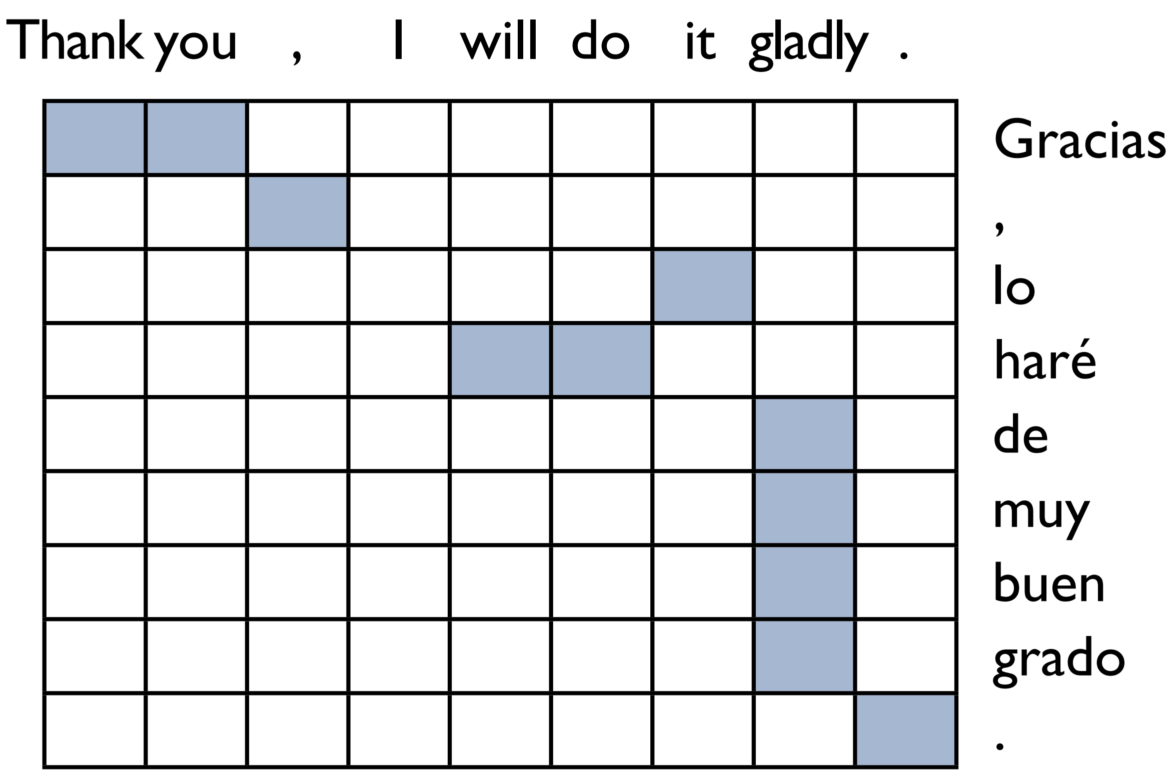
The Steps in a Modern Translation System

Learn a model

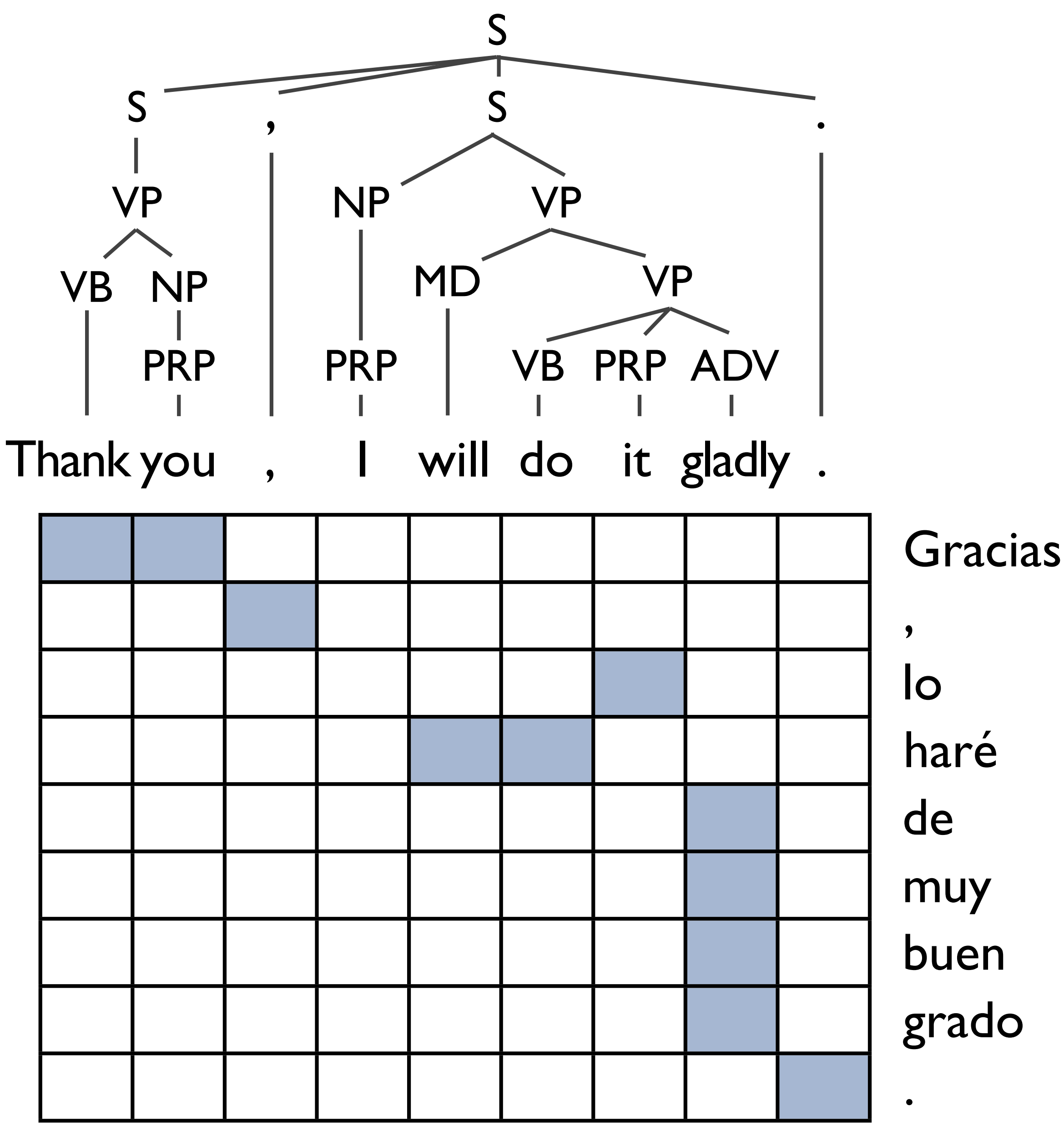
Apply the model

Choose a translation

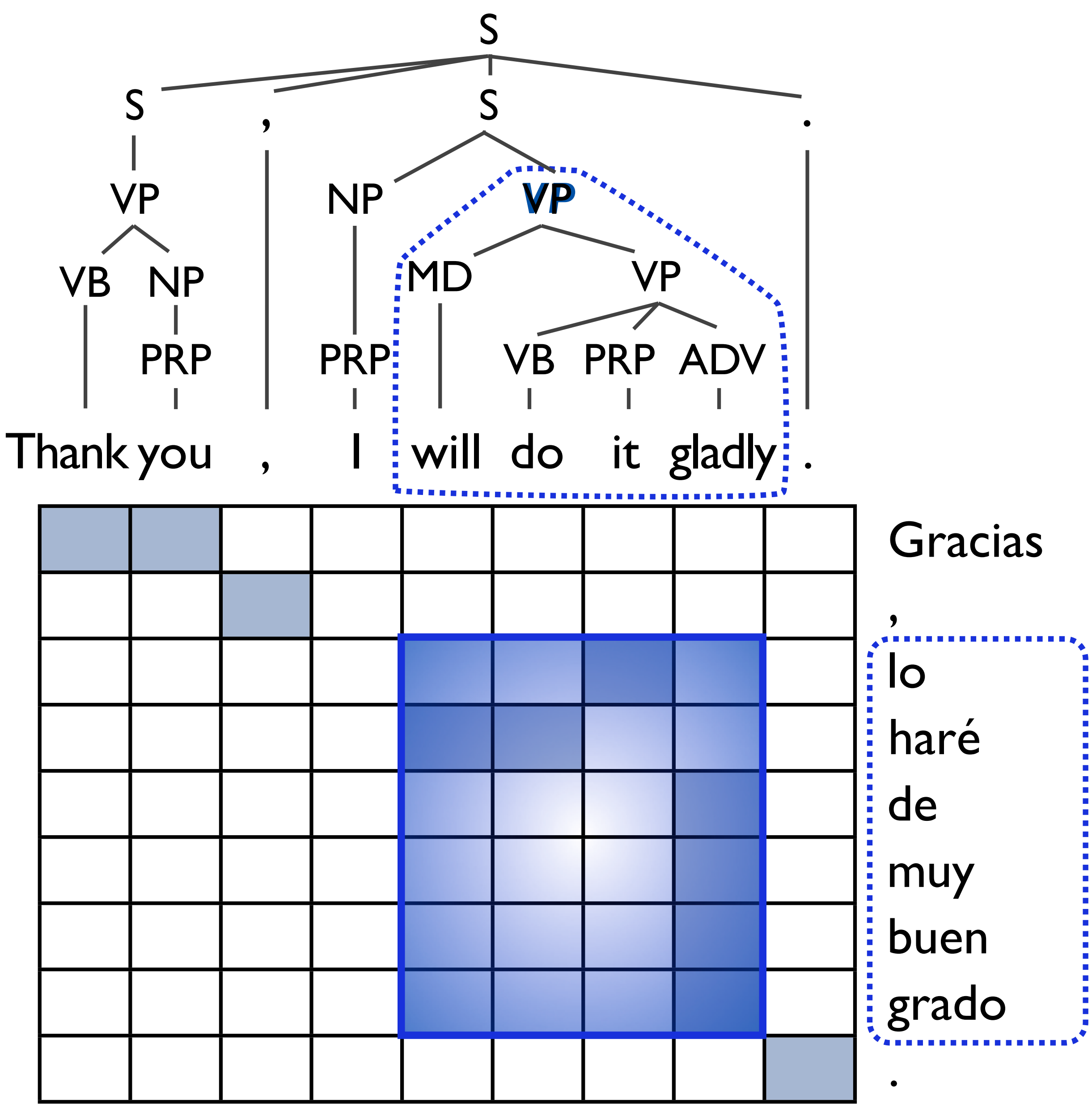
Extracting Translation Rules



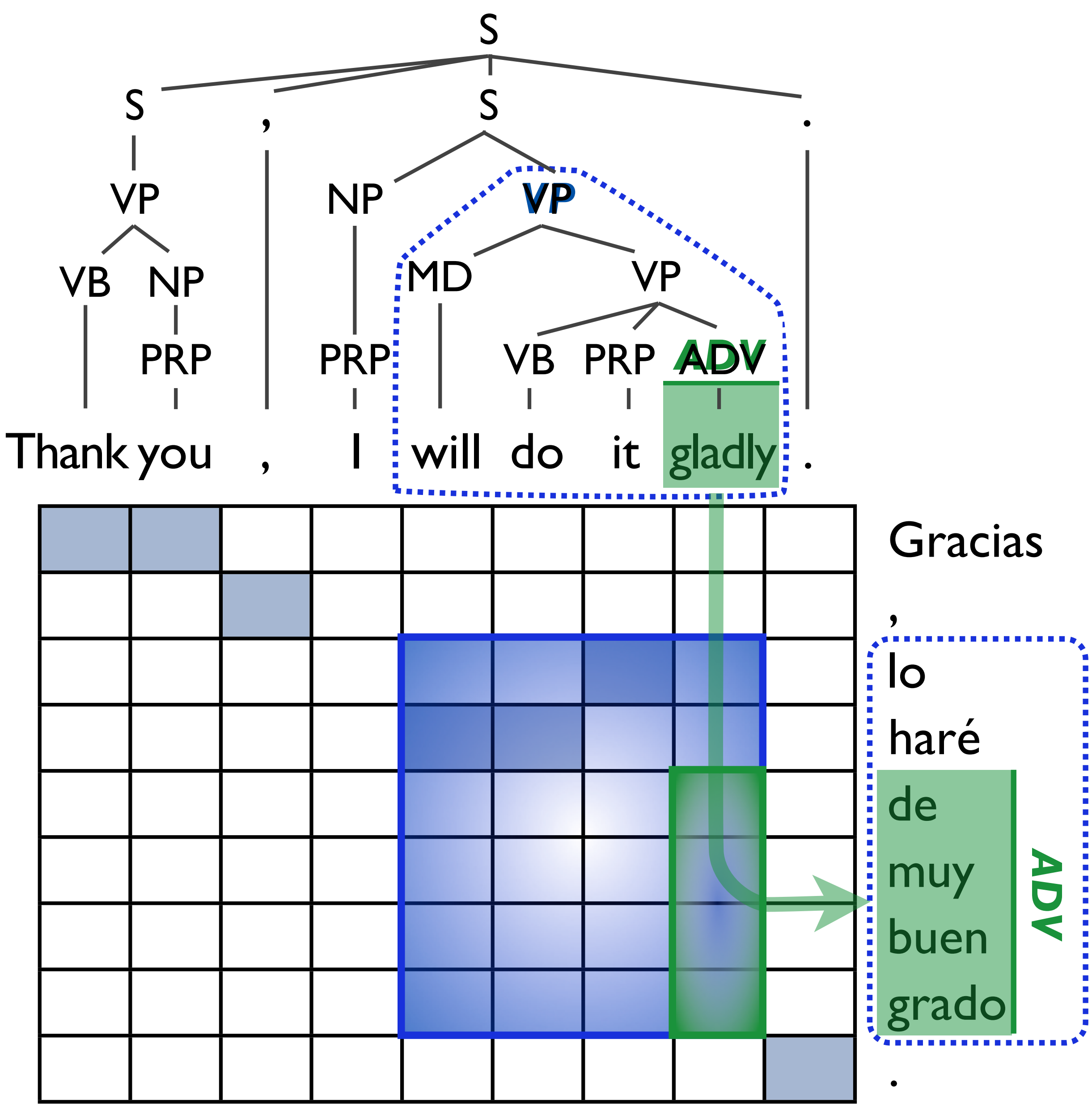
Extracting Translation Rules



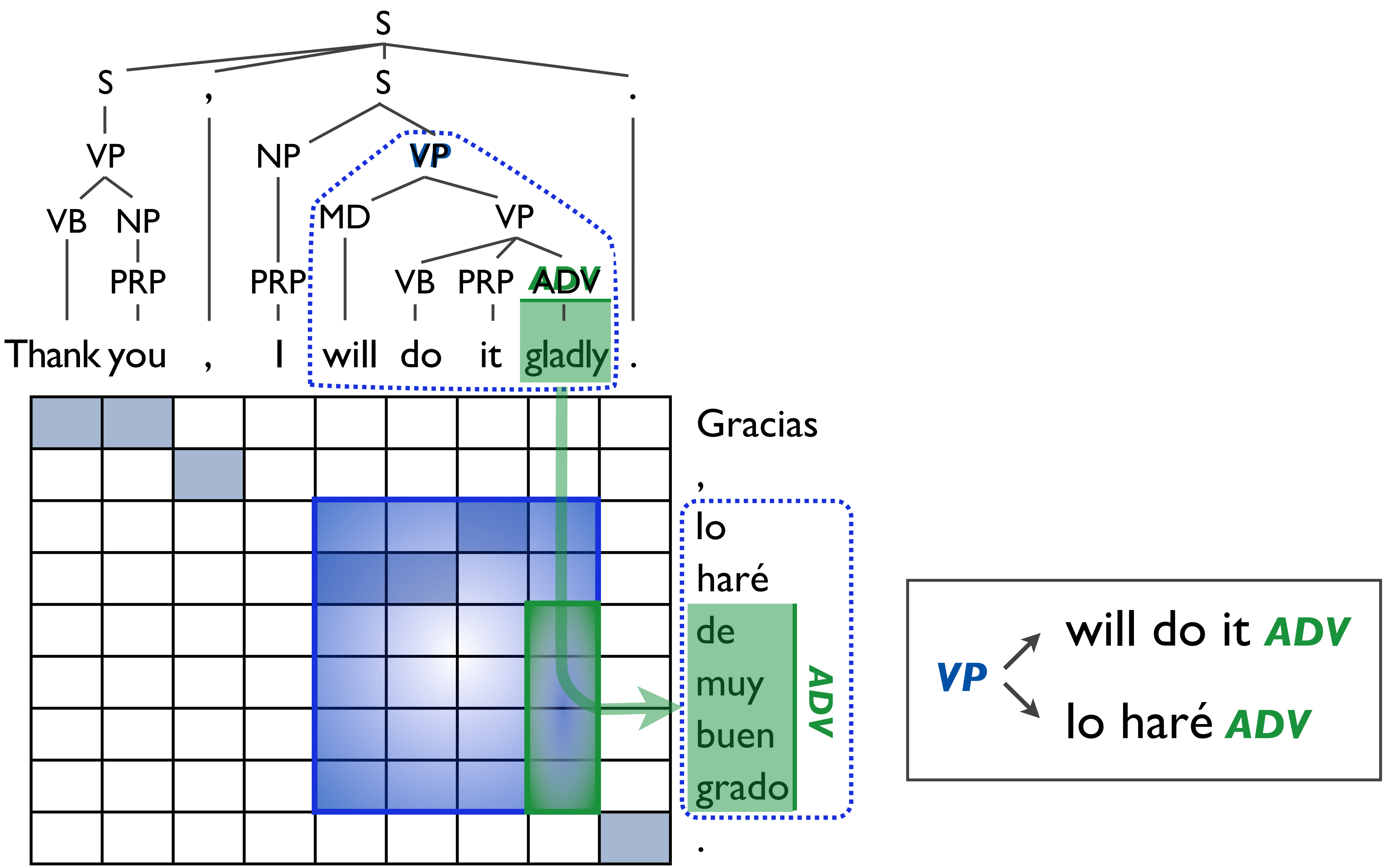
Extracting Translation Rules



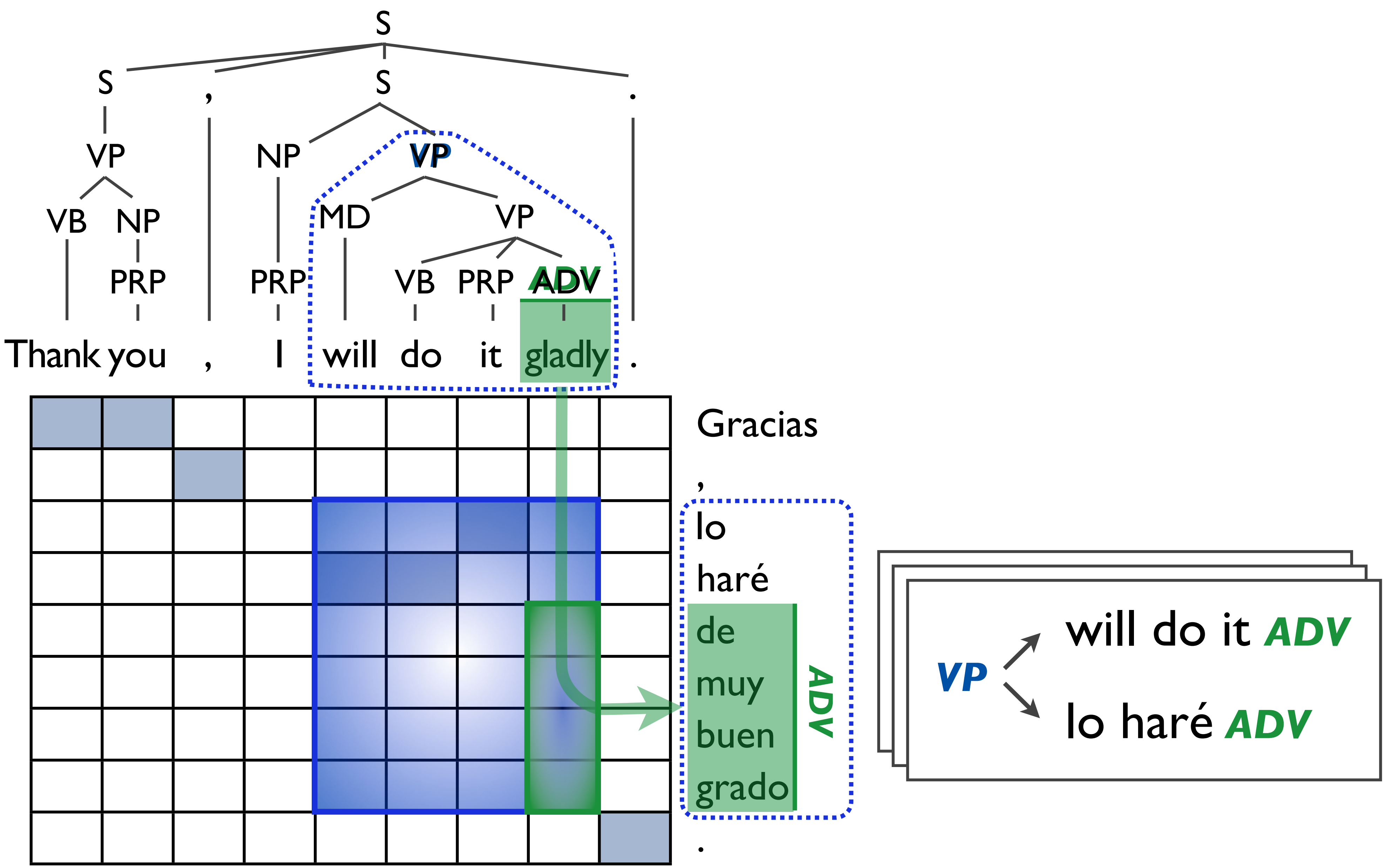
Extracting Translation Rules



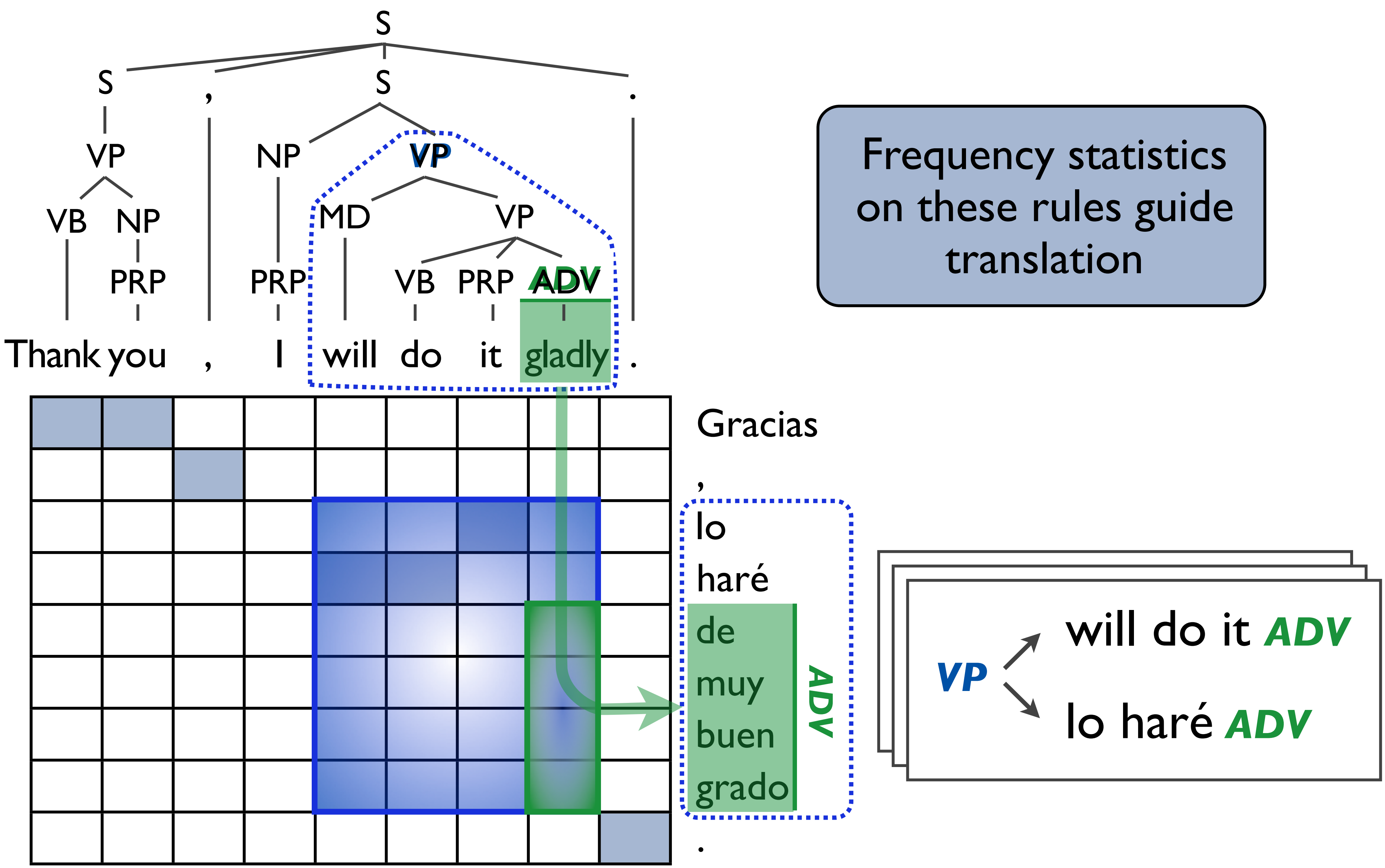
Extracting Translation Rules



Extracting Translation Rules



Extracting Translation Rules



Frequency statistics on these rules guide translation

Synchronous Context-Free Grammars

Grammar

Derivation

Translation:

Source: Mi dormitorio nuevo no es ni grande ni pequeño

Synchronous Context-Free Grammars

NN → *bedroom*
NN → dormitorio

Grammar

Derivation

Translation:

Source: Mi dormitorio nuevo no es ni grande ni pequeño

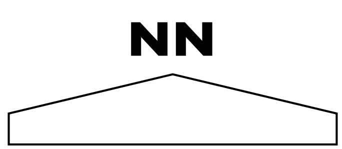
Synchronous Context-Free Grammars

NN ↗ *bedroom*
↘ dormitorio

Grammar

Derivation

Translation:



Source: Mi dormitorio nuevo no es ni grande ni pequeño

Synchronous Context-Free Grammars

NN → *bedroom*
NN → dormitorio

Grammar

Derivation

NN

Translation: *bedroom*

NN

Source: Mi dormitorio nuevo no es ni grande ni pequeño

Synchronous Context-Free Grammars

NN ↗ *bedroom*
NN ↘ dormitorio

JJ ↗ *new*
JJ ↘ nuevo

JJ ↗ *big*
JJ ↘ grande

JJ ↗ *small*
JJ ↘ pequeño

Grammar

Derivation

NN

Translation: *bedroom*

NN

Source: Mi dormitorio nuevo no es ni grande ni pequeño

Synchronous Context-Free Grammars

JJ → *new* / nuevo

NN → *bedroom* / dormitorio

JJ → *big* / grande

JJ → *small* / pequeño

Grammar

Derivation

Translation: JJ *new* NN *bedroom* JJ *big* JJ *small*

NN JJ JJ JJ

Source: Mi dormitorio nuevo no es ni grande ni pequeño

Synchronous Context-Free Grammars

NN → *bedroom* / dormitorio

JJ → *new* / nuevo

JJ → *big* / grande

JJ → *small* / pequeño

NP → *My* **JJ NN** / Mi **NN JJ**

Grammar

Derivation

Translation: **JJ** *new* **NN** *bedroom* **JJ** *big* **JJ** *small*

NN **JJ** **JJ** **JJ**

Source: Mi dormitorio nuevo no es ni grande ni pequeño

Synchronous Context-Free Grammars

NN → *bedroom* / dormitorio

JJ → *new* / nuevo

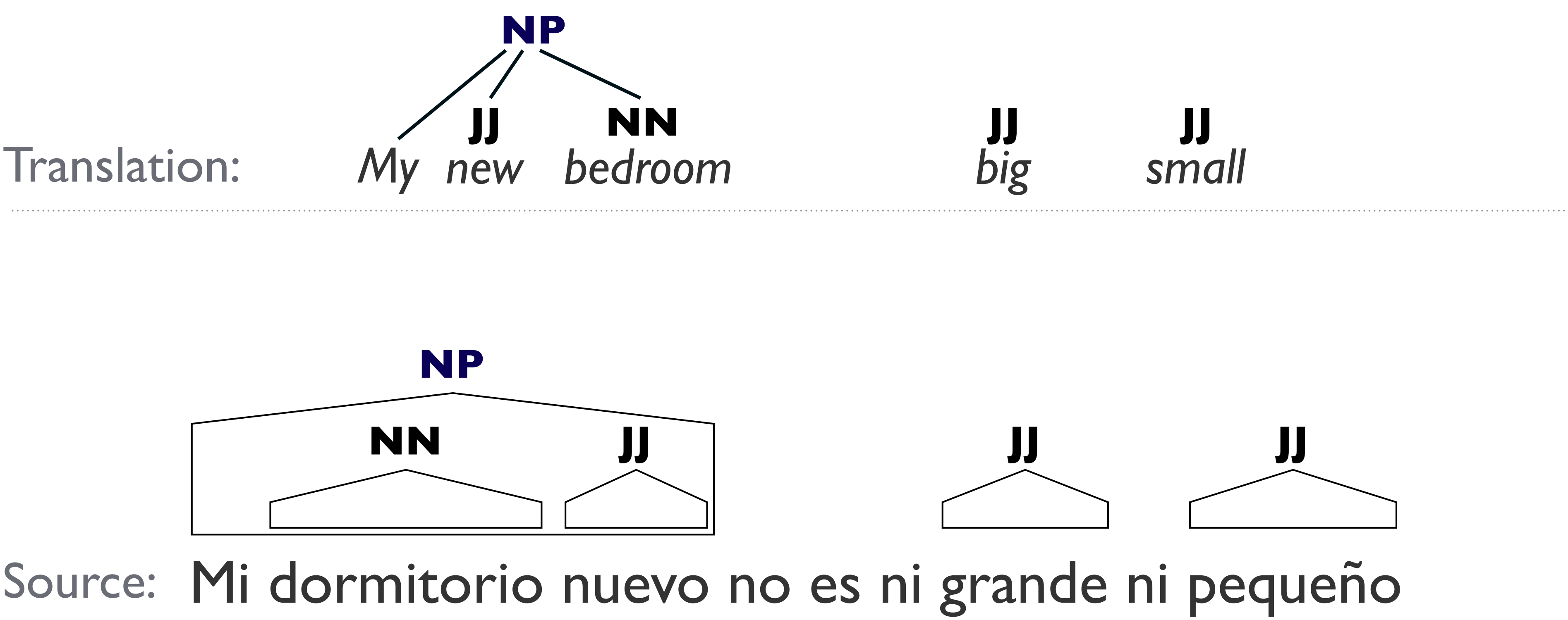
JJ → *big* / grande

JJ → *small* / pequeño

NP → *My* **JJ NN** / Mi **NN JJ**

Grammar

Derivation



Synchronous Context-Free Grammars

NN → *bedroom* / dormitorio

JJ → *new* / nuevo

JJ → *big* / grande

JJ → *small* / pequeño

NP → *My* **JJ NN** / Mi **NN JJ**

S → **NP** *is neither* **JJ** *nor* **JJ** / **NP** no es ni **JJ** ni **JJ**

Grammar

Derivation

NP

JJ **NN** **JJ** **JJ**

Translation: *My new bedroom* *big* *small*

NP

NN **JJ** **JJ** **JJ**

Source: Mi dormitorio nuevo no es ni grande ni pequeño

Synchronous Context-Free Grammars

NN → *bedroom* / dormitorio

JJ → *new* / nuevo

JJ → *big* / grande

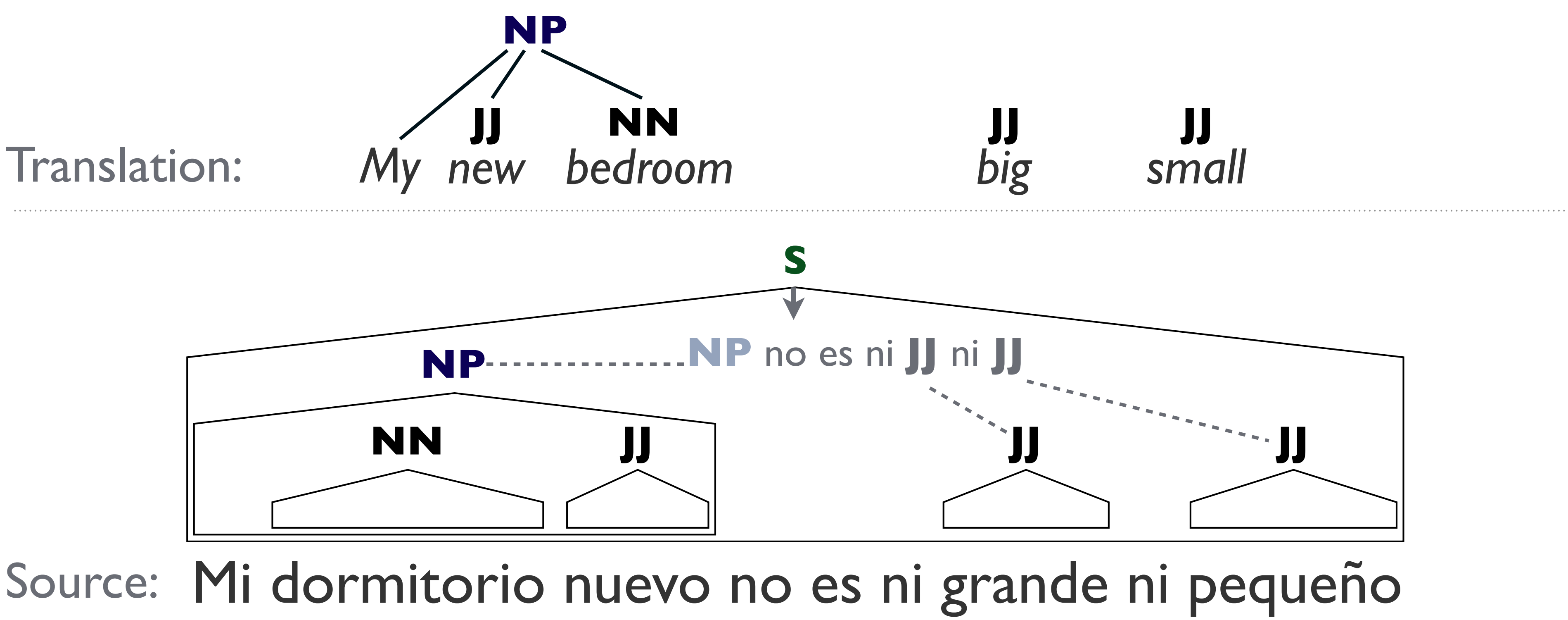
JJ → *small* / pequeño

NP → *My* **JJ NN** / Mi **NN JJ**

S → **NP** *is neither* **JJ** *nor* **JJ** / **NP** no es ni **JJ** ni **JJ**

Grammar

Derivation



Synchronous Context-Free Grammars

NN → *bedroom* / dormitorio

JJ → *new* / nuevo

JJ → *big* / grande

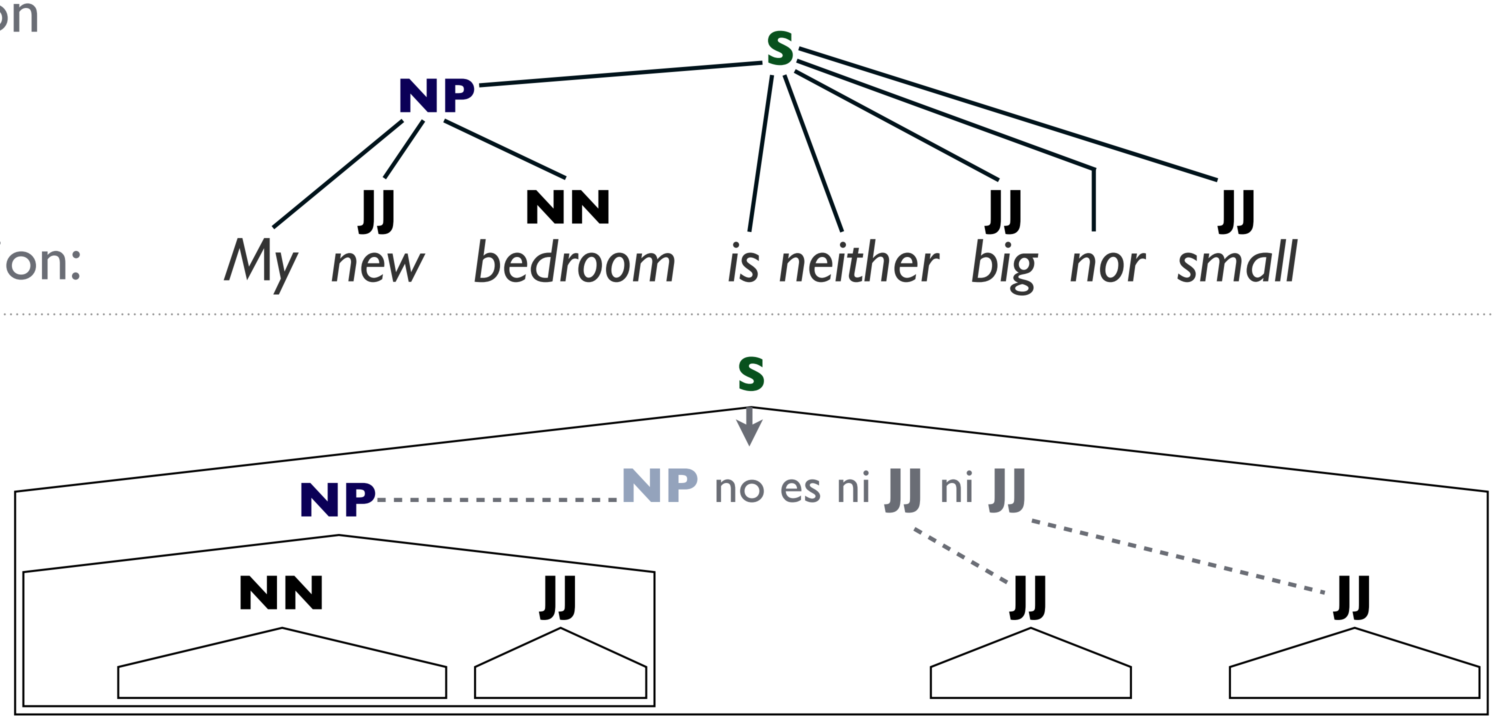
JJ → *small* / pequeño

NP → *My* **JJ NN** / Mi **NN JJ**

S → **NP** *is neither* **JJ** *nor* **JJ** / **NP** no es ni **JJ** ni **JJ**

Grammar

Derivation



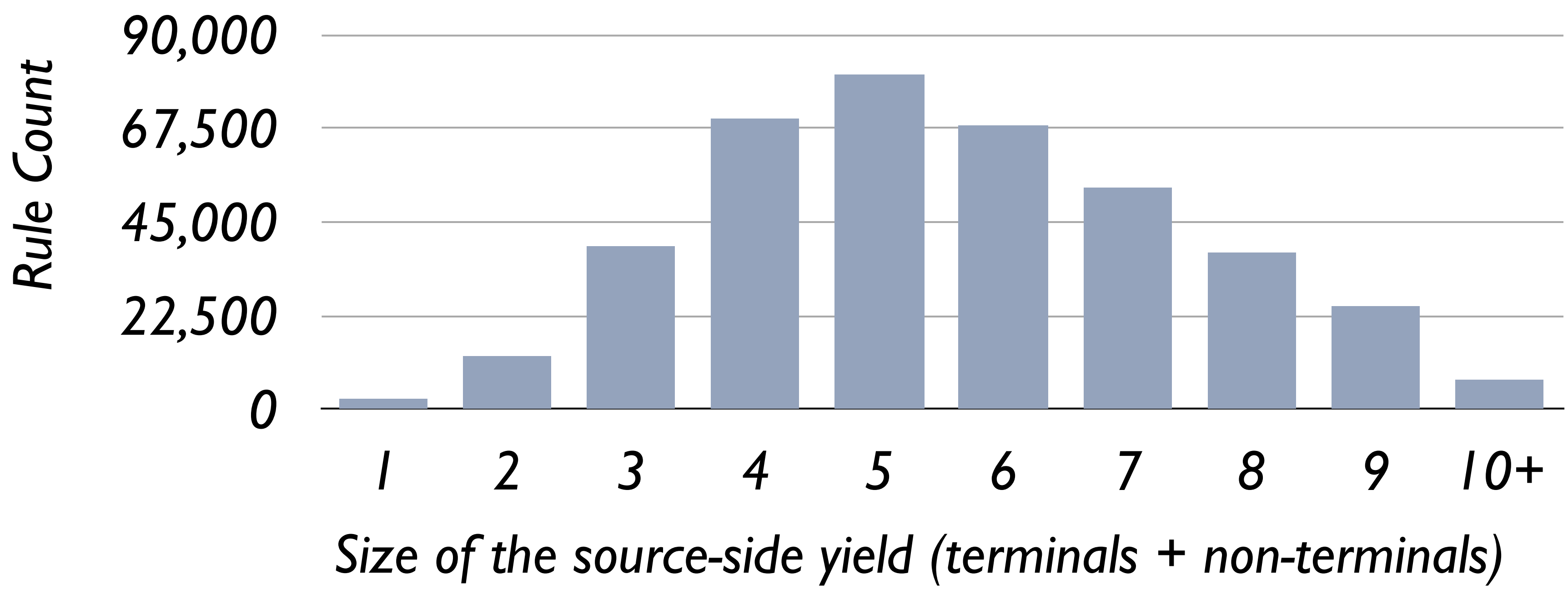
Translation: *My new bedroom is neither big nor small*

Source: Mi dormitorio nuevo no es ni grande ni pequeño

The Size of the Grammar

A grammar learned from 220 million words of
Arabic-to-English example translations:

332,000 rules match a 30-word sentence to be translated

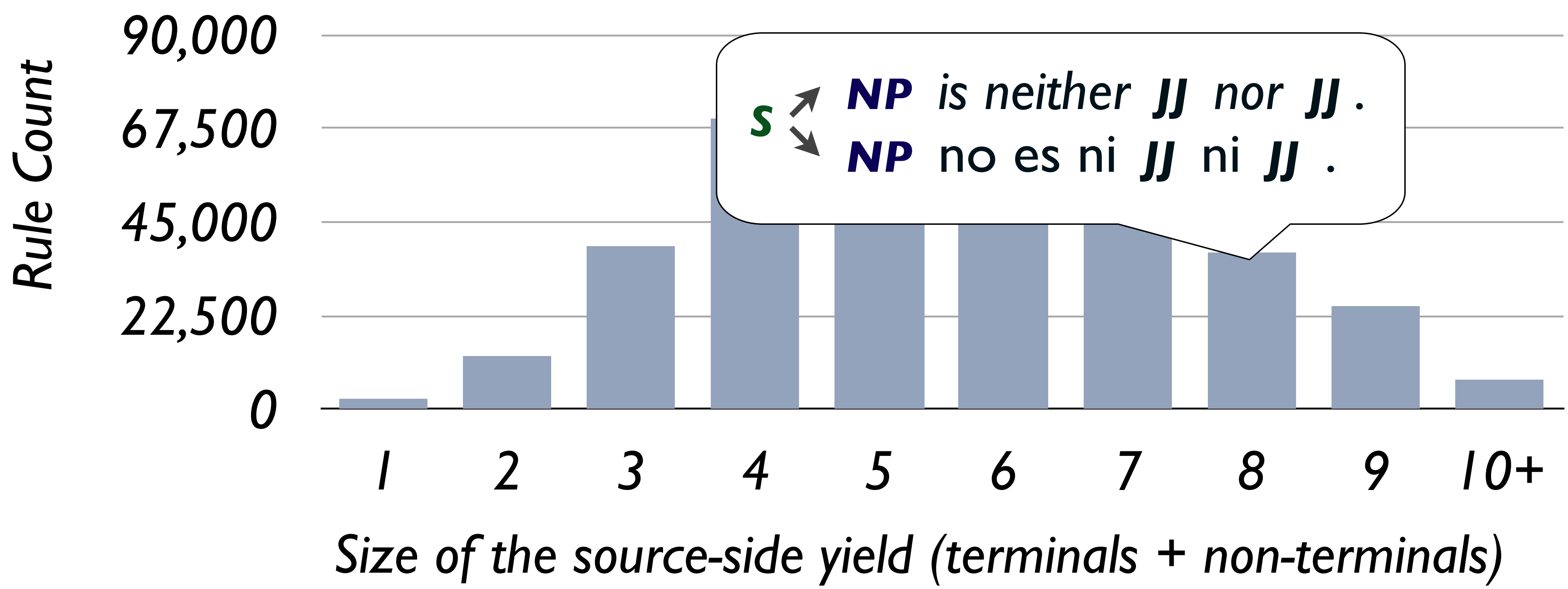


John DeNero, Adam Pauls, Mohit Bansal, and Dan Klein. *Efficient Parsing for Transducer Grammars*, NAACL 2009.

The Size of the Grammar

A grammar learned from 220 million words of Arabic-to-English example translations:

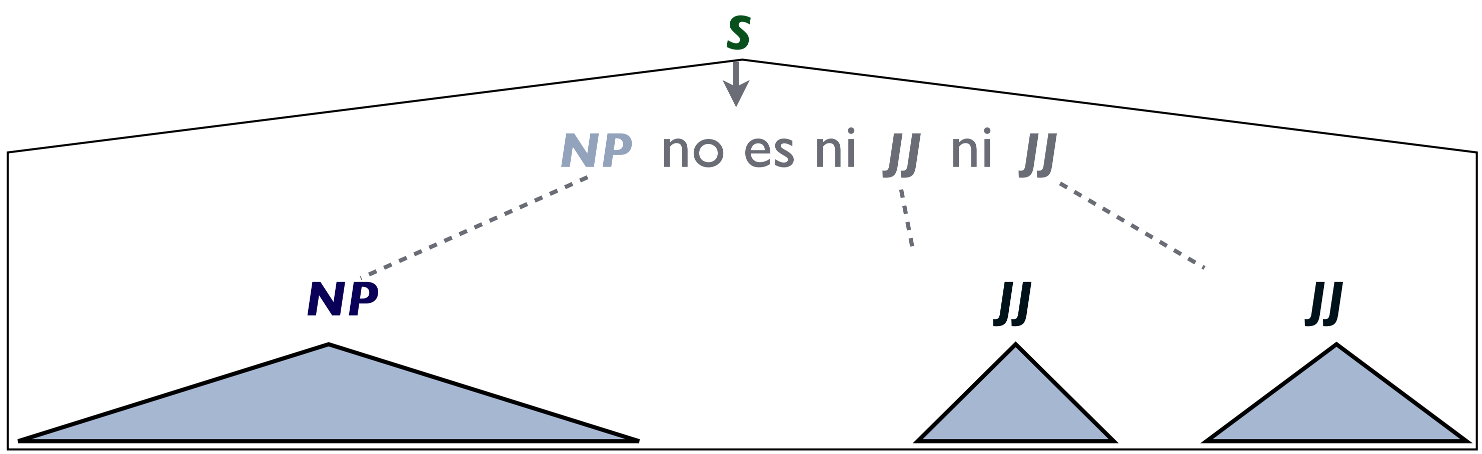
332,000 rules match a 30-word sentence to be translated



John DeNero, Adam Pauls, Mohit Bansal, and Dan Klein. *Efficient Parsing for Transducer Grammars*, NAACL 2009.

The Structure of the Grammar

$S \rightarrow NP$ no es ni JJ ni JJ



Mi dormitorio nuevo no es ni grande ni pequeño

The Structure of the Grammar

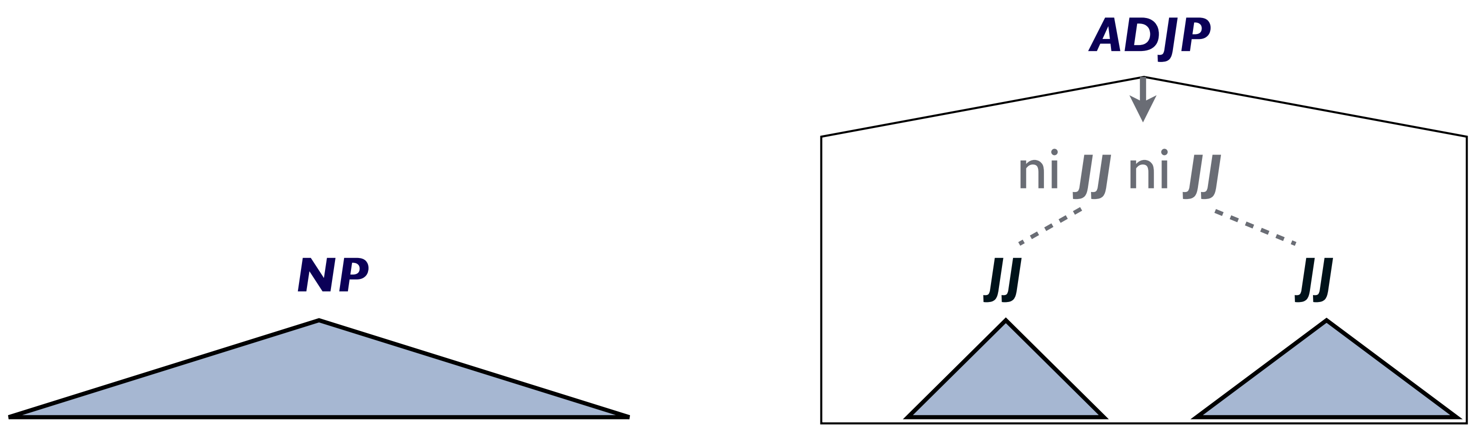
S → ***NP*** no es ni ***JJ*** ni ***JJ***

Mi dormitorio nuevo no es ni grande ni pequeño

The Structure of the Grammar

S → ***NP*** no es ni ***JJ*** ni ***JJ***

ADJP → ni ***JJ*** ni ***JJ***



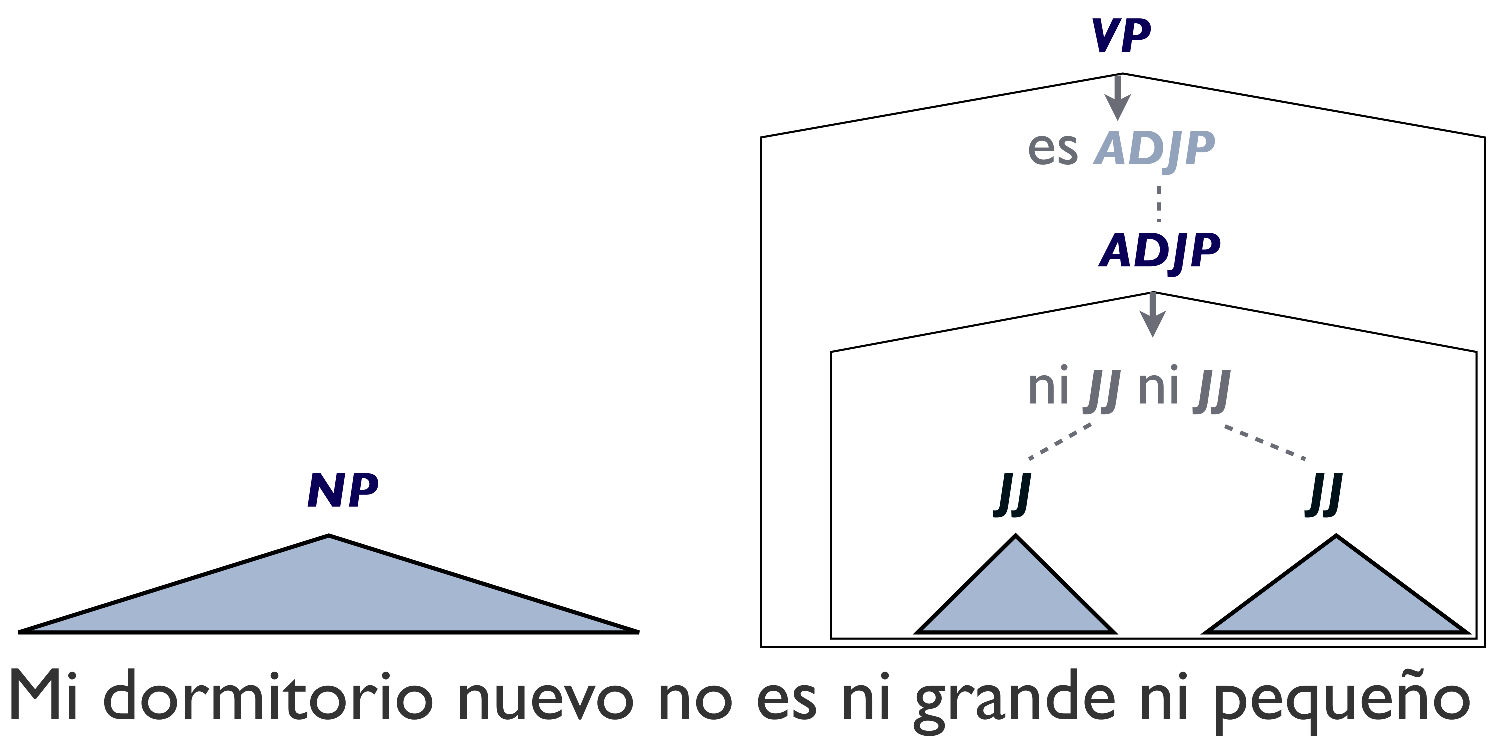
Mi dormitorio nuevo no es ni grande ni pequeño

The Structure of the Grammar

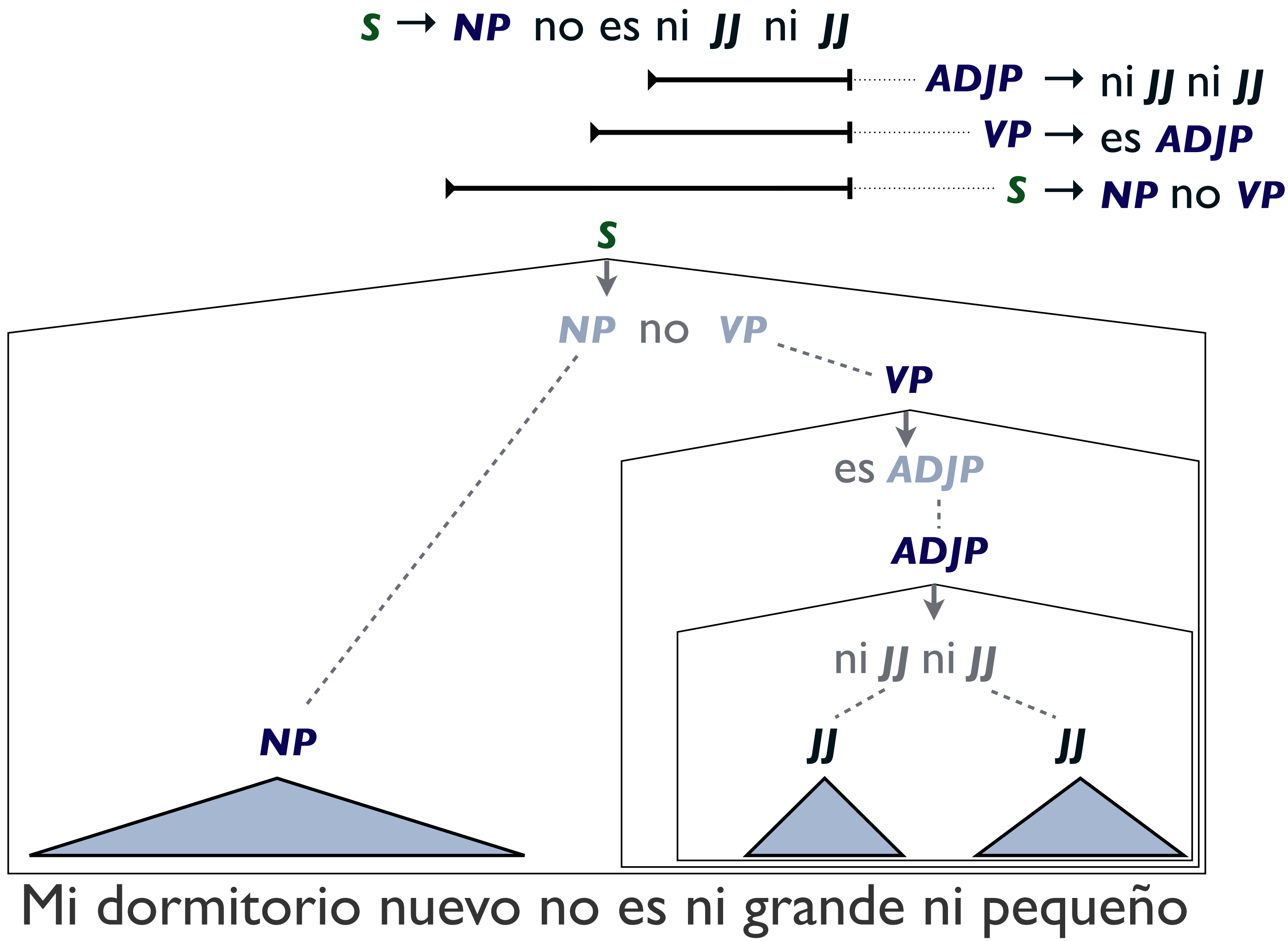
S → ***NP*** no es ni ***JJ*** ni ***JJ***

ADJP → ni ***JJ*** ni ***JJ***

VP → es ***ADJP***



The Structure of the Grammar



Coarse-to-Fine Translation

Mi dormitorio nuevo no es ni grande ni pequeño

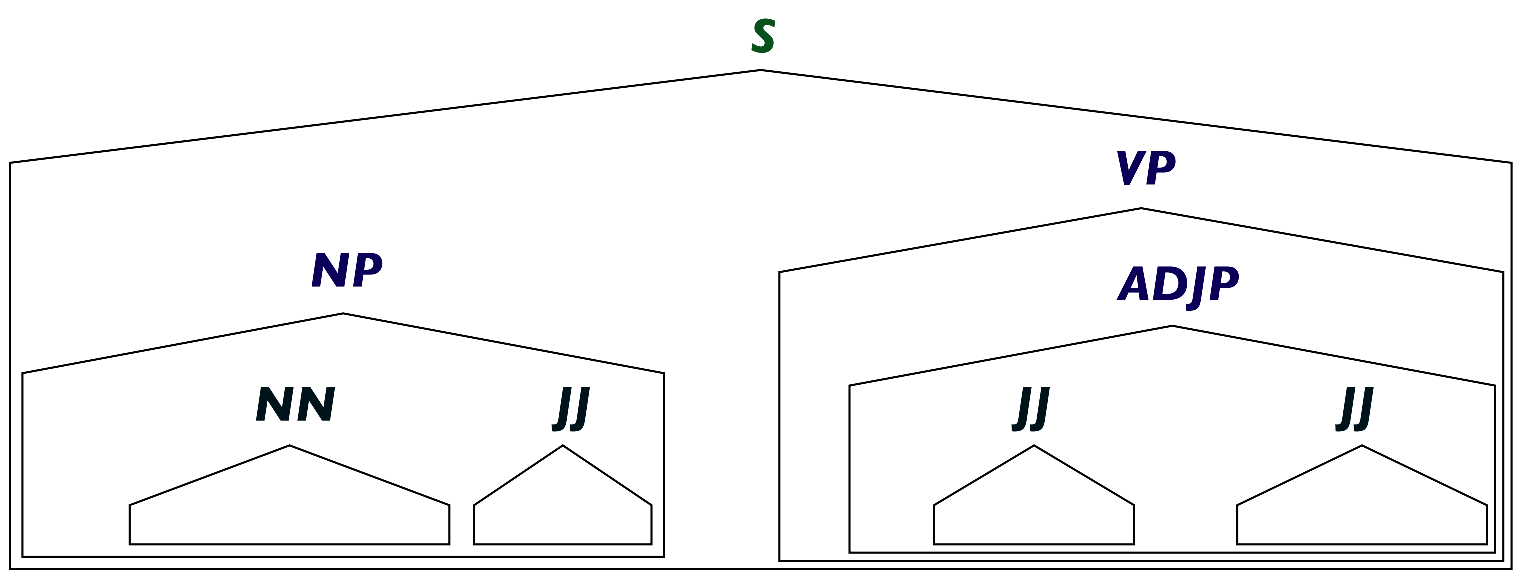
Coarse-to-Fine Translation

1. *Apply a subset of the grammar with only small rules*

Mi dormitorio nuevo no es ni grande ni pequeño

Coarse-to-Fine Translation

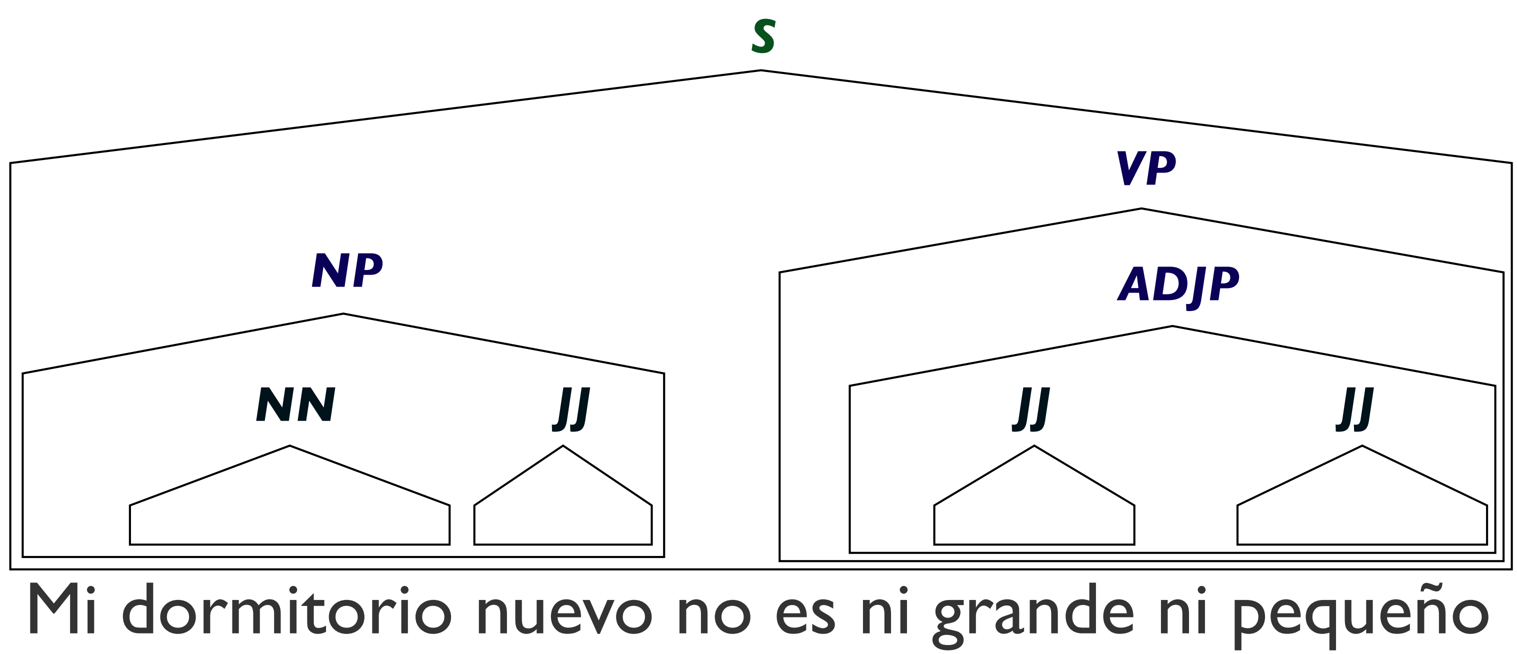
1. *Apply a subset of the grammar with only small rules*



Mi dormitorio nuevo no es ni grande ni pequeño

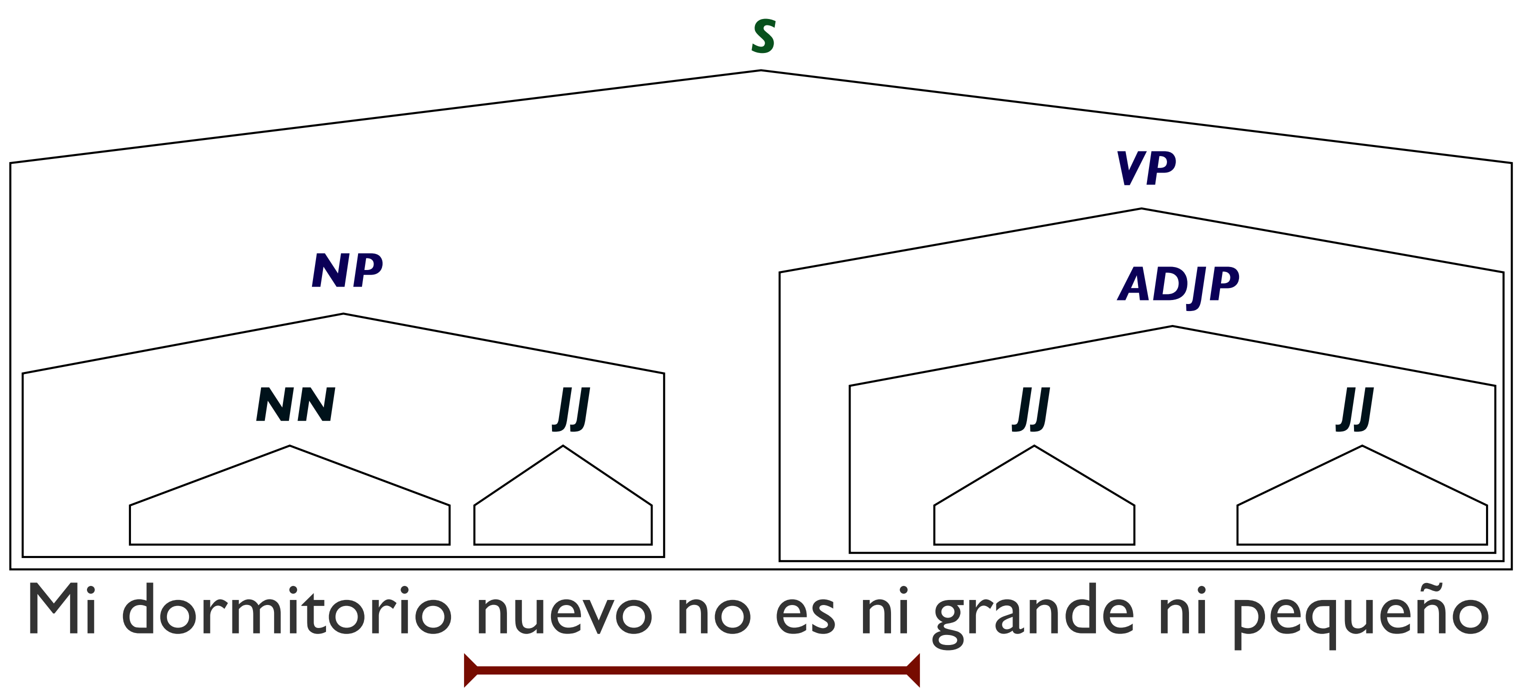
Coarse-to-Fine Translation

1. *Apply a subset of the grammar with only small rules*
2. *Prune away unlikely portions of the search space*



Coarse-to-Fine Translation

1. *Apply a subset of the grammar with only small rules*
2. *Prune away unlikely portions of the search space*



Coarse-to-Fine Translation

1. *Apply a subset of the grammar with only small rules*
2. *Prune away unlikely portions of the search space*

Mi dormitorio nuevo no es ni grande ni pequeño

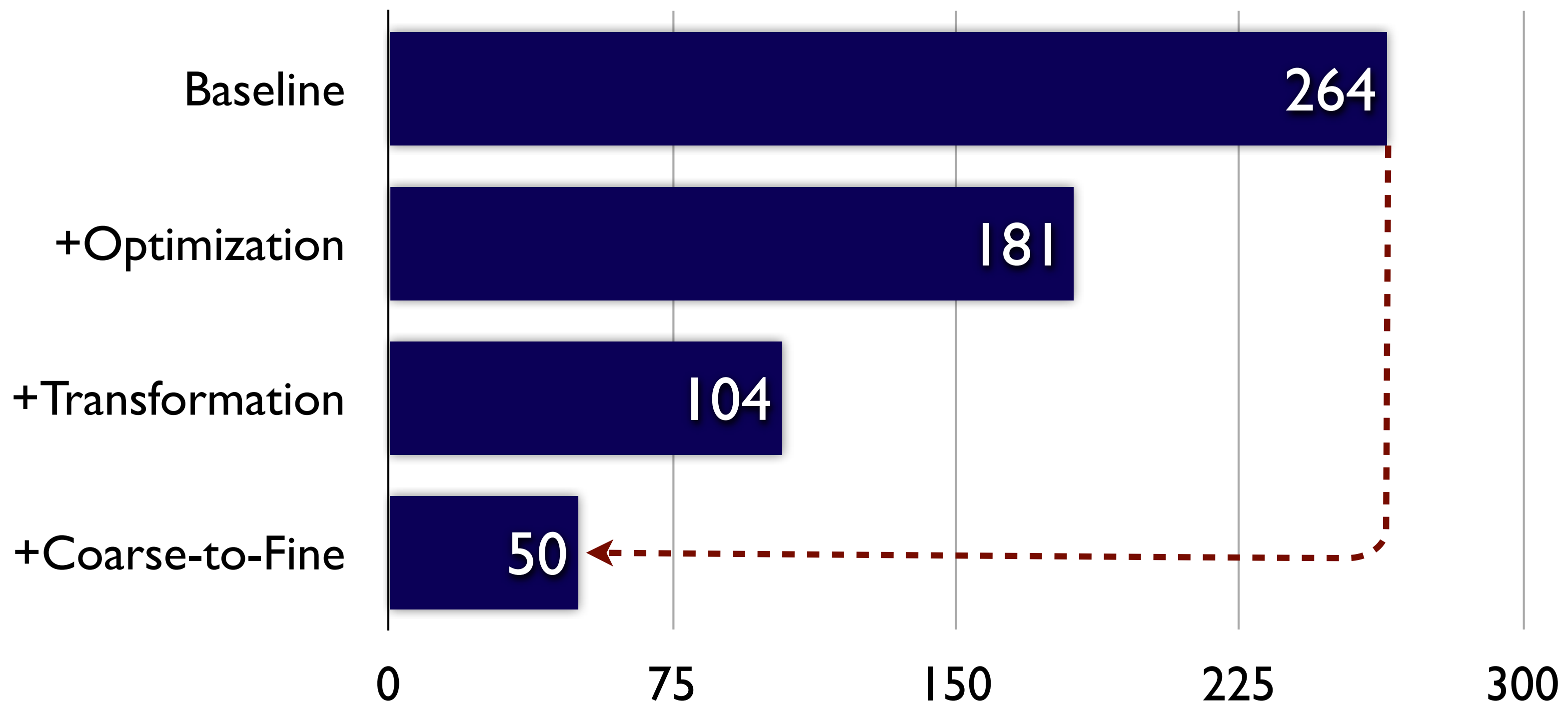
Coarse-to-Fine Translation

1. *Apply a subset of the grammar with only small rules*
2. *Prune away unlikely portions of the search space*
3. *Apply the full translation grammar to the pruned space*

Mi dormitorio nuevo no es ni grande ni pequeño

Experimental Results

Minutes required to analyze a 300 sentence test set



5x speed-up with the largest translation grammars in use today (ISI Syntax-Based MT System) [DeNero et al. NAACL '09]*

* John DeNero, Adam Pauls, Mohit Bansal, and Dan Klein. *Efficient Parsing for Transducer Grammars*, NAACL 2009.

The Steps in a Modern Translation System

Learn a model

Apply the model

Choose a translation

The Steps in a Modern Translation System

Learn a model

Apply the model

Choose a translation

- Fully exploiting large data sets requires searching over very large spaces

The Steps in a Modern Translation System

Learn a model | **Apply the model** | Choose a translation

- Fully exploiting large data sets requires searching over very large spaces
- Coarse-to-fine inference is a powerful technique for doing so

The Steps in a Modern Translation System

Learn a model

Apply the model

Choose a translation

Even the Best Models are Wrong

Yo lo haré después

NOVEL SENTENCE

Model

Even the Best Models are Wrong

Yo lo haré después

NOVEL SENTENCE

Model

Later do it I will

I will later do it

That I’ll do later

Later that I’ll do

...

Even the Best Models are Wrong

Yo lo haré después

NOVEL SENTENCE

Model

Later do it I will

I will later do it

That I'll do later

Later that I'll do

...

+ Samples from output space

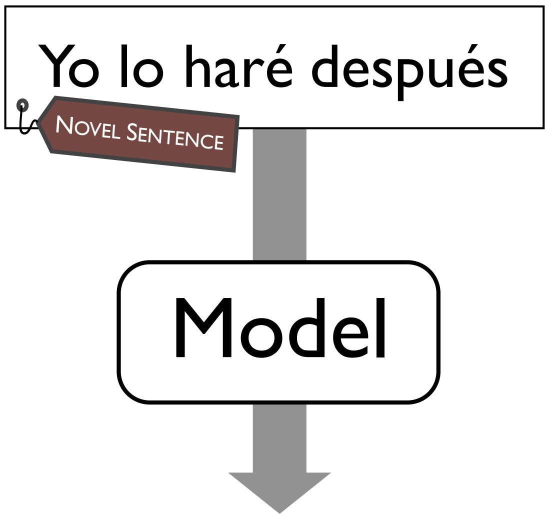
× Samples near maximum

◆ Highest scoring translation

Translation Quality (BLEU)

Total model score for 1000 sentences

Even the Best Models are Wrong



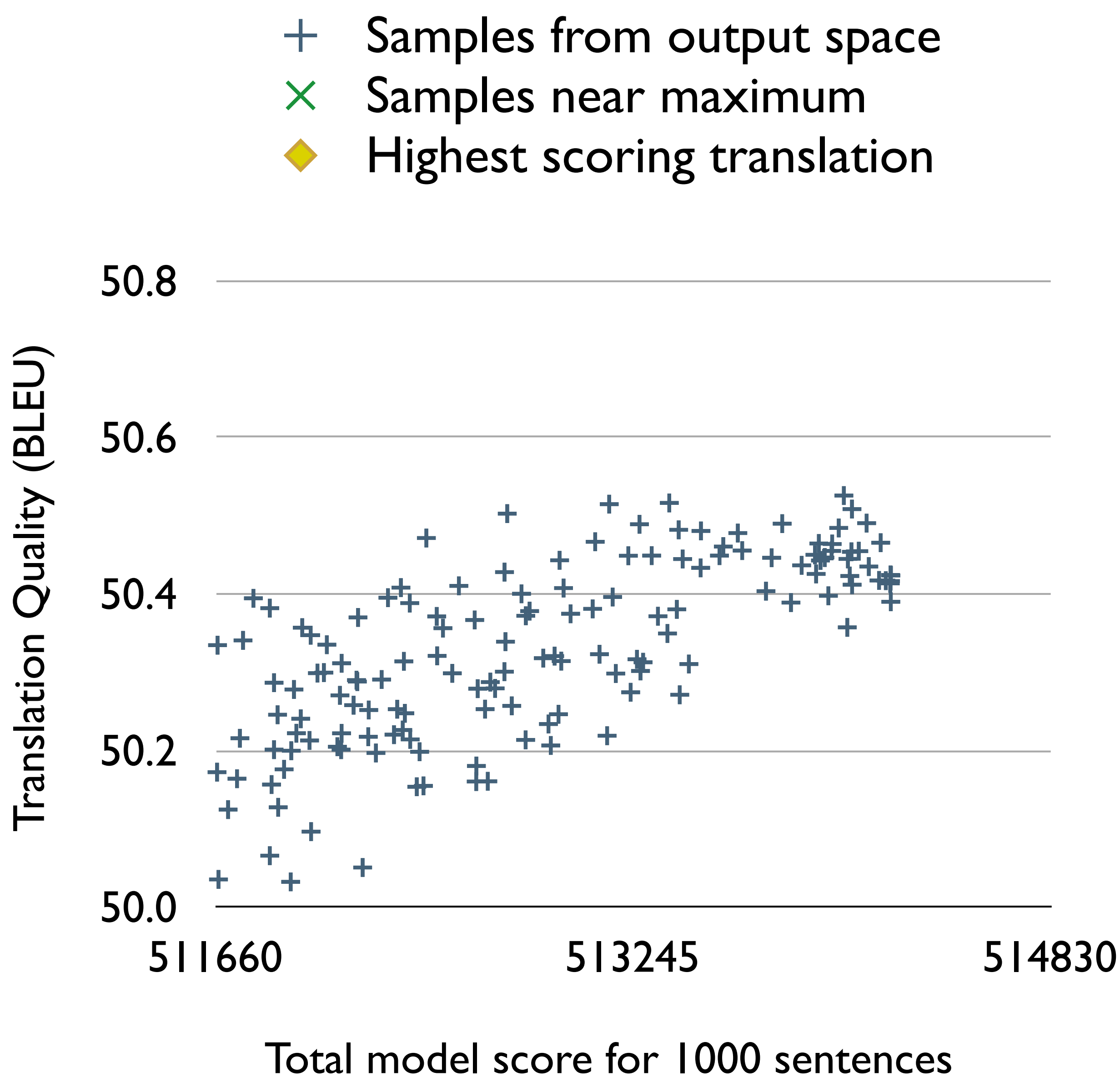
Later do it I will

I will later do it

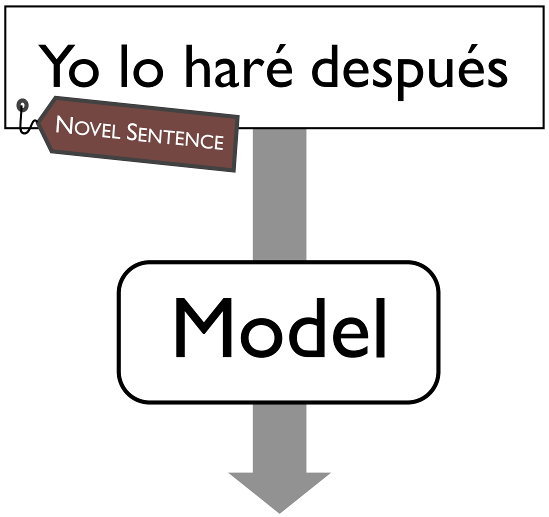
That I'll do later

Later that I'll do

...



Even the Best Models are Wrong



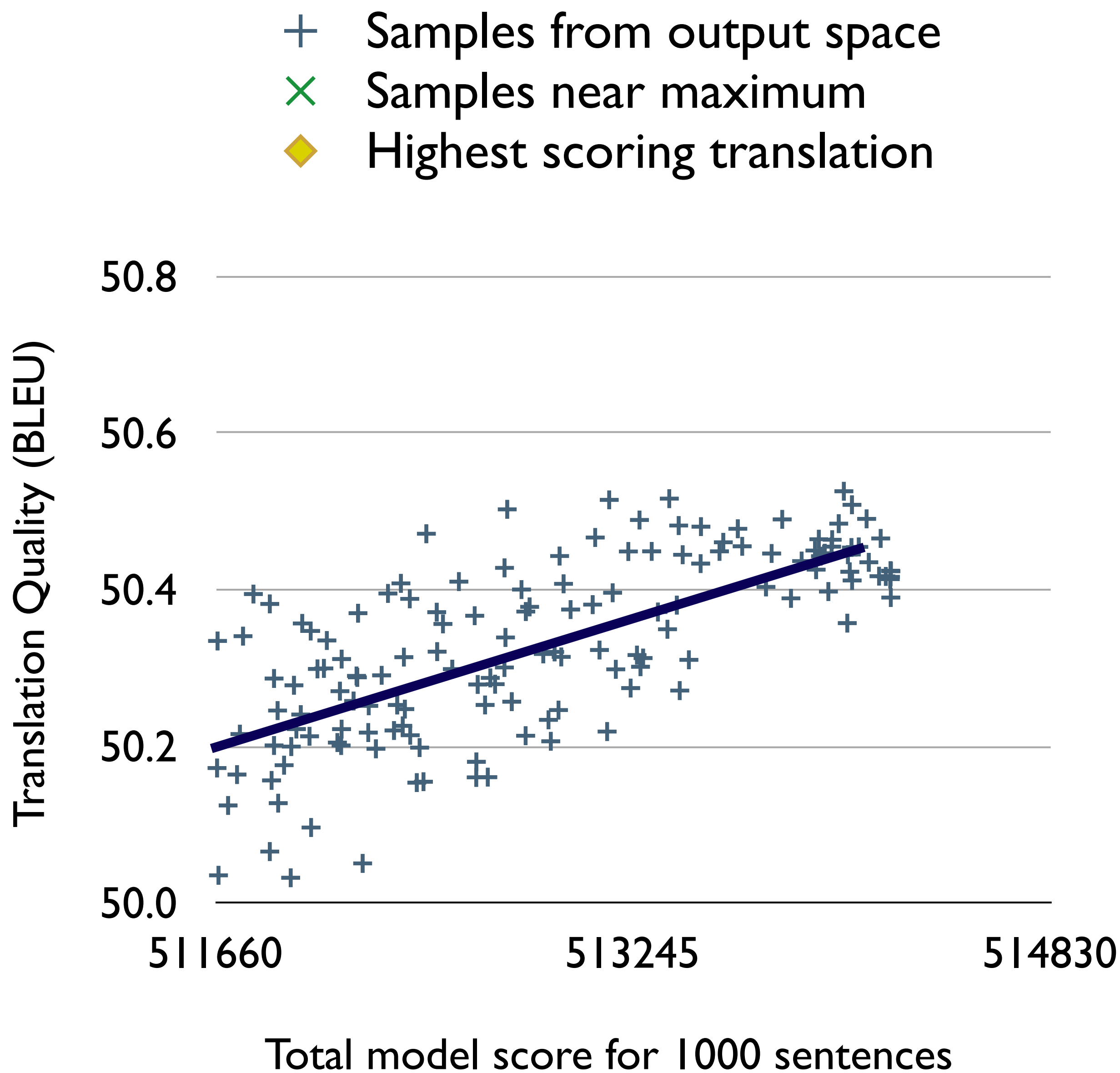
Later do it I will

I will later do it

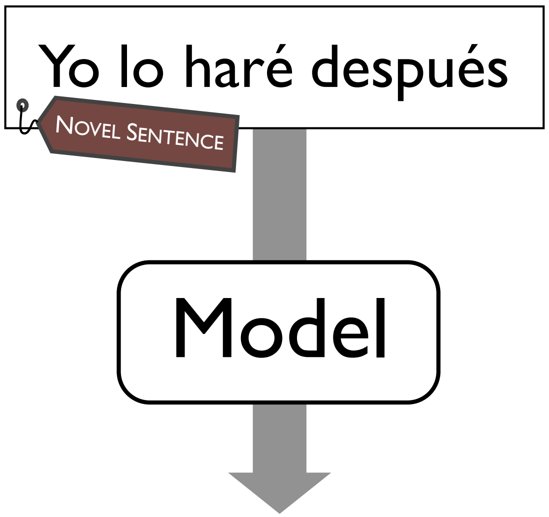
That I'll do later

Later that I'll do

...



Even the Best Models are Wrong



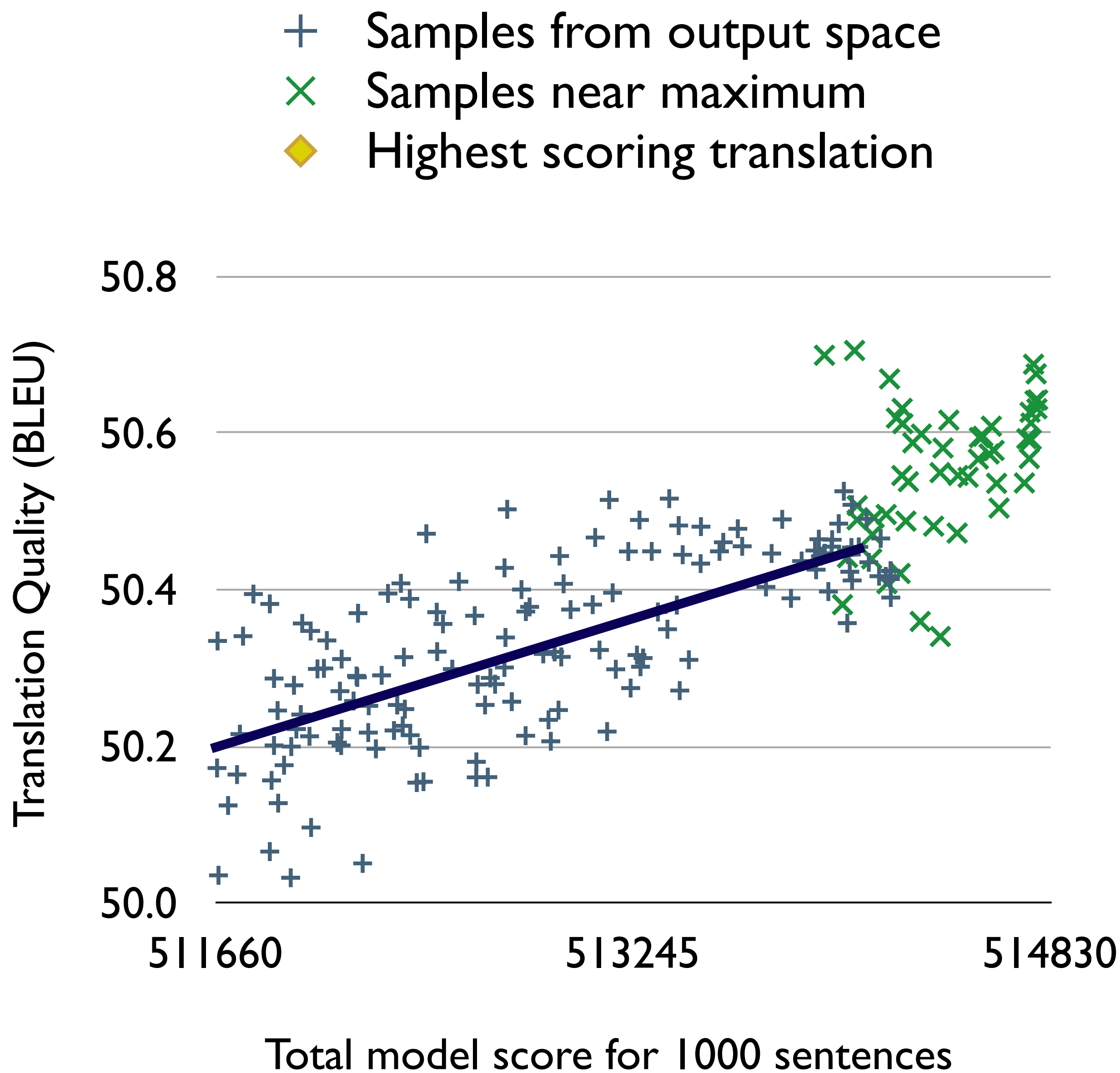
Later do it I will

I will later do it

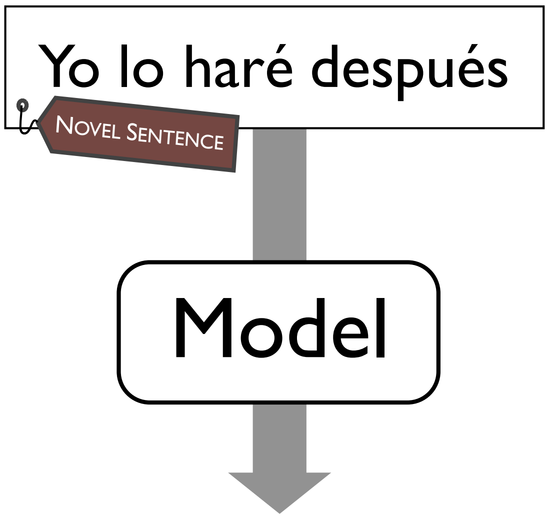
That I'll do later

Later that I'll do

...



Even the Best Models are Wrong



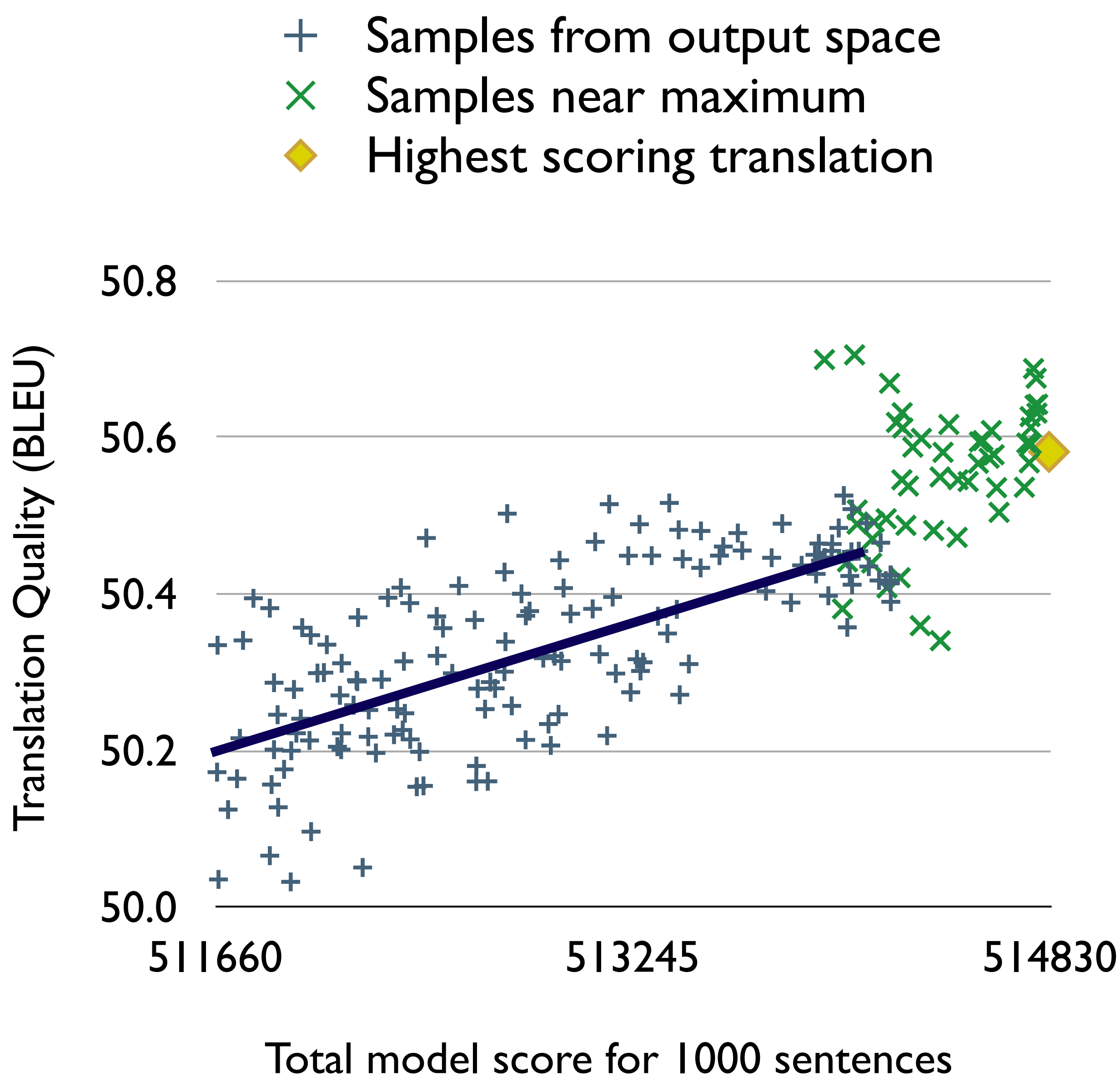
- Later do it I will

  I will later do it

  That I'll do later

  Later that I'll do

  ...



Consensus by Averaging Over Sentences

Yo lo haré después

NOVEL SENTENCE

Model

Later do it I will

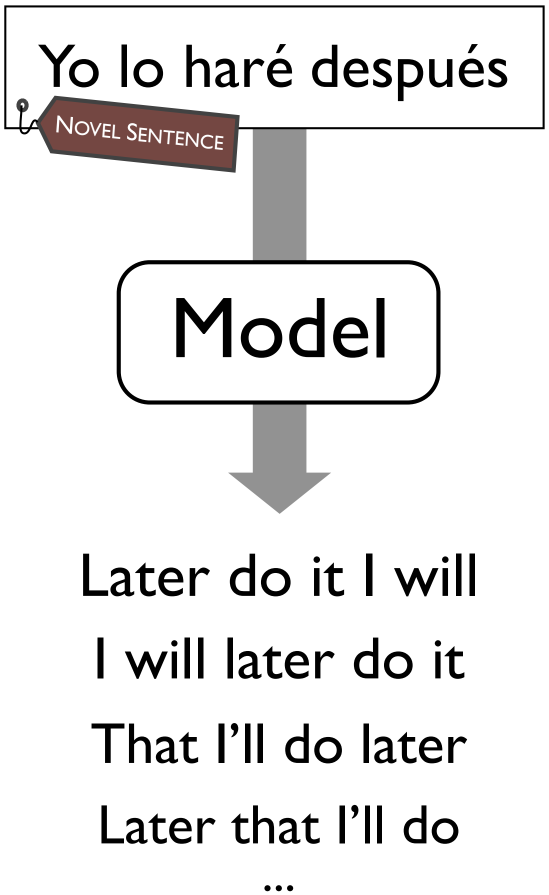
I will later do it

That I'll do later

Later that I'll do

...

Consensus by Averaging Over Sentences



Intuition: "Happy families are all alike; every unhappy family is unhappy in its own way."   [Tolstoy. 1877]*

* Leo Tolstoy. Анна Каренина. 1877.

Consensus by Averaging Over Sentences

Yo lo haré después

NOVEL SENTENCE

Model

Later do it I will

I will later do it

That I'll do later

Later that I'll do

...

Intuition: "Happy families are all alike; every unhappy family is unhappy in its own way." [Tolstoy. 1877]*

Idea: Average over sentences to find the phrases that are alike. [DeNero et al. ACL '09]**

* Leo Tolstoy. Анна Каренина. 1877.

** John DeNero, David Chiang, and Kevin Knight. *Fast Consensus Decoding over Translation Forests*, ACL 2009.

Consensus by Averaging Over Sentences

Yo lo haré después

NOVEL SENTENCE

Model

Later do it I will

I will later do it

That I'll do later

Later that I'll do

...

Intuition: "Happy families are all alike; every unhappy family is unhappy in its own way." [Tolstoy. 1877]*

Idea: Average over sentences to find the phrases that are alike. [DeNero et al. ACL '09]**

"Later" "do" ... "do it" "I'll" "do later" ..."do it I will"

* Leo Tolstoy. Анна Каренина. 1877.

** John DeNero, David Chiang, and Kevin Knight. *Fast Consensus Decoding over Translation Forests*, ACL 2009.

Consensus by Averaging Over Sentences

Yo lo haré después

NOVEL SENTENCE

Model

Later do it I will

I will later do it

That I'll do later

Later that I'll do

...

Intuition: "Happy families are all alike; every unhappy family is unhappy in its own way." [Tolstoy. 1877]*

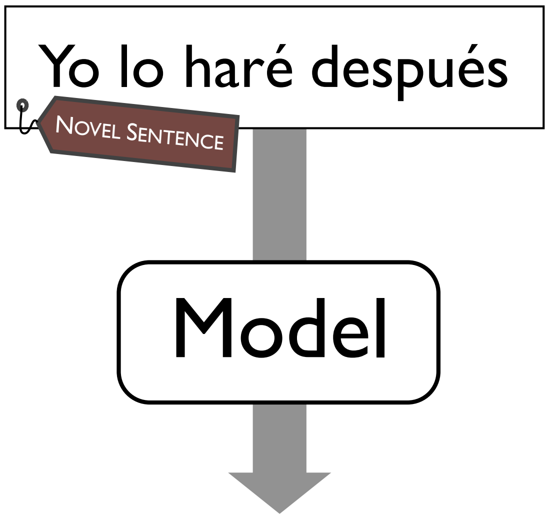
Idea: Average over sentences to find the phrases that are alike. [DeNero et al. ACL '09]**

"Later"	"do"	...	"do it"	"I'll"	"do later"	...	"do it I will"
1	1		1	0	0		1

* Leo Tolstoy. Анна Каренина. 1877.

** John DeNero, David Chiang, and Kevin Knight. *Fast Consensus Decoding over Translation Forests*, ACL 2009.

Consensus by Averaging Over Sentences



Later do it I will

I will later do it

That I'll do later

Later that I'll do

...

Intuition: "Happy families are all alike; every unhappy family is unhappy in its own way." [Tolstoy. 1877]*

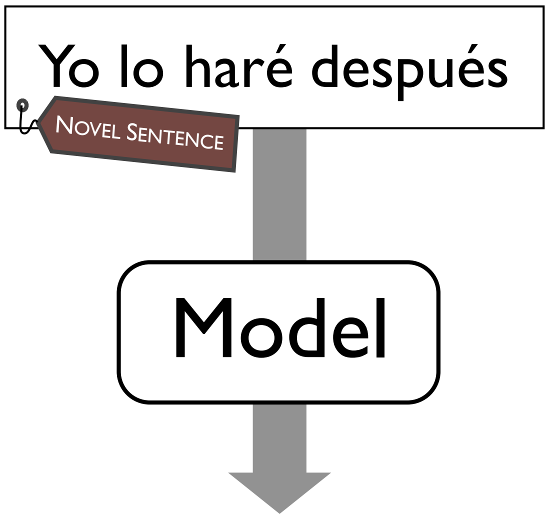
Idea: Average over sentences to find the phrases that are alike. [DeNero et al. ACL '09]**

"Later"	"do"	...	"do it"	"I'll"	"do later"	...	"do it I will"
1	1		1	0	0		1
1	1		1	0	0		0

* Leo Tolstoy. Анна Каренина. 1877.

** John DeNero, David Chiang, and Kevin Knight. *Fast Consensus Decoding over Translation Forests*, ACL 2009.

Consensus by Averaging Over Sentences



Later do it I will
I will later do it
That I'll do later
Later that I'll do
...

Intuition: "Happy families are all alike; every unhappy family is unhappy in its own way." [Tolstoy. 1877]*

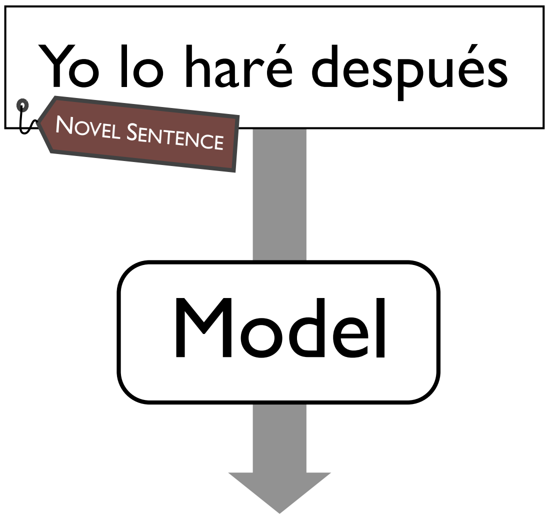
Idea: Average over sentences to find the phrases that are alike. [DeNero et al. ACL '09]**

"Later"	"do"	...	"do it"	"I'll"	"do later"	...	"do it I will"
1	1		1	0	0		1
1	1		1	0	0		0
1	1		0	1	1		0
1	1		0	1	0		0

* Leo Tolstoy. Анна Каренина. 1877.

** John DeNero, David Chiang, and Kevin Knight. *Fast Consensus Decoding over Translation Forests*, ACL 2009.

Consensus by Averaging Over Sentences



Intuition: "Happy families are all alike; every unhappy family is unhappy in its own way." [Tolstoy. 1877]*

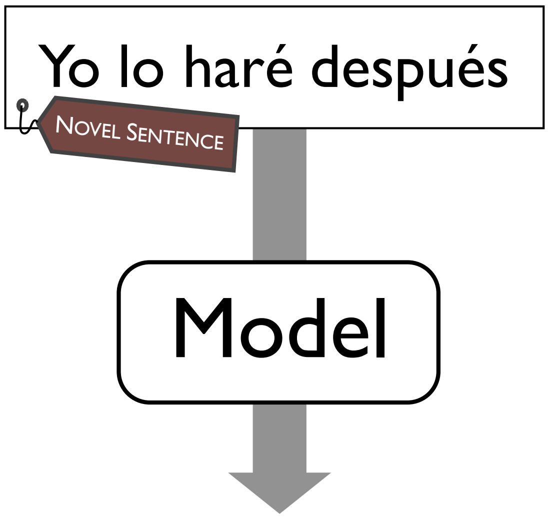
Idea: Average over sentences to find the phrases that are alike. [DeNero et al. ACL '09]**

		"Later"	"do"	...	"do it"	"I'll"	"do later"	...	"do it I will"
Later do it I will	0.12	1	1		1	0	0		1
I will later do it	0.10	1	1		1	0	0		0
That I'll do later	0.07	1	1		0	1	1		0
Later that I'll do	0.07	1	1		0	1	0		0
...	...								

* Leo Tolstoy. Анна Каренина. 1877.

** John DeNero, David Chiang, and Kevin Knight. *Fast Consensus Decoding over Translation Forests*, ACL 2009.

Consensus by Averaging Over Sentences



Intuition: "Happy families are all alike; every unhappy family is unhappy in its own way." [Tolstoy. 1877]*

Idea: Average over sentences to find the phrases that are alike. [DeNero et al. ACL '09]**

		"Later"	"do"	...	"do it"	"I'll"	"do later"	...	"do it I will"
Later do it I will	0.12	1	1		1	0	0		1
I will later do it	0.10	1	1		1	0	0		0
That I'll do later	0.07	1	1		0	1	1		0
Later that I'll do	0.07	1	1		0	1	0		0
...	...								
Expected output	1.00	0.97	0.98		0.54	0.41	0.34		0.12

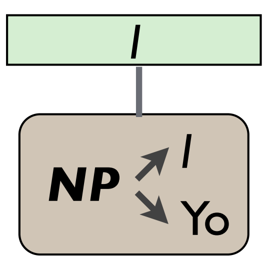
* Leo Tolstoy. Анна Каренина. 1877.

** John DeNero, David Chiang, and Kevin Knight. *Fast Consensus Decoding over Translation Forests*, ACL 2009.

Phrase Expectations from Forests

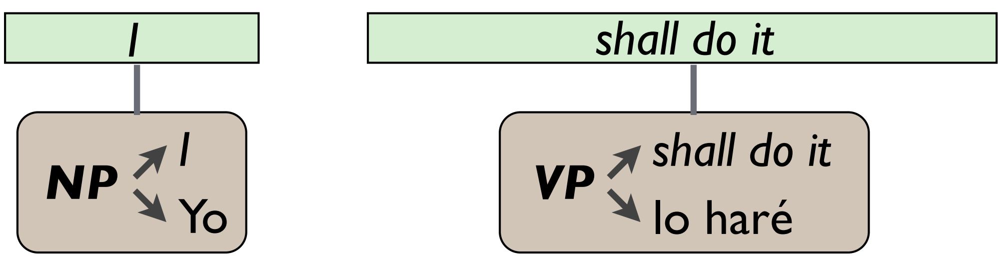
Yo lo haré después

Phrase Expectations from Forests



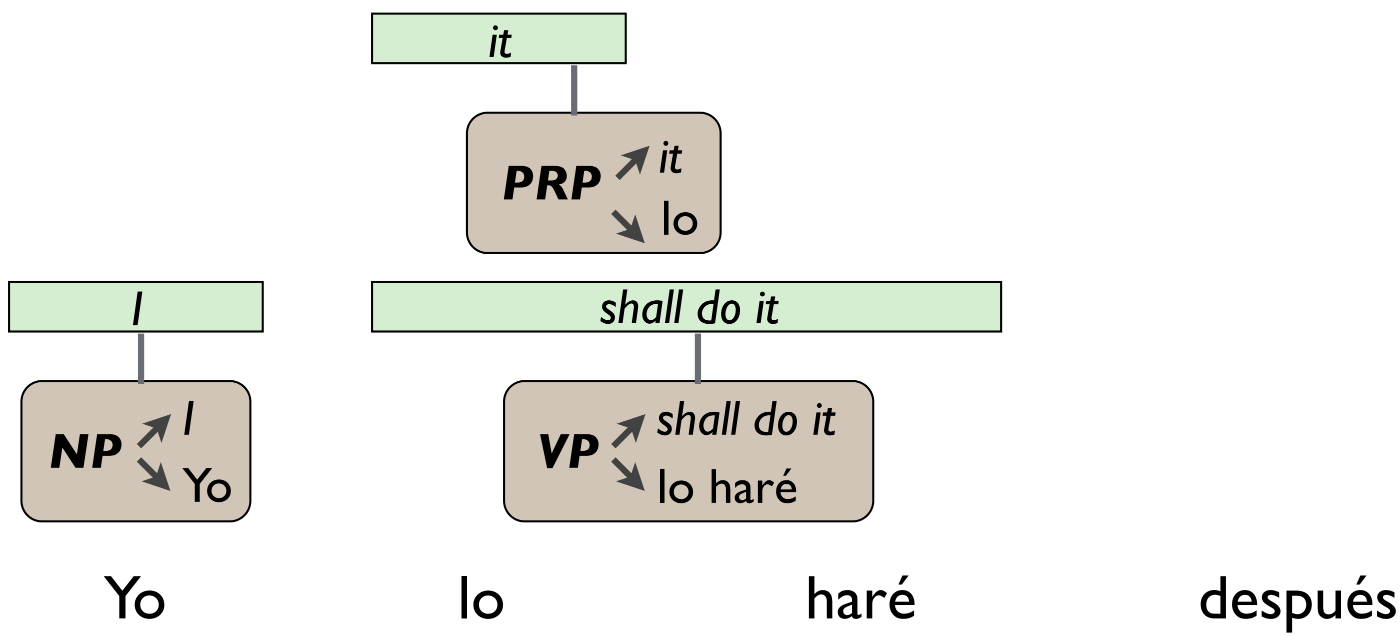
Yo lo haré después

Phrase Expectations from Forests

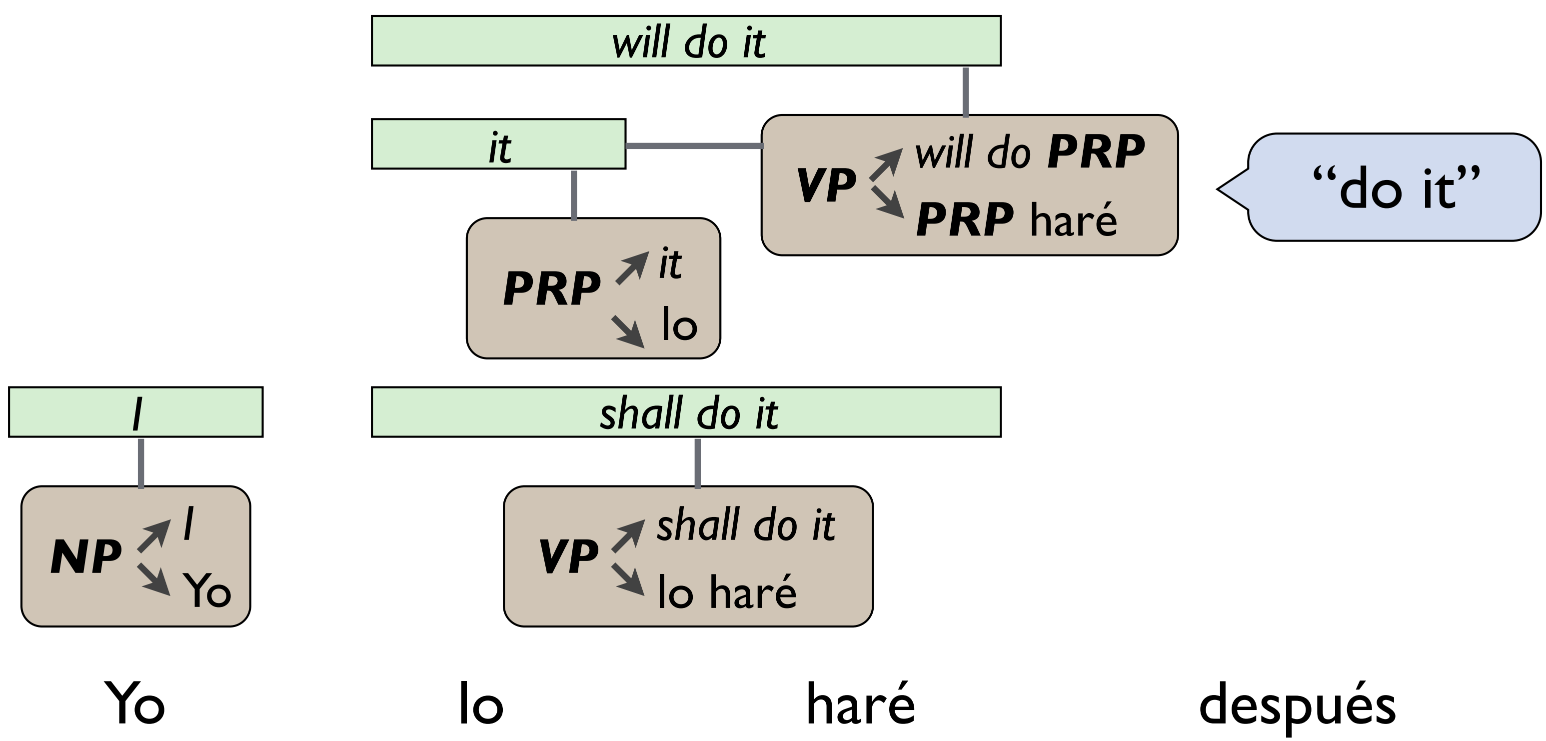


Yo lo haré después

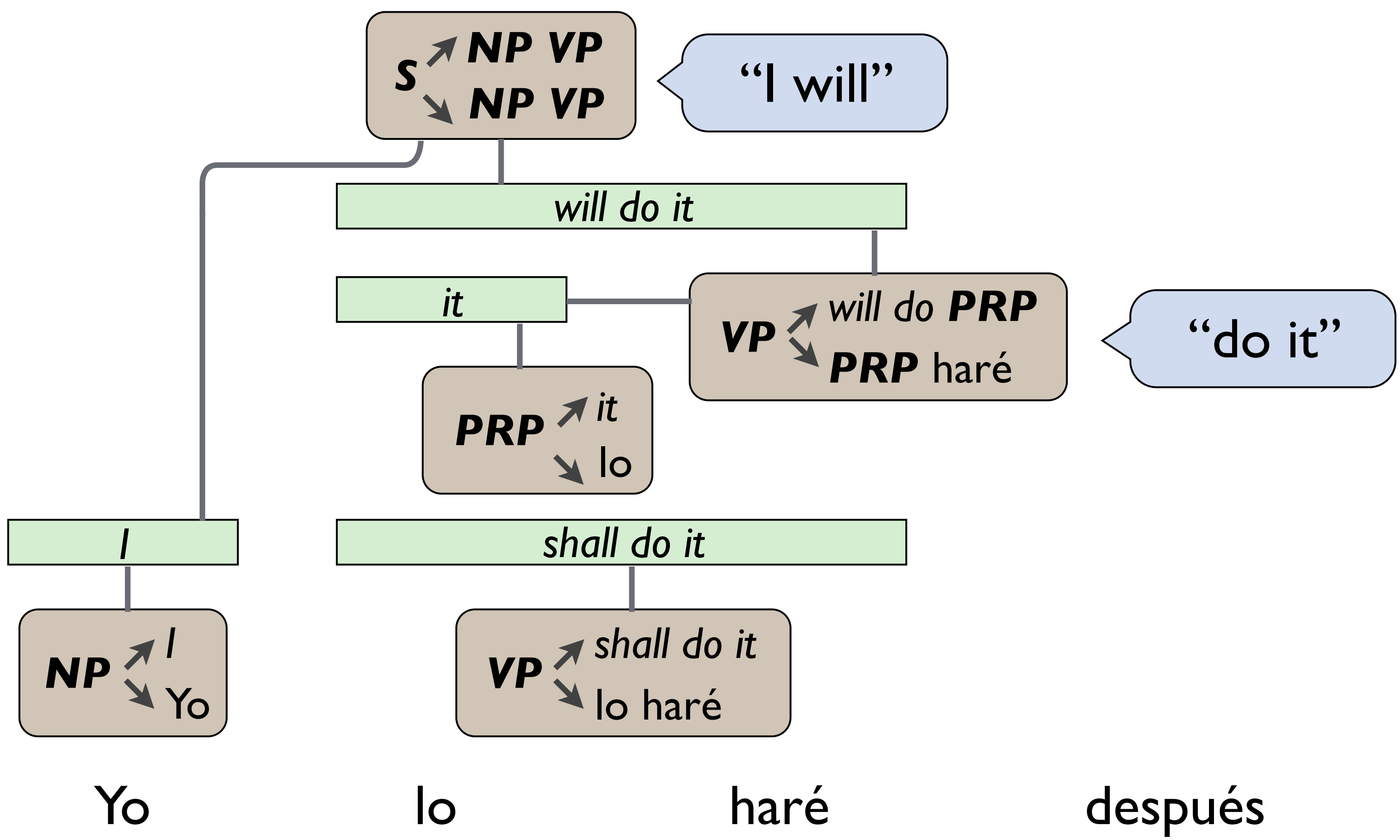
Phrase Expectations from Forests



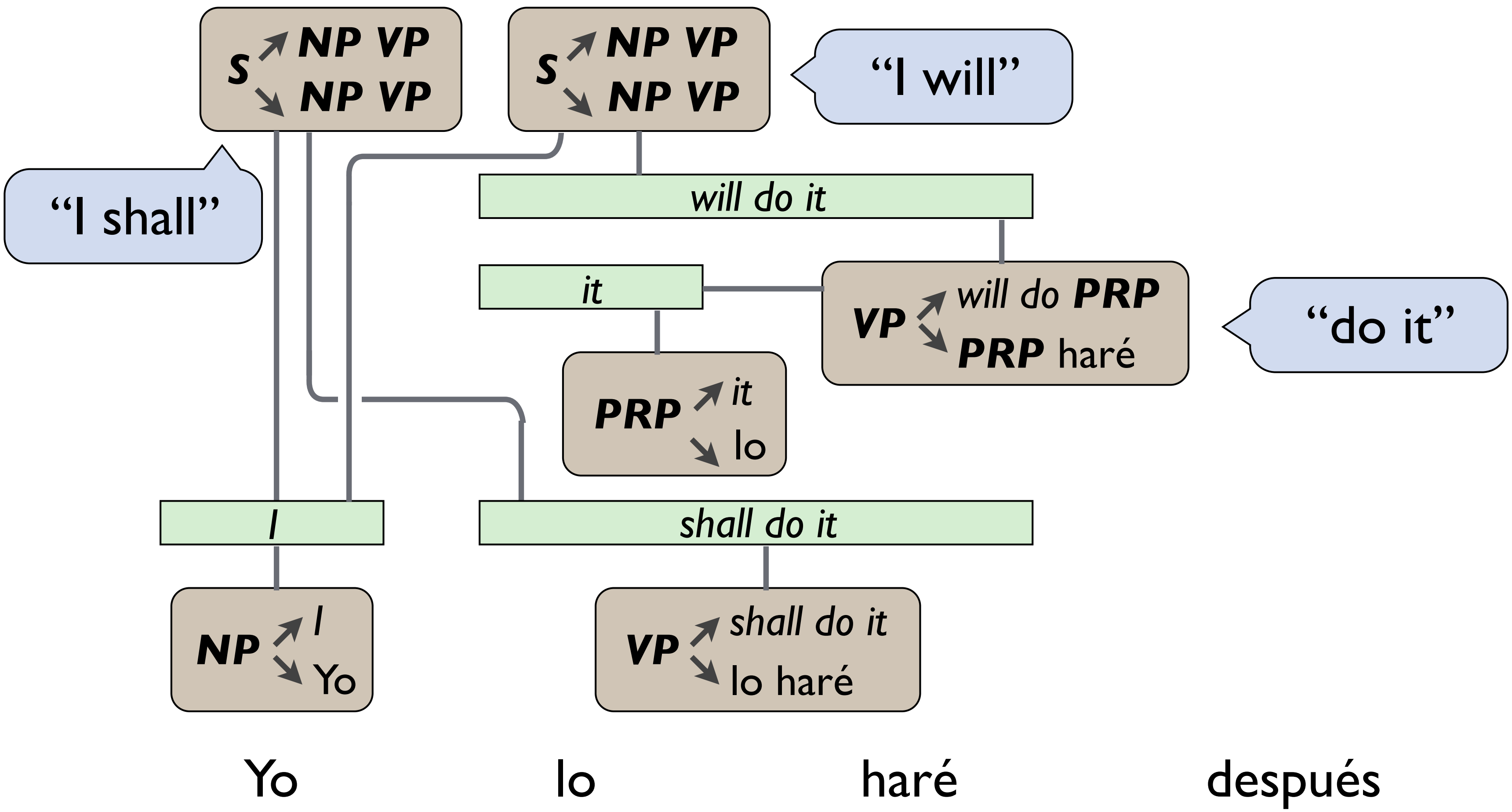
Phrase Expectations from Forests



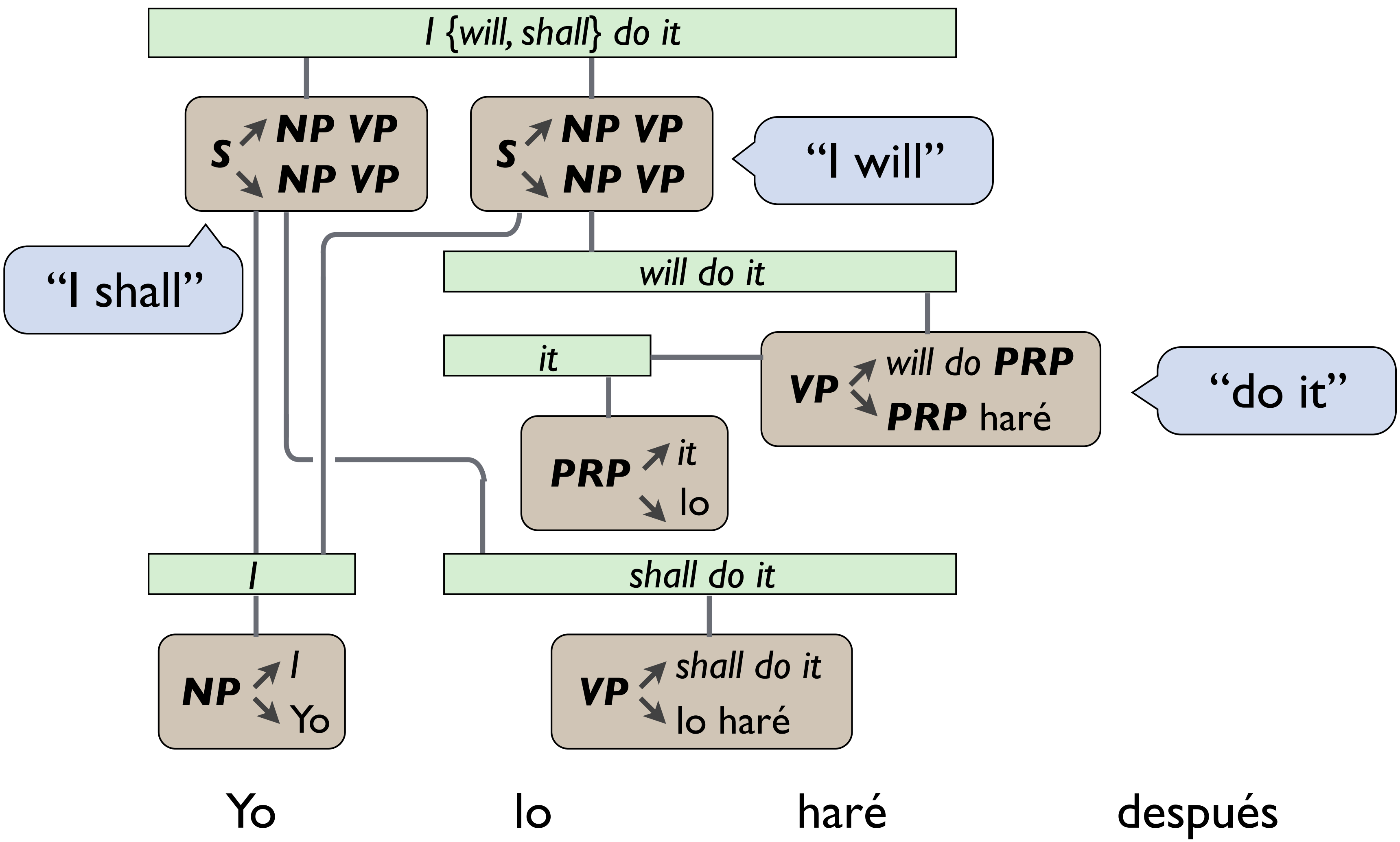
Phrase Expectations from Forests



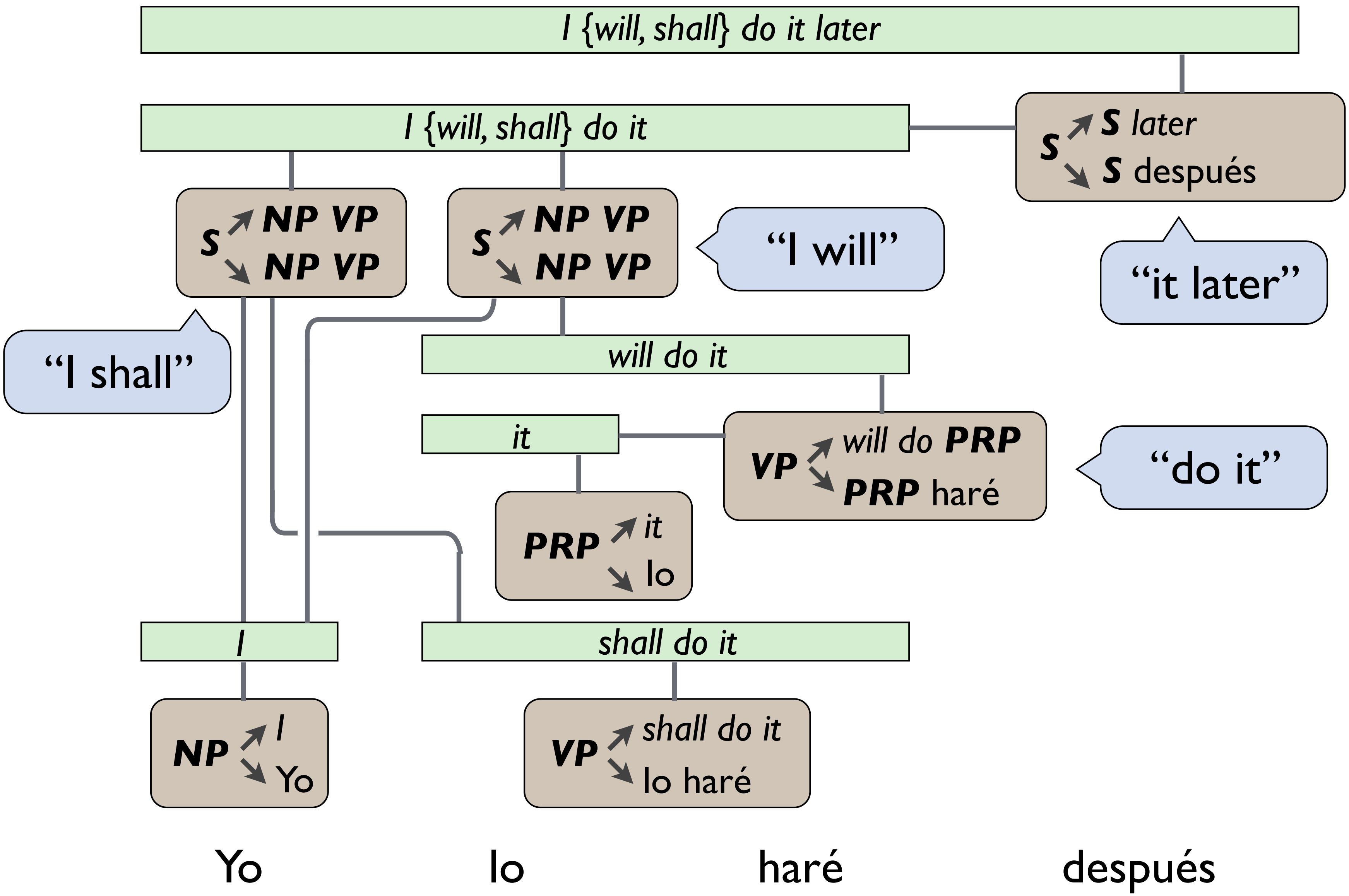
Phrase Expectations from Forests



Phrase Expectations from Forests



Phrase Expectations from Forests

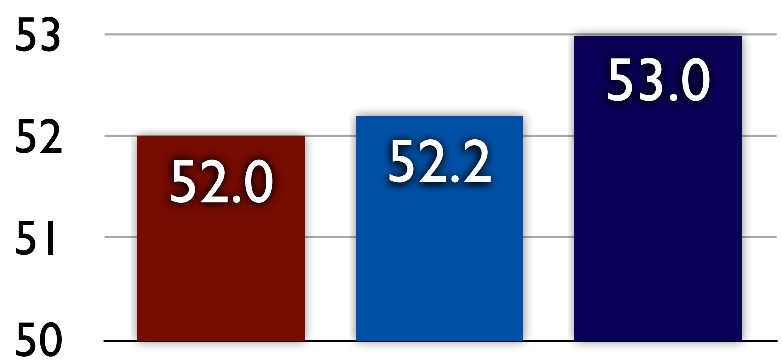


Single System Translation Results

Translation quality in ISI's Full-Scale
Arabic-to-English Hierarchical Translation System

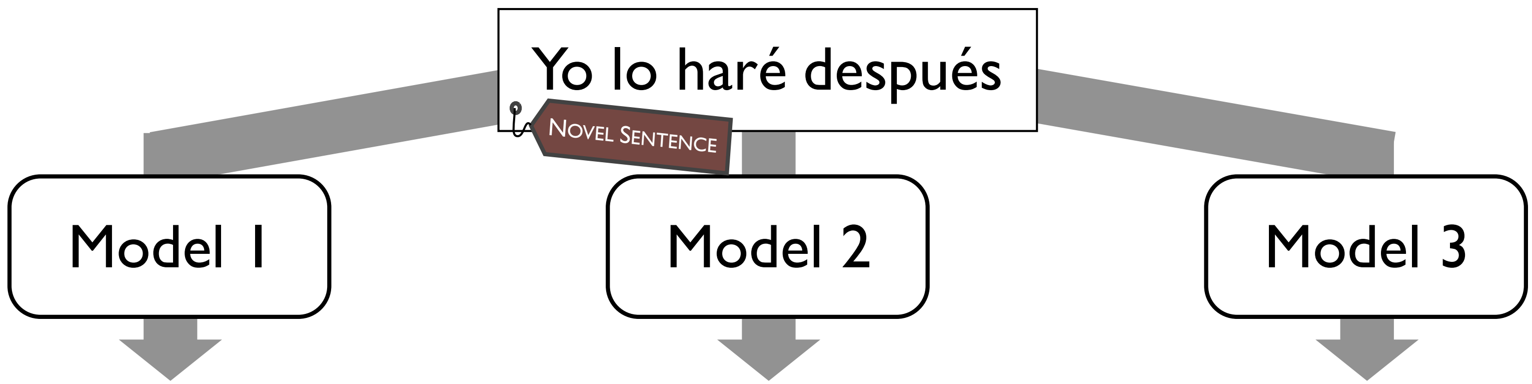
- Model-Only Baseline
- Consensus from a List [DeNero et al. ACL '09]*
- Consensus from a Forest [DeNero et al. ACL '09]*

Translation Quality (BLEU)

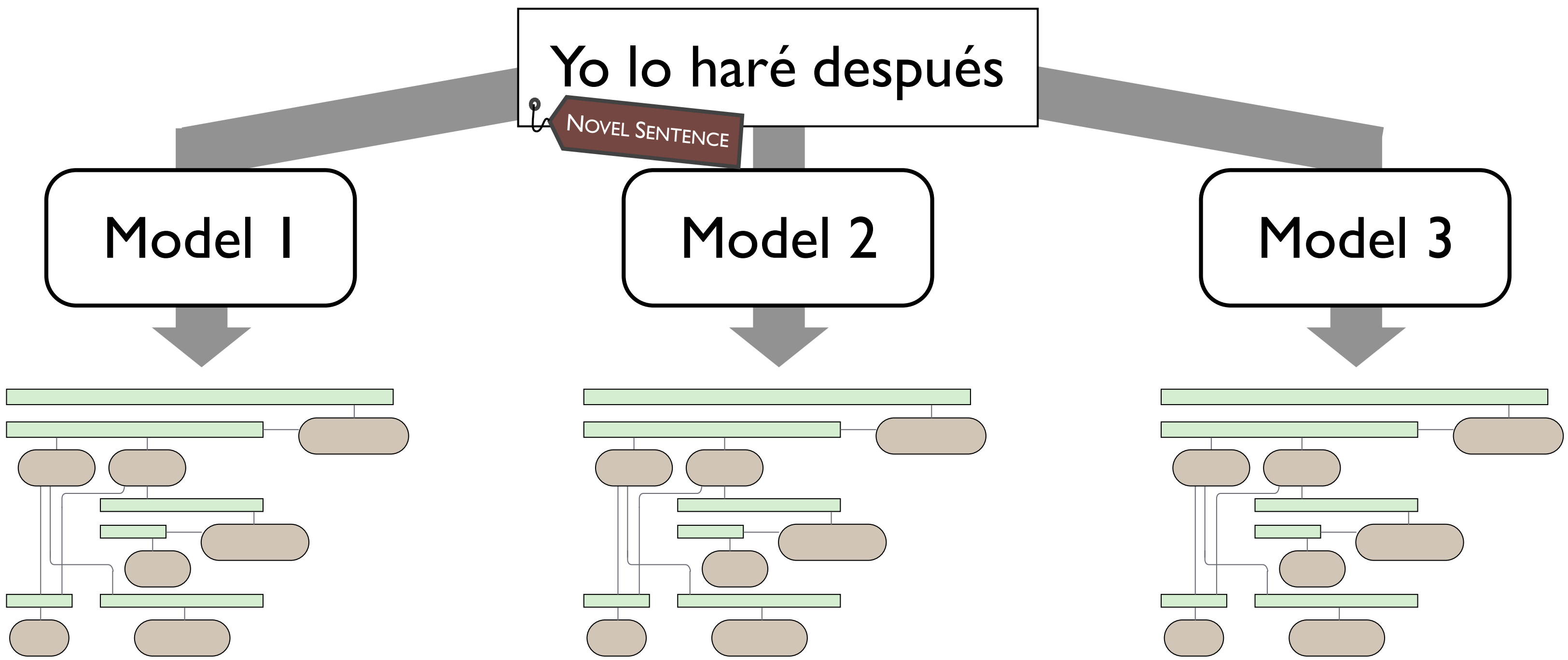


* John DeNero, David Chiang, and Kevin Knight. *Fast Consensus Decoding over Translation Forests*, ACL 2009.

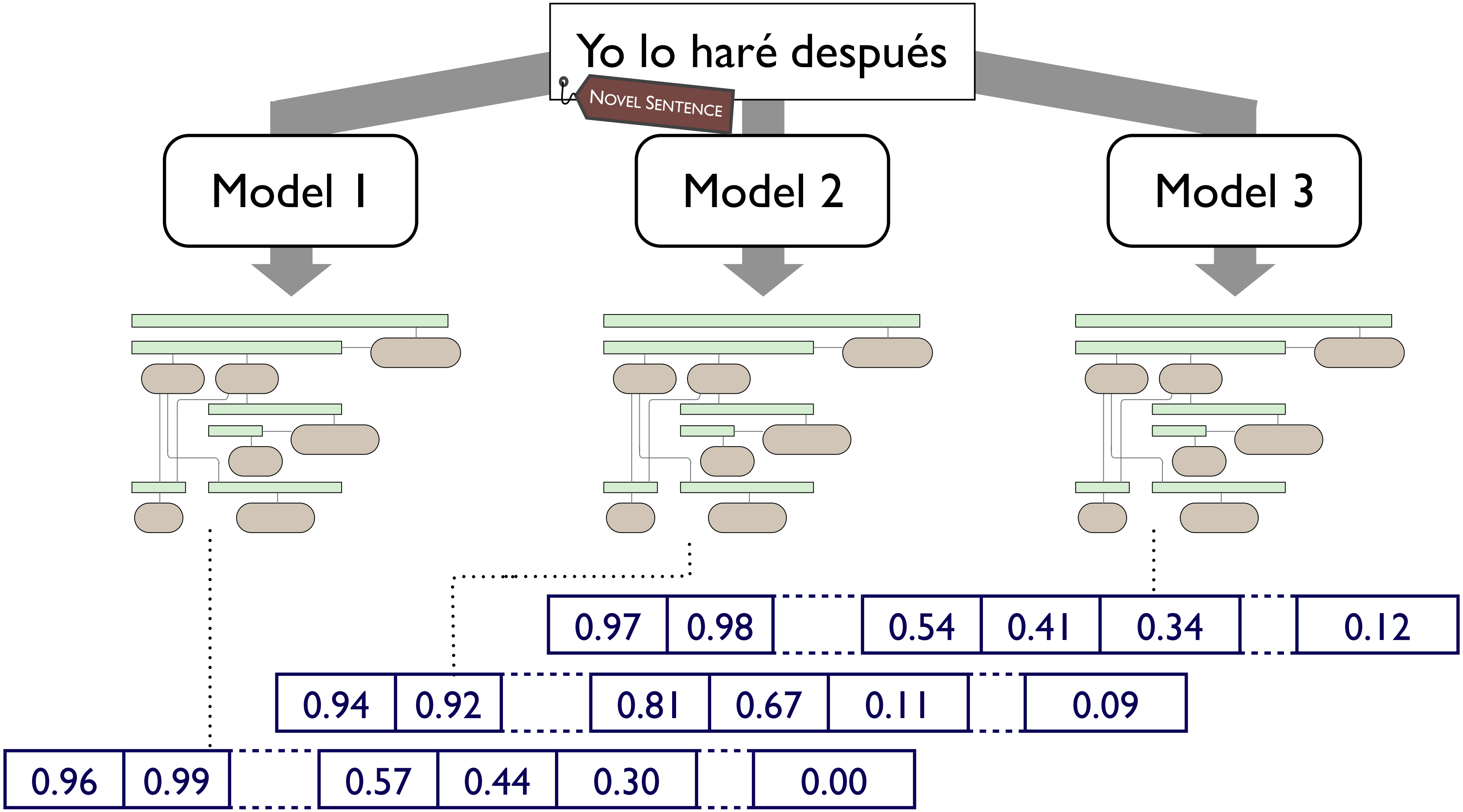
Translating Using Multiple Systems



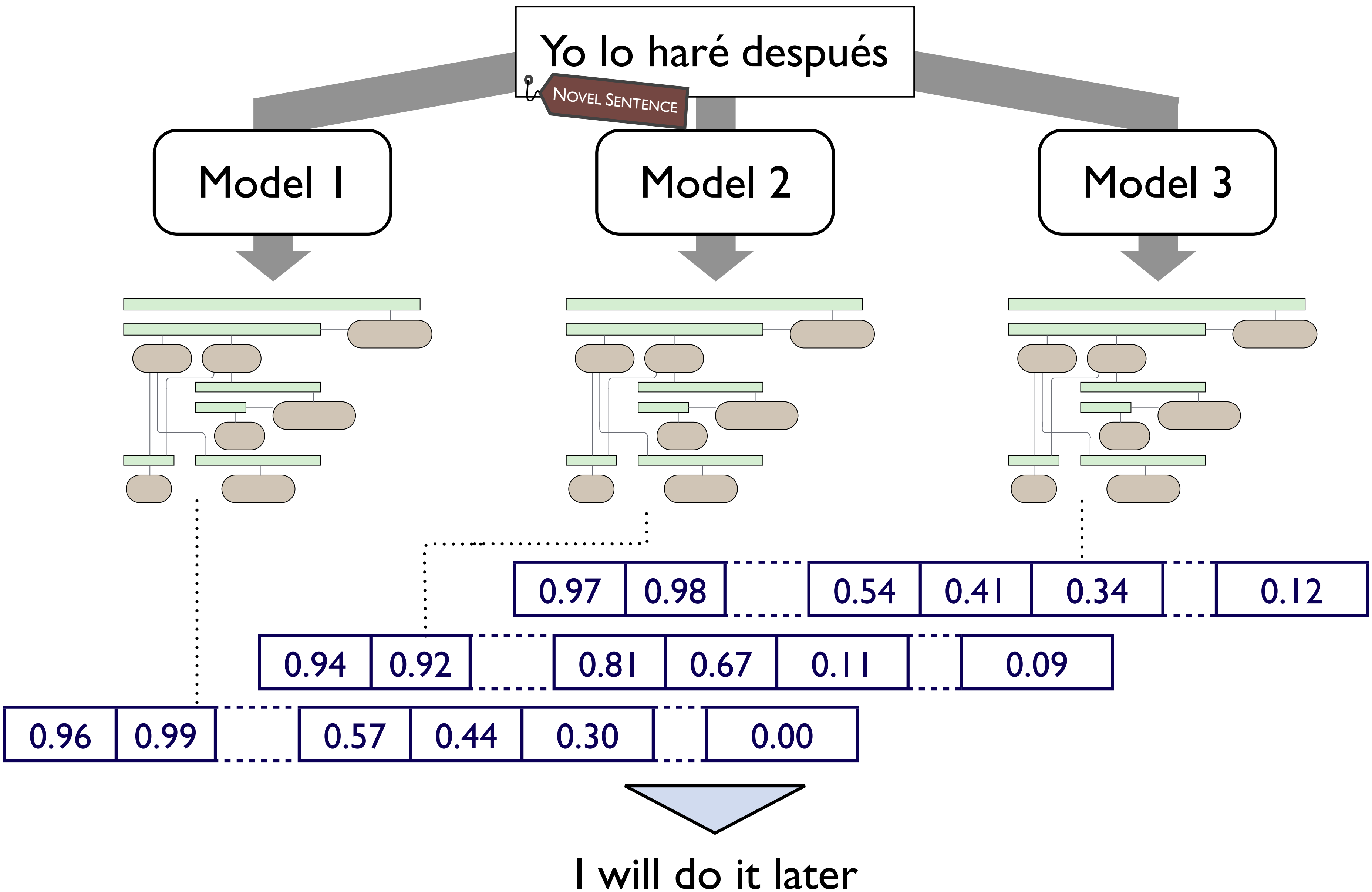
Translating Using Multiple Systems



Translating Using Multiple Systems



Translating Using Multiple Systems



Consensus Modeling FAQ

Consensus Modeling FAQ

Q: How do we combine different models?

Consensus Modeling FAQ

Q: How do we combine different models?

A: Train a linear consensus model scoring a derivation *d*:

$$\underbrace{\sum_{i=1}^{I}}_{\text{Models}} \left[\underbrace{w_i^{(\alpha)} \alpha_i(d)}_{\text{Which model?}} + \underbrace{\sum_{n=1}^{4}}_{\text{Phrase lengths}} \underbrace{w_i^{(n)} v_i^{(n)}(d)}_{\text{Expected counts}} \right] + \underbrace{w^{(b)} \cdot b(d)}_{\text{Model score}} + \underbrace{w^{(\ell)} \cdot \ell(d)}_{\text{Length}}$$

Consensus Modeling FAQ

Q: How do we combine different models?

A: Train a linear consensus model scoring a derivation *d*:

$$\underbrace{\sum_{i=1}^{I}}_{\text{Models}} \left[\underbrace{w_i^{(\alpha)} \alpha_i(d)}_{\text{Which model?}} + \underbrace{\sum_{n=1}^{4}}_{\text{Phrase lengths}} \underbrace{w_i^{(n)} v_i^{(n)}(d)}_{\text{Expected counts}} \right] + \underbrace{w^{(b)} \cdot b(d)}_{\text{Model score}} + \underbrace{w^{(\ell)} \cdot \ell(d)}_{\text{Length}}$$

Q: What output sentences are considered?

Consensus Modeling FAQ

Q: How do we combine different models?

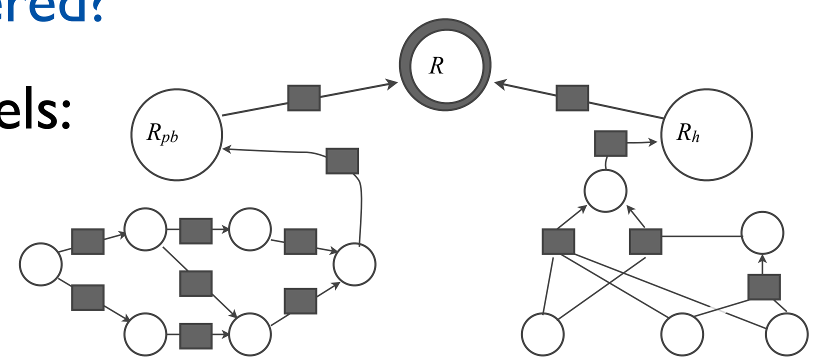
A: Train a linear consensus model scoring a derivation *d*:

$$\sum_{i=1}^{I}\left[w_i^{(\alpha)}\alpha_i(d)+\sum_{n=1}^{4}w_i^{(n)}v_i^{(n)}(d)\right]+w^{(b)}\cdot b(d)+w^{(\ell)}\cdot \ell(d)$$

Models | Which model? | Phrase lengths | Expected counts | Model score | Length

Q: What output sentences are considered?

A: The union of output spaces of models:

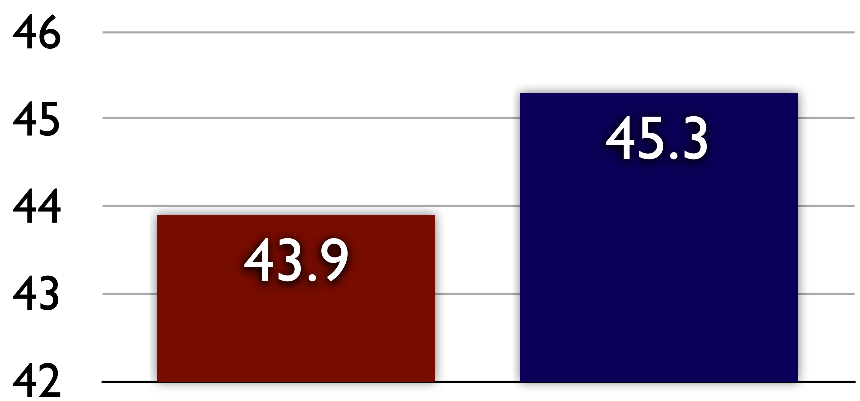


Multi-System Translation Results

Google's Full-Scale Research Translation System for Arabic-to-English

- Best Single-System Model-Only Baseline
- Multi-System Forest-Based Consensus [DeNero et al. NAACL '10]*

Translation quality (BLEU)



* John DeNero, Shankar Kumar, Ciprian Chelba, and Franz Och. *Model Combination for Machine Translation*, NAACL 2010.

The Steps in a Modern Translation System

Learn a model

Apply the model

Choose a translation

The Steps in a Modern Translation System

Learn a model

Apply the model

Choose a translation

- Statistical models provide distributions over outputs

The Steps in a Modern Translation System

Learn a model

Apply the model

Choose a translation

- Statistical models provide distributions over outputs
- Leveraging those distributions improves performance

The Steps in a Modern Translation System

Learn a model

Apply the model

Choose a translation

- Statistical models provide distributions over outputs
- Leveraging those distributions improves performance
- Compact representations can enable large-scale computation

Summary of Translation Research

Learn a model

Apply the model

Choose a translation

Summary of Translation Research

- Large-context models
- Non-parametric models

Learn a model	Apply the model	Choose a translation

[DeNero et al. EMNLP ’08]

[DeNero & Klein. ACL ’10]

Summary of Translation Research

- Large-context models
- Non-parametric models
- Coarse-to-fine

Learn a model	Apply the model	Choose a translation
[DeNero et al. EMNLP '08]	[DeNero et al. NAACL '09]	
[DeNero & Klein. ACL '10]		

Summary of Translation Research

- Large-context models
- Non-parametric models

Learn a model

[DeNero et al. EMNLP ’08]

[DeNero & Klein. ACL ’10]

- Coarse-to-fine

Apply the model

[DeNero et al. NAACL ’09]

- Full distributions
- Compact encodings

Choose a translation

[DeNero et al. ACL ’09]

[DeNero et al. NAACL ’10]

Summary of Translation Research

- Large-context models
- Non-parametric models

- Coarse-to-fine

- Full distributions
- Compact encodings

Learn a model

Apply the model

Choose a translation

[DeNero et al. EMNLP '08]

[DeNero & Klein. ACL '10]

[DeNero et al. NAACL '09]

[DeNero et al. ACL '09]

[DeNero et al. NAACL '10]

Are we done yet?

Summary of Translation Research

- Large-context models
- Non-parametric models

- Coarse-to-fine

- Full distributions
- Compact encodings

Learn a model

Apply the model

Choose a translation

[DeNero et al. EMNLP '08]

[DeNero & Klein. ACL '10]

[DeNero et al. NAACL '09]

[DeNero et al. ACL '09]

[DeNero et al. NAACL '10]

Are we done yet?

- Morphology in alignment modeling

Summary of Translation Research

- Large-context models
- Non-parametric models

Learn a model

[DeNero et al. EMNLP '08]

[DeNero & Klein. ACL '10]

- Coarse-to-fine

Apply the model

[DeNero et al. NAACL '09]

- Full distributions
- Compact encodings

Choose a translation

[DeNero et al. ACL '09]

[DeNero et al. NAACL '10]

Are we done yet?

- Morphology in alignment modeling
- Unsupervised composed phrase learning

Summary of Translation Research

- Large-context models
- Non-parametric models

Learn a model

[DeNero et al. EMNLP '08]

[DeNero & Klein. ACL '10]

- Coarse-to-fine

Apply the model

[DeNero et al. NAACL '09]

- Full distributions
- Compact encodings

Choose a translation

[DeNero et al. ACL '09]

[DeNero et al. NAACL '10]

Are we done yet?

- Morphology in alignment modeling
- Unsupervised composed phrase learning
- Adding information to consensus models

Acknowledgements

John Thank you!

Juan Gracias!

and many thanks to my excellent coauthors on this work:

Berkeley: Mohit Bansal, John Blitzer, Alex Bouchard-Côté, Aria Haghighi, Dan Klein, and Adam Pauls

Information Sciences Institute: David Chiang and Kevin Knight

Google: Ciprian Chelba, Shankar Kumar, and Franz Och