Sampling Alignment Structure under a Bayesian Translation Model

University of California
Berkeley

John DeNero
Alexandre Bouchard-Côté
Dan Klein

Training a Machine Translation System

Input:

Gracias , lo haré de muy buen grado .
Thank you , I shall do so gladly .

Training a Machine Translation System

Input:

Gracias , lo haré de muy buen grado .
Thank you , I shall do so gladly .

First, we learn word alignments,

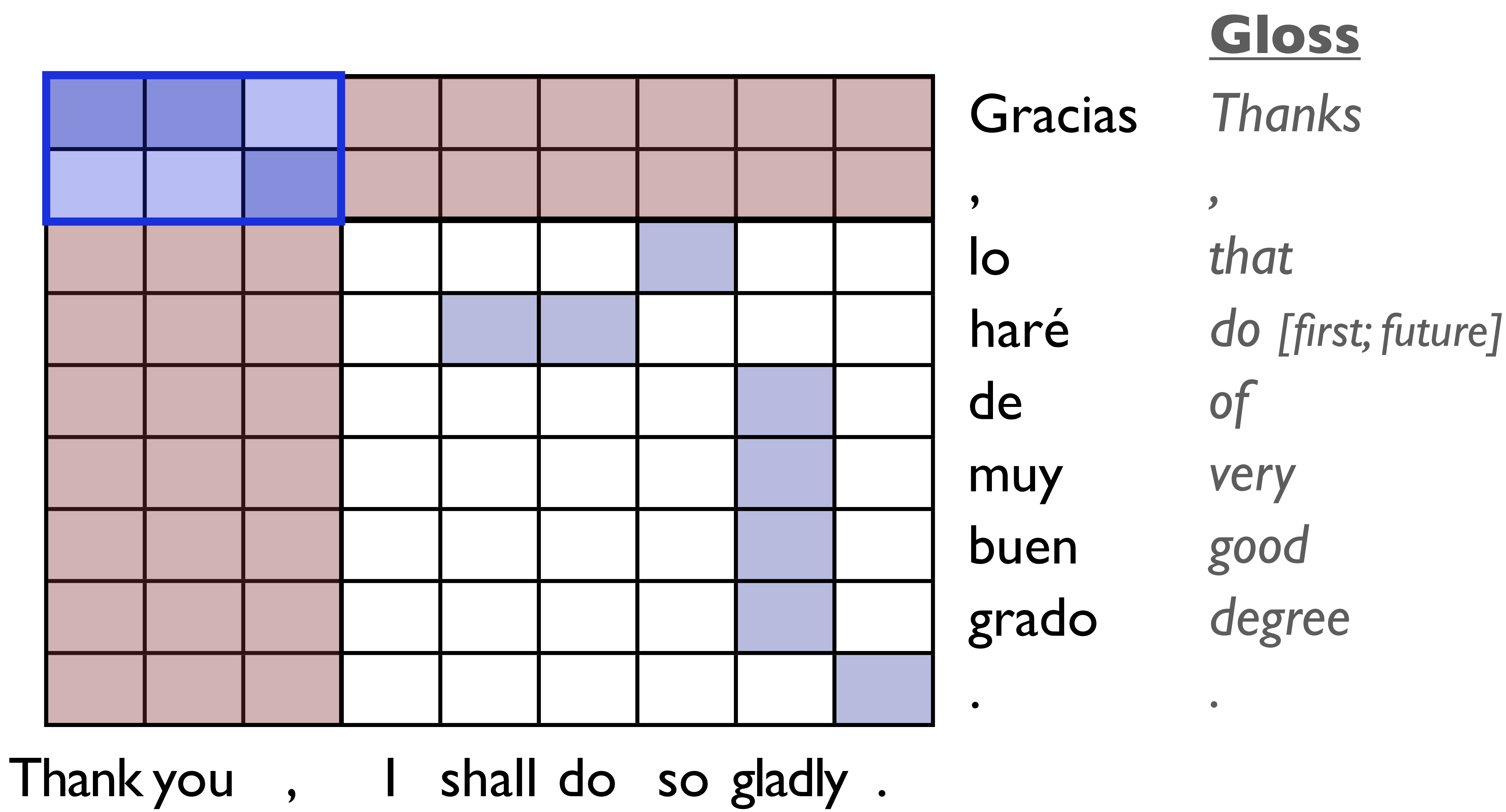
Training a Machine Translation System

Input:

Gracias , lo haré de muy buen grado .
Thank you , I shall do so gladly .

First, we learn word alignments,

then we infer aligned phrases.



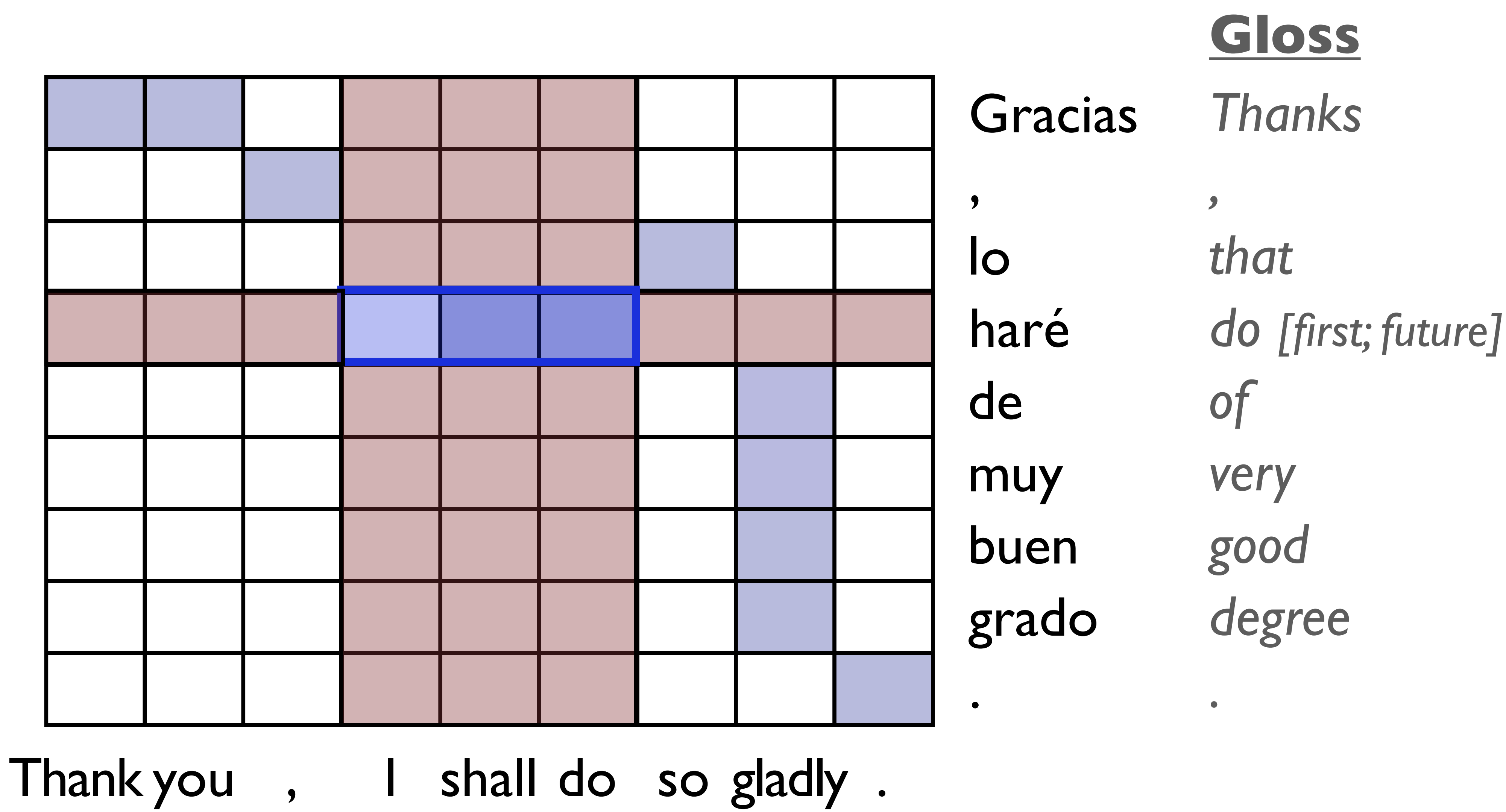
Training a Machine Translation System

Input:

Gracias , lo haré de muy buen grado .
Thank you , I shall do so gladly .

First, we learn word alignments,

then we infer aligned phrases.

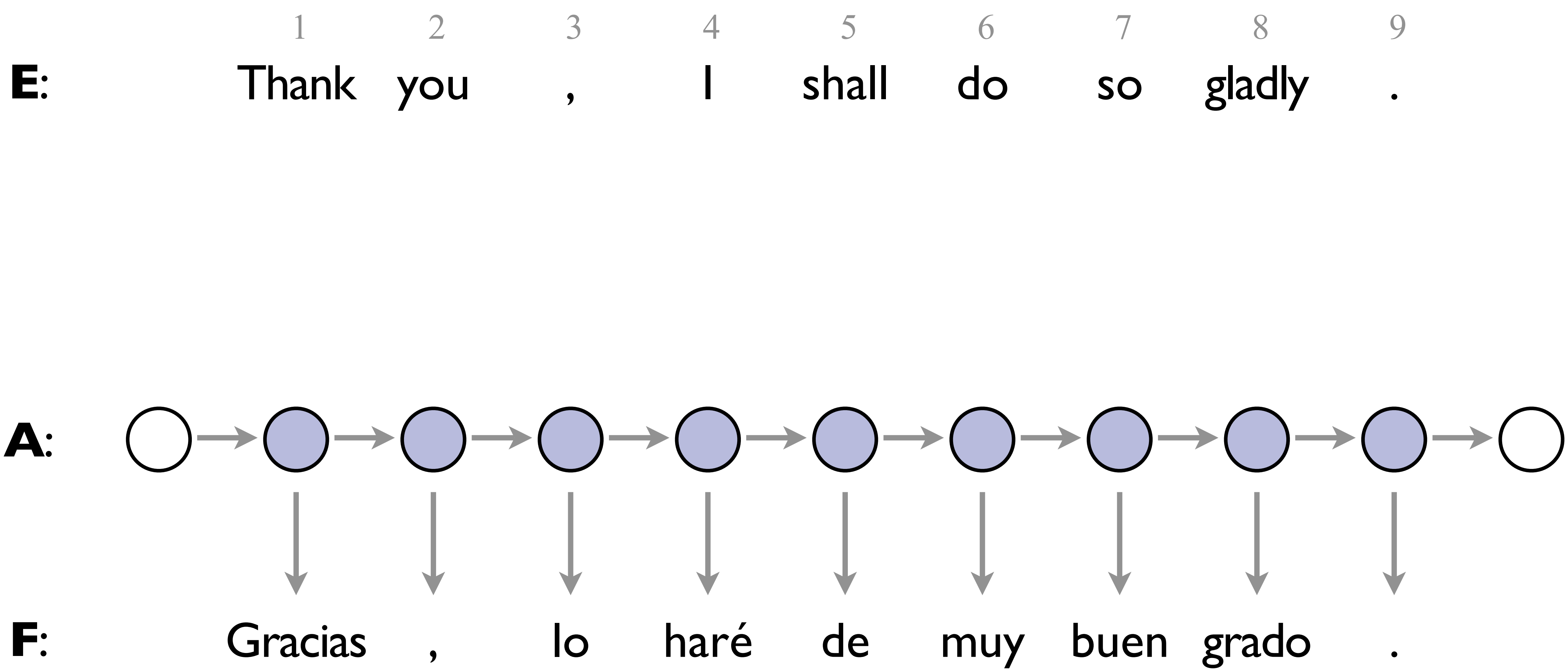


Statistical Word Alignment

E: Thank you , I shall do so gladly .

F: Gracias , lo haré de muy buen grado .

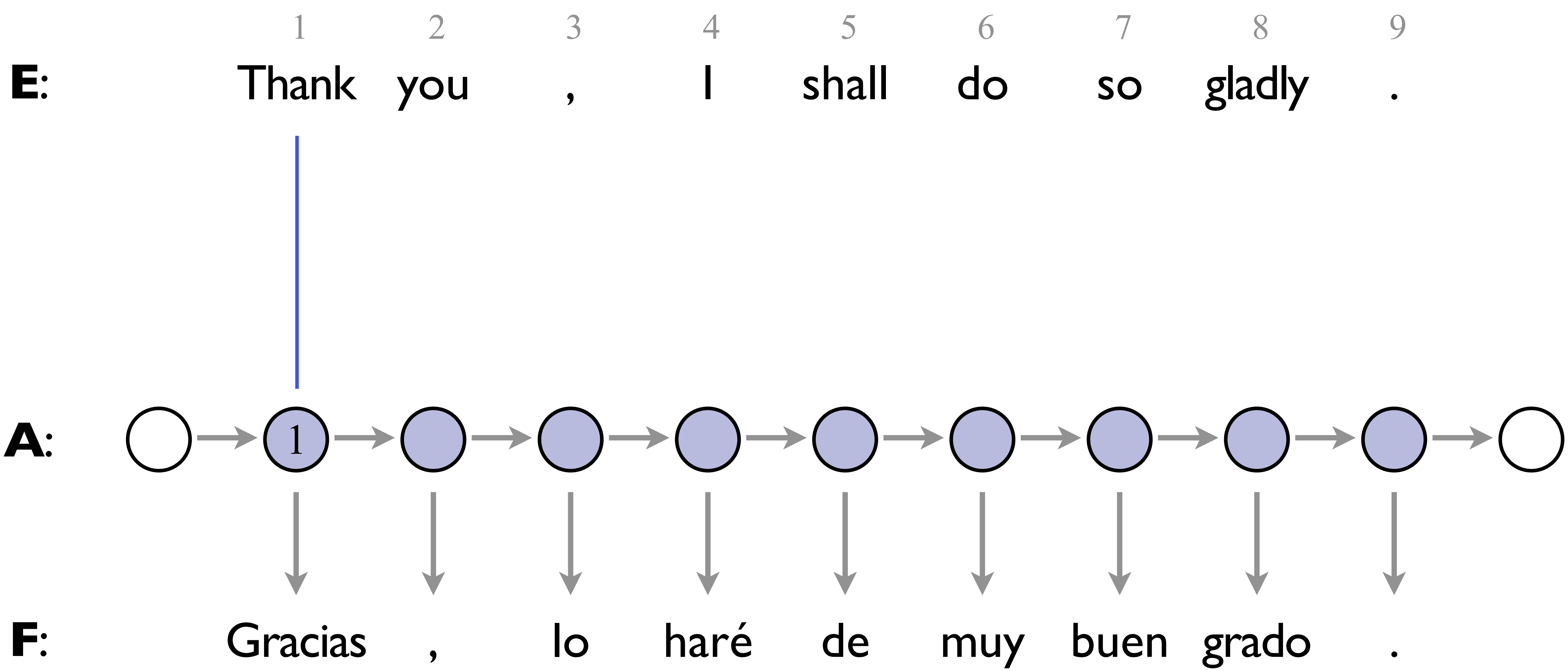
Statistical Word Alignment



Model Parameters

Emissions: $P(F_1 = \text{Gracias} \mid E_{A_1} = \text{Thank})$ *Transitions:* $P(A_2 = 3 \mid A_1 = 1)$

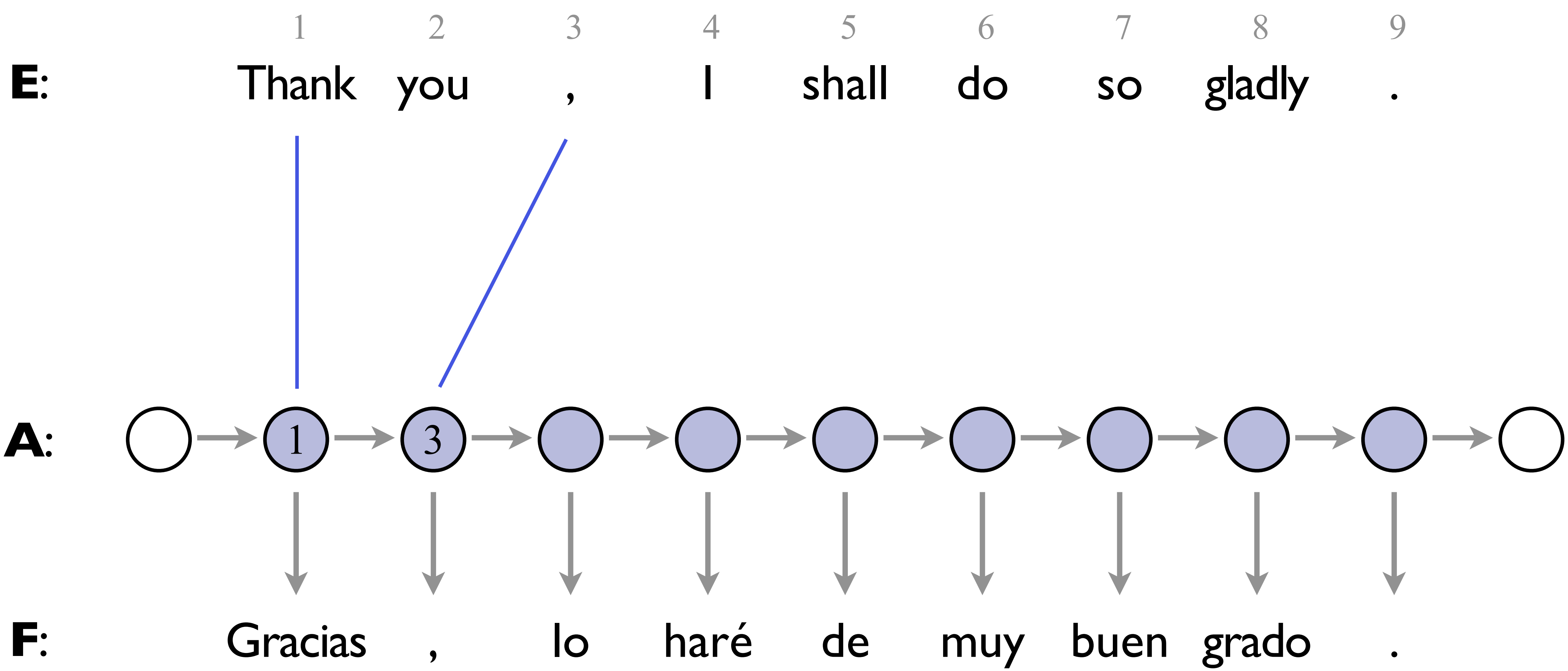
Statistical Word Alignment



Model Parameters

Emissions: $P(F_1 = \text{Gracias} \mid E_{A_1} = \text{Thank})$ *Transitions:* $P(A_2 = 3 \mid A_1 = 1)$

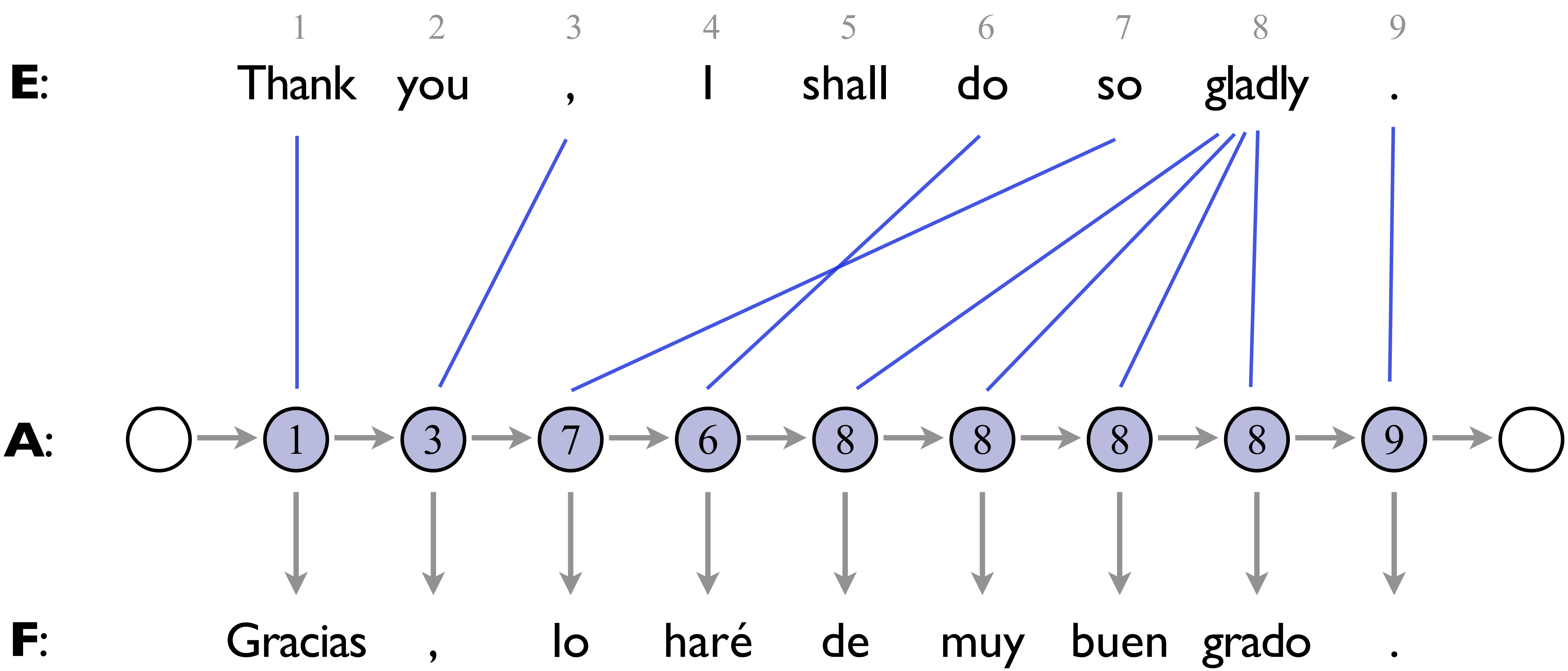
Statistical Word Alignment



Model Parameters

Emissions: $P(F_1 = \text{Gracias} \mid E_{A_1} = \text{Thank})$ *Transitions:* $P(A_2 = 3 \mid A_1 = 1)$

Statistical Word Alignment

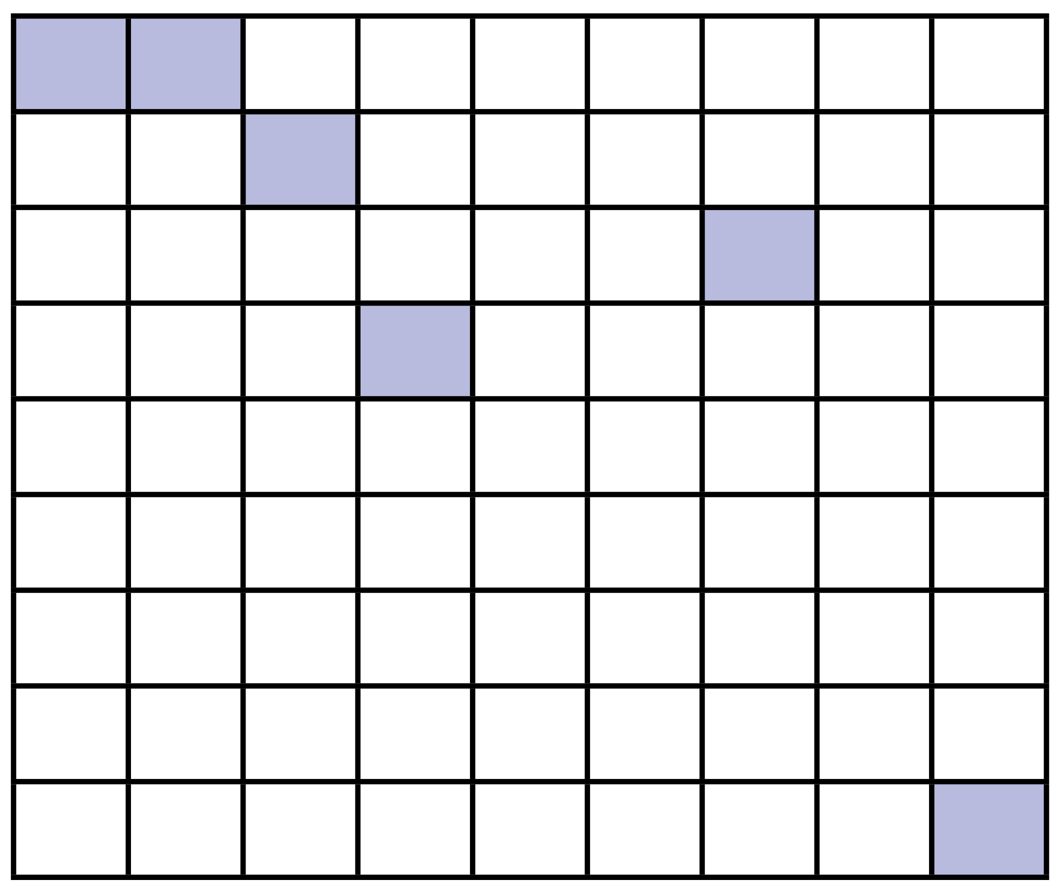


Model Parameters

Emissions: $P(F_1 = \text{Gracias} \mid E_{A_1} = \text{Thank})$ *Transitions:* $P(A_2 = 3 \mid A_1 = 1)$

What Happens in Practice

A real word alignment
(GIZA++ Model 4 with
grow-diag-final combination)

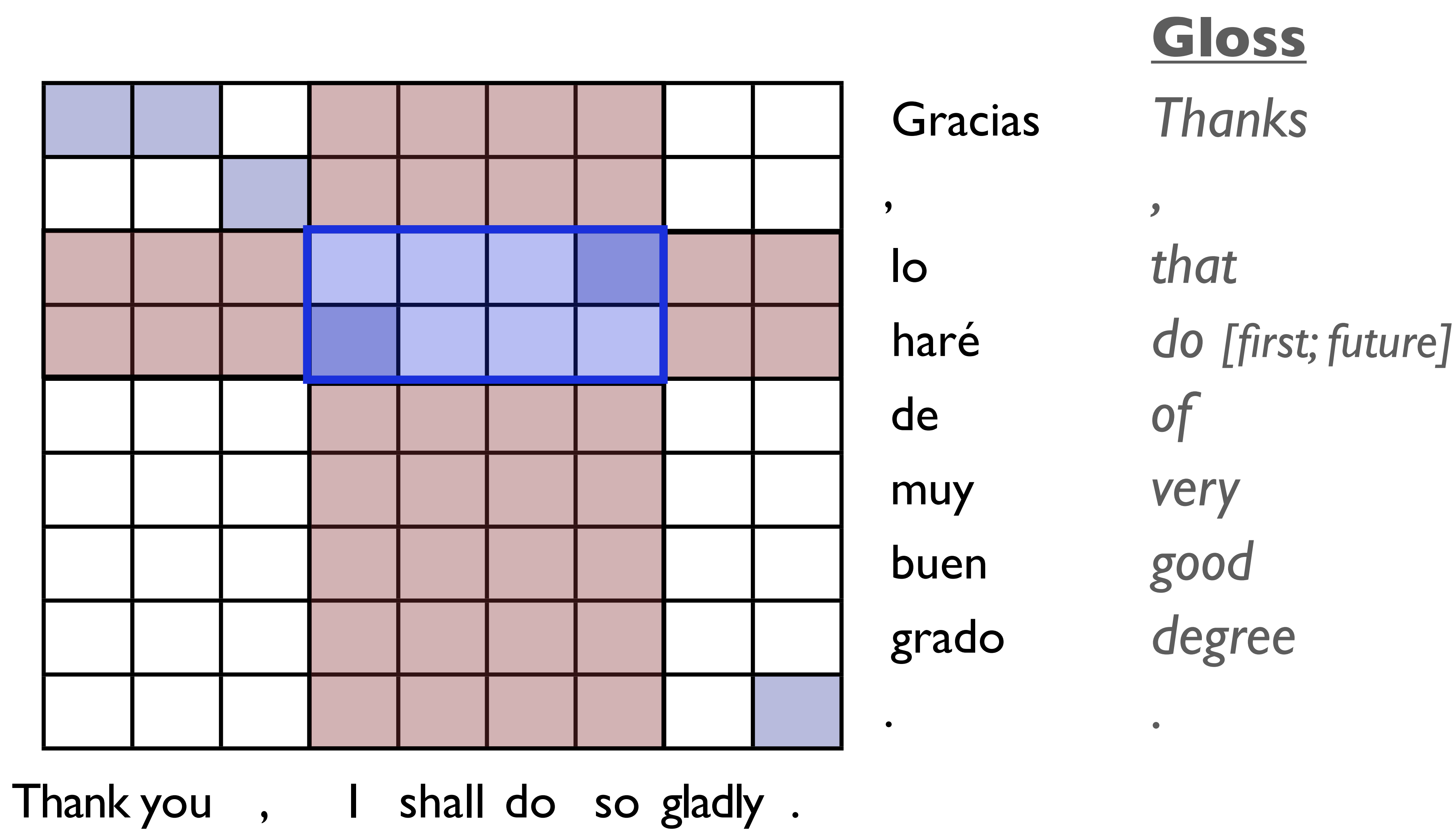


Gloss

Thanks
,
that
do [first; future]
of
very
good
degree
.

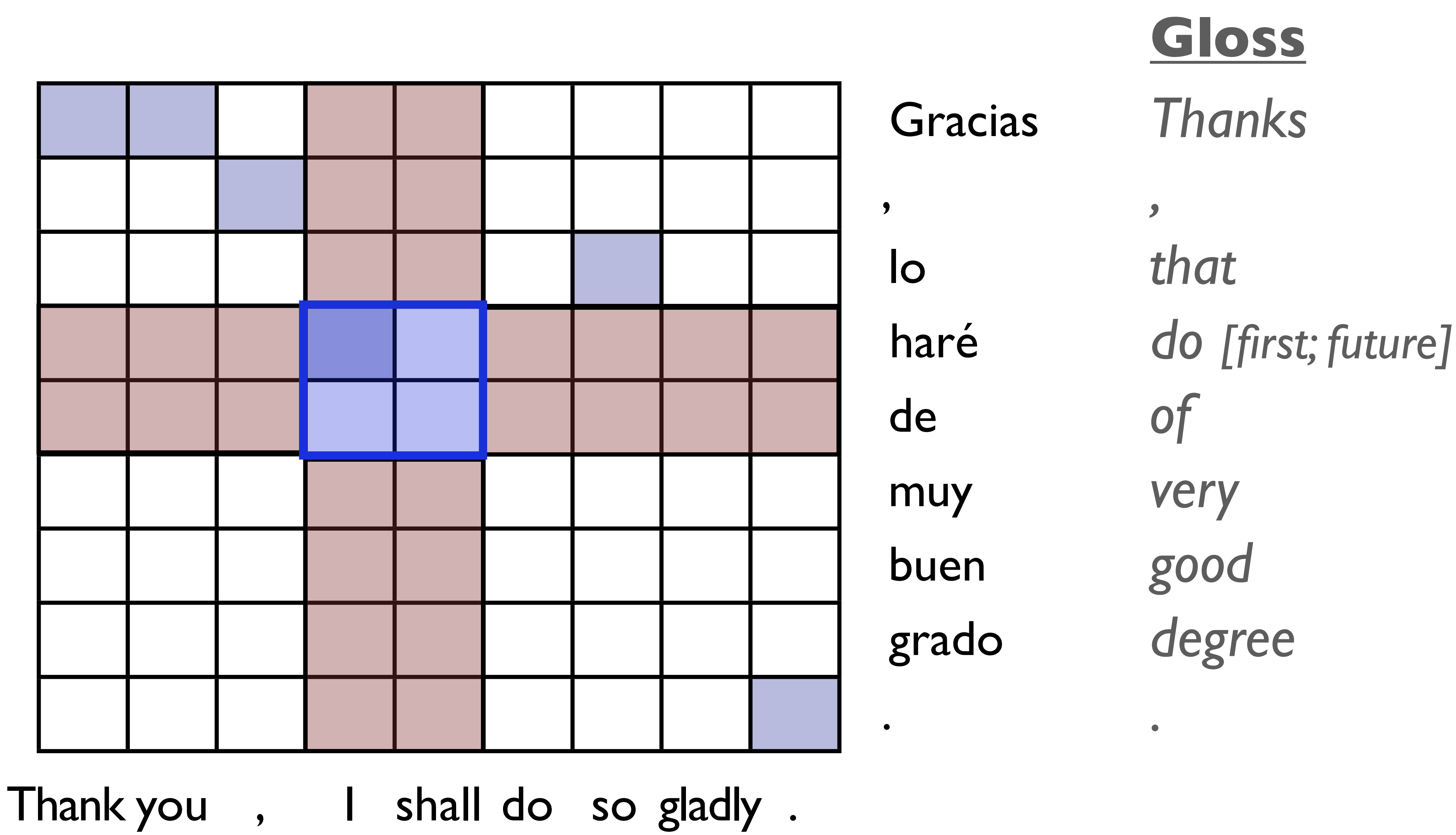
What Happens in Practice

A real word alignment
(GIZA++ Model 4 with
grow-diag-final combination)



What Happens in Practice

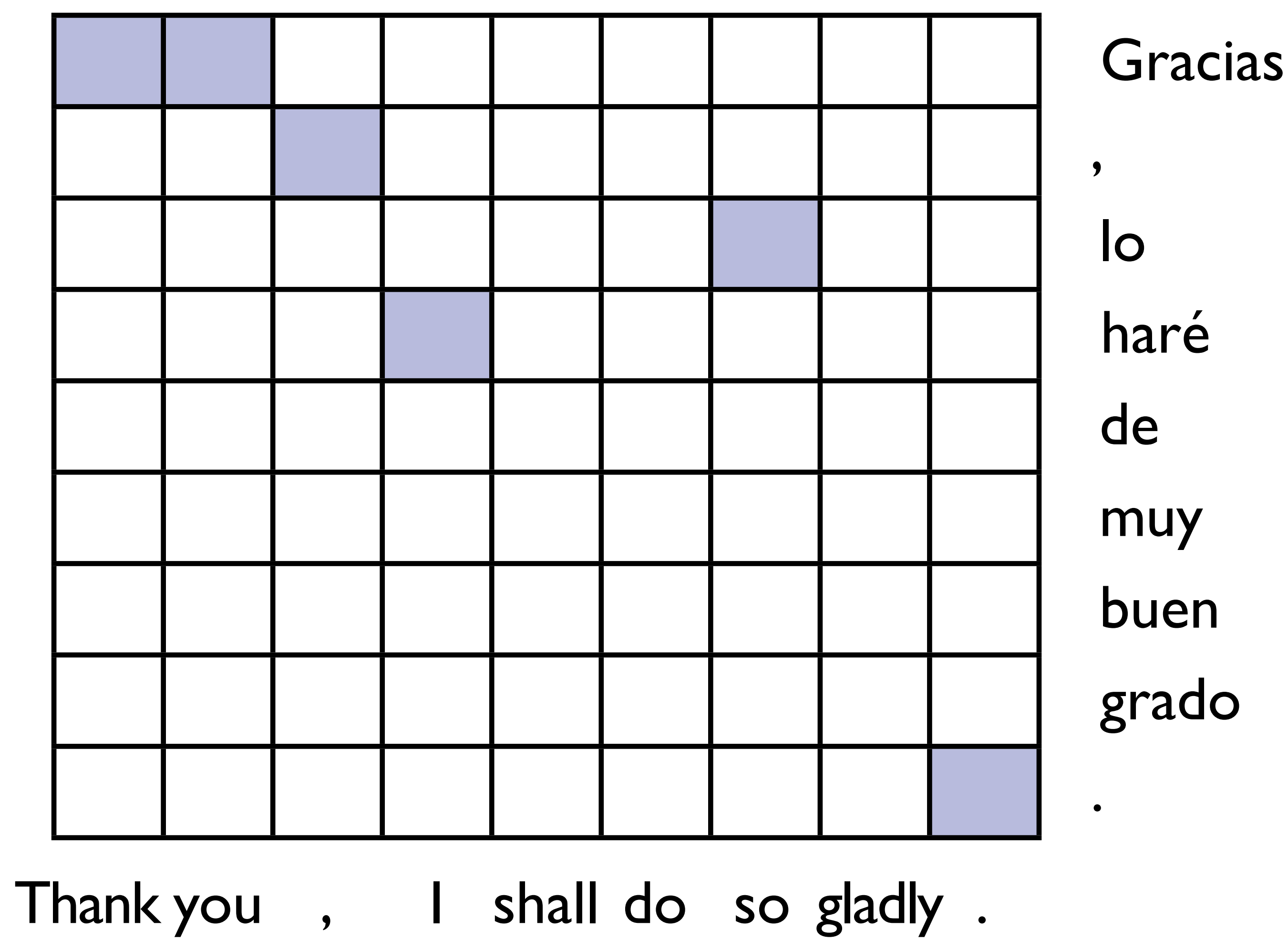
A real word alignment
(GIZA++ Model 4 with
grow-diag-final combination)



What Happens in Practice

A real word alignment

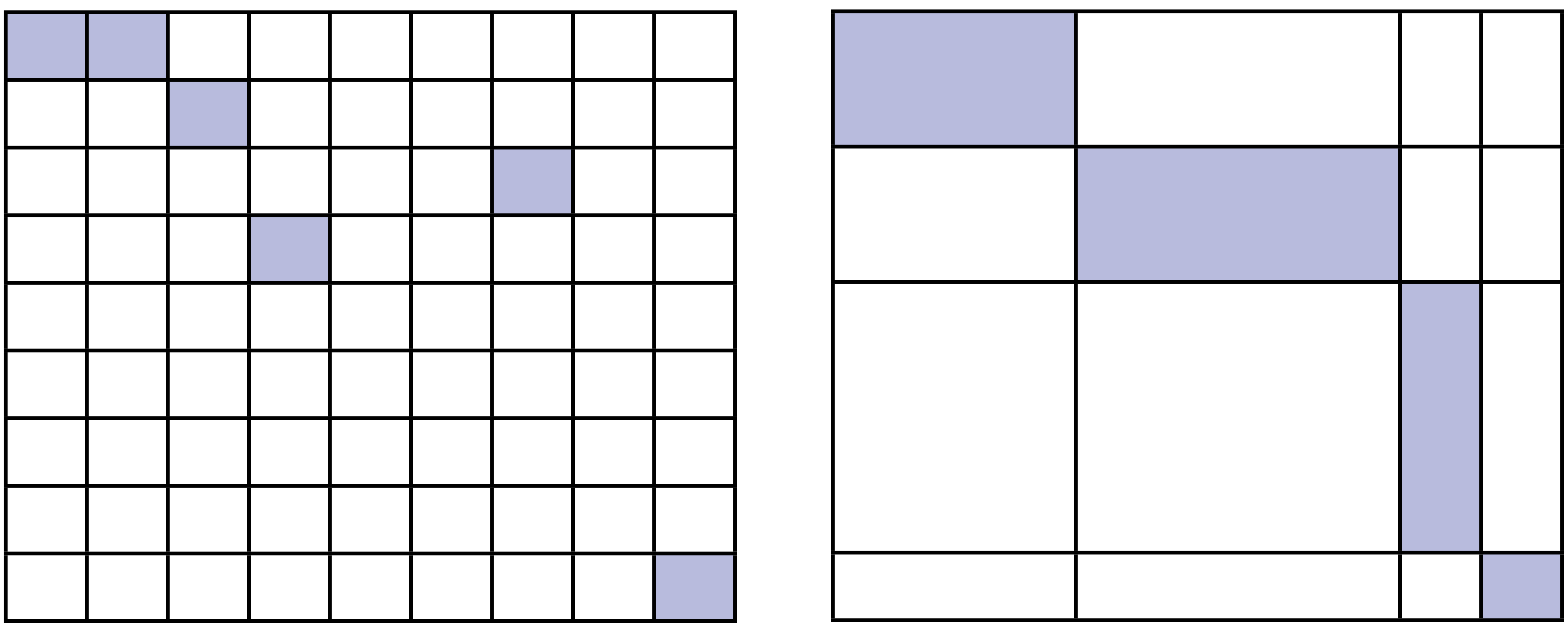
(GIZA++ Model 4 with
grow-diag-final combination)



What Happens in Practice

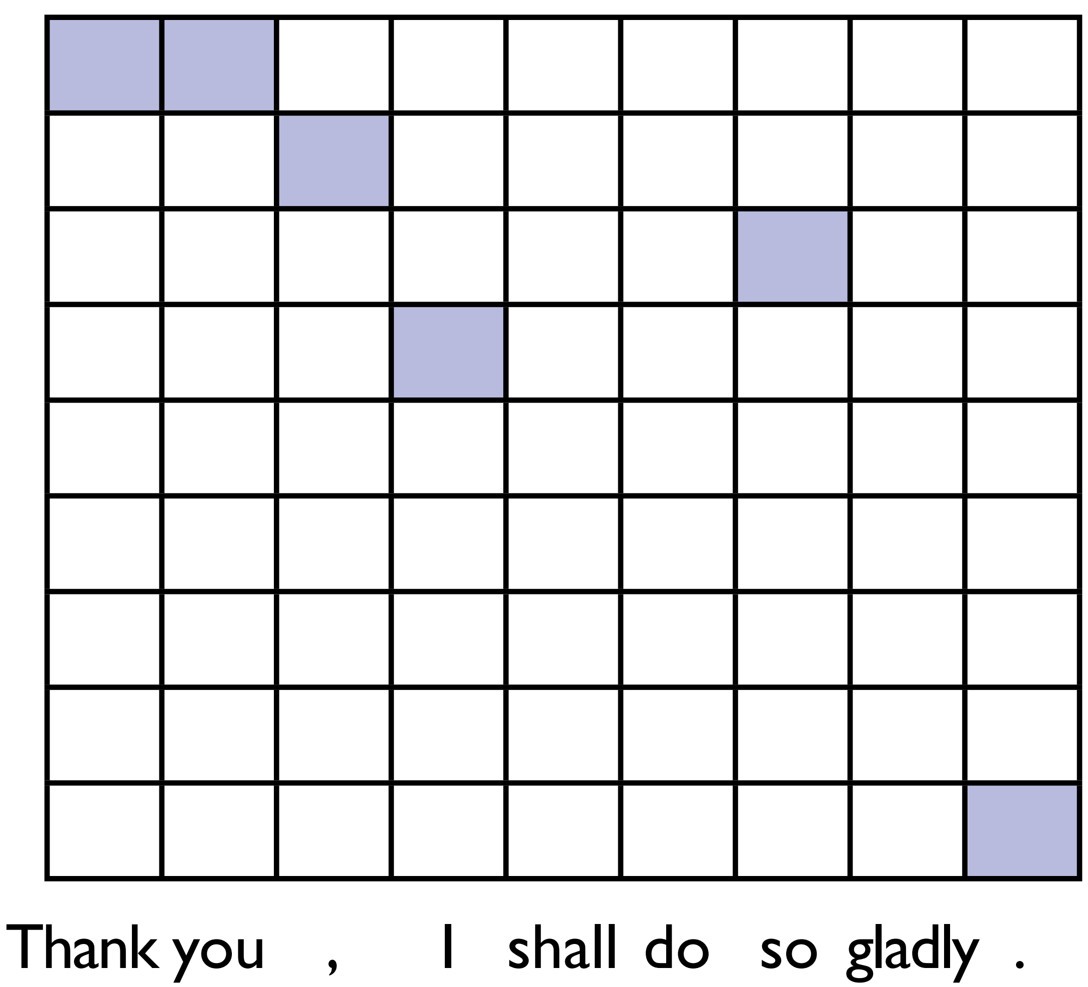
A real word alignment
(GIZA++ Model 4 with grow-diag-final combination)

A sampled phrase alignment
(our system)

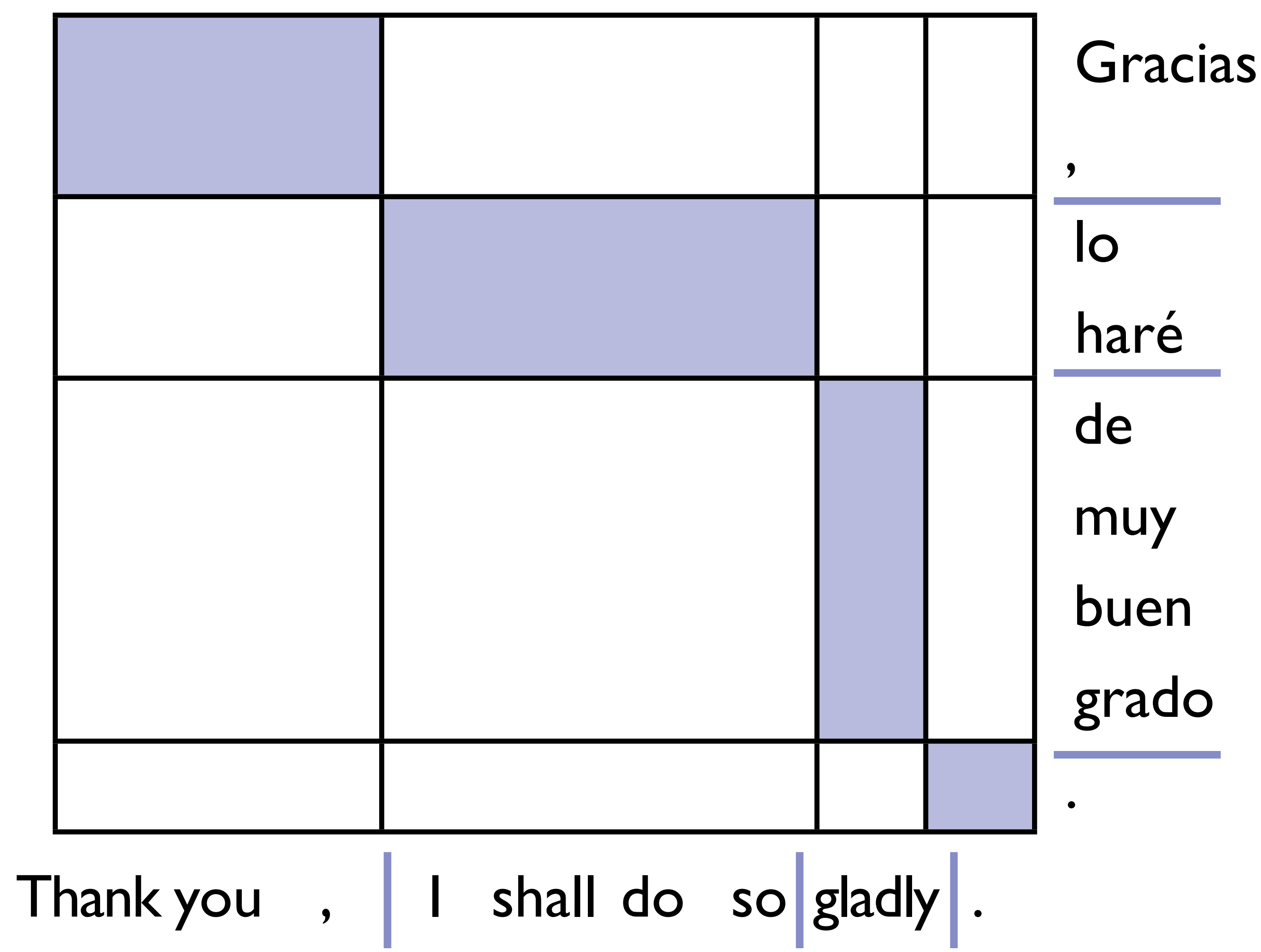


What Happens in Practice

A real word alignment
(GIZA++ Model 4 with grow-diag-final combination)



A sampled phrase alignment
(our system)



A Brief History of Phrase Alignment

The Challenge:

Train models that explicitly align phrases, not just words

A Brief History of Phrase Alignment

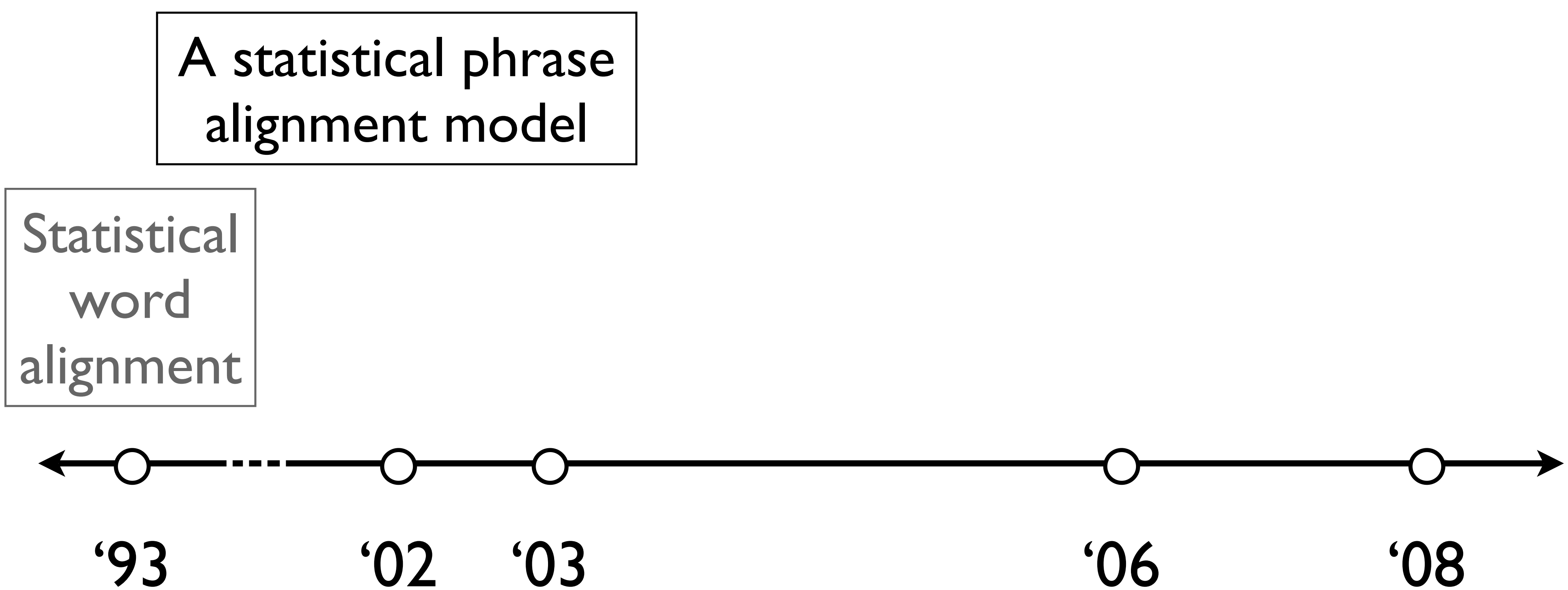
The Challenge:

Train models that explicitly align phrases, not just words

A Brief History of Phrase Alignment

The Challenge:

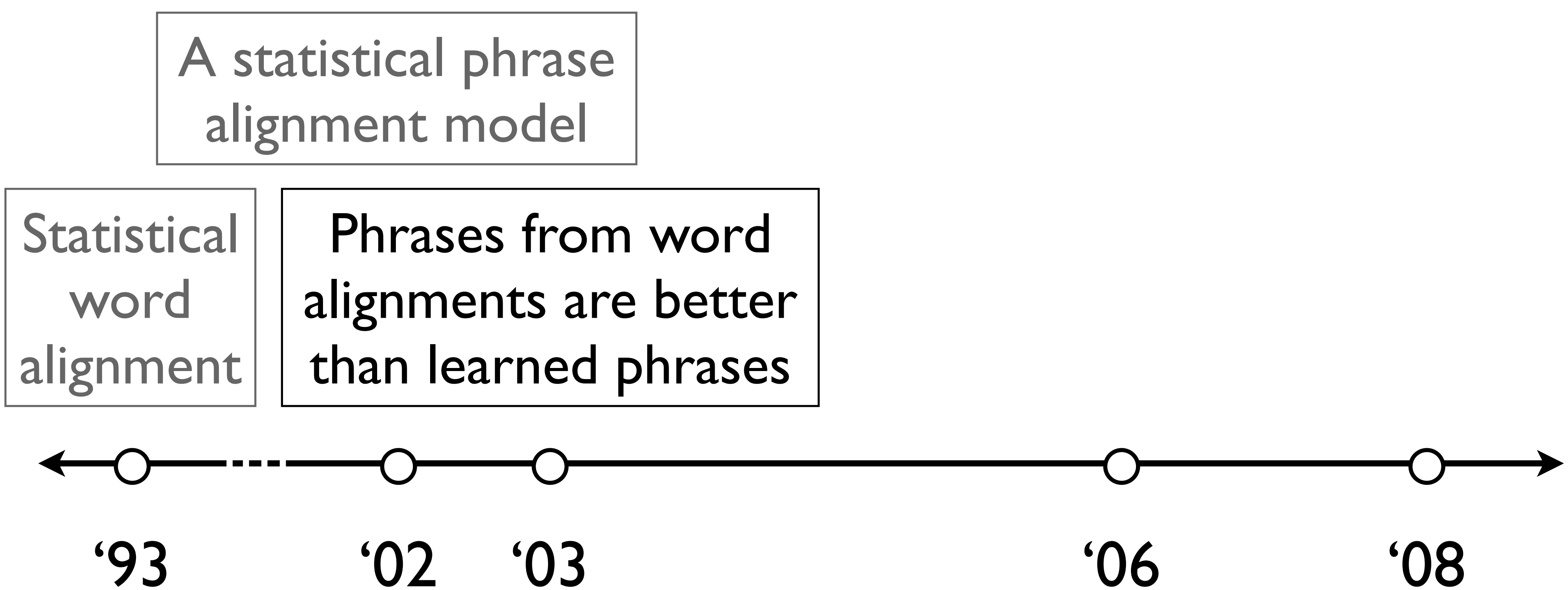
Train models that explicitly align phrases, not just words



A Brief History of Phrase Alignment

The Challenge:

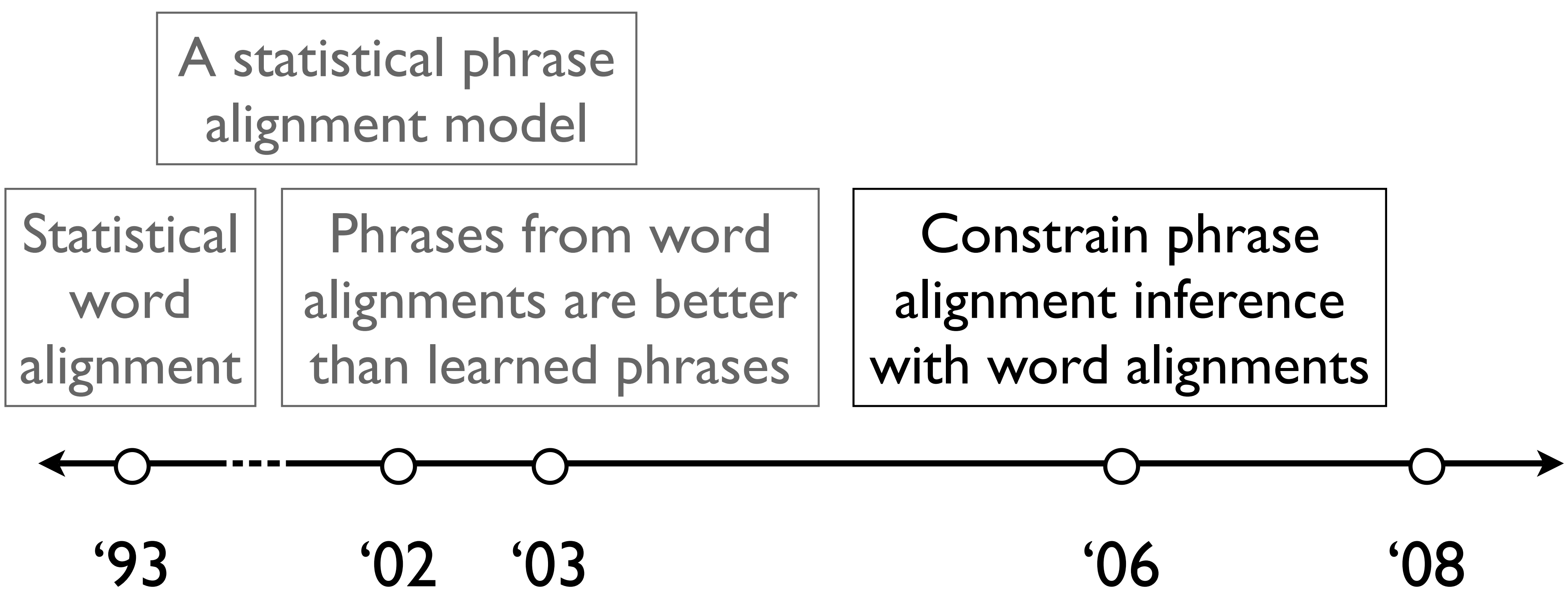
Train models that explicitly align phrases, not just words



A Brief History of Phrase Alignment

The Challenge:

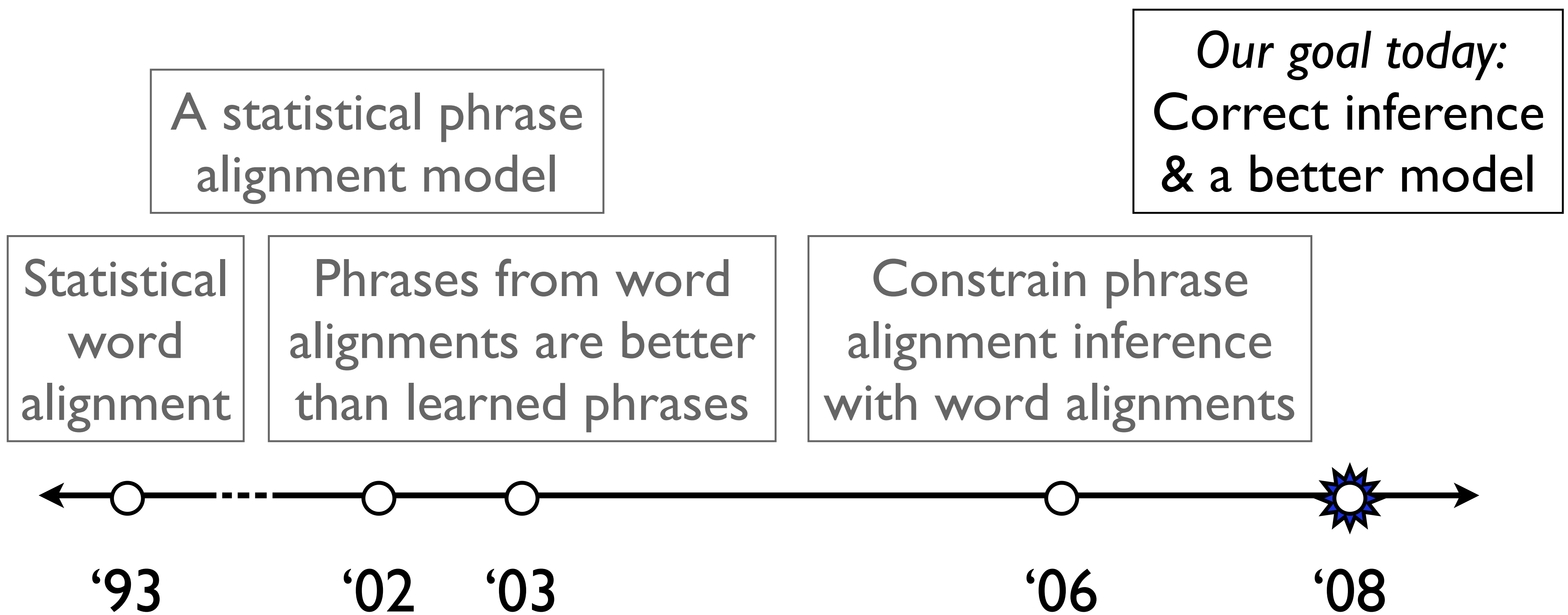
Train models that explicitly align phrases, not just words



A Brief History of Phrase Alignment

The Challenge:

Train models that explicitly align phrases, not just words



A Generative Phrase Alignment Model

Process for generating a sentence pair:

A Generative Phrase Alignment Model

Process for generating a sentence pair:

Choose number of phrase pairs

A Generative Phrase Alignment Model

Process for generating a sentence pair:

Choose number of phrase pairs

Generate each phrase pair

日本 冻结	提供 援助	向 俄
Japan to freeze	aid	to Russia

A Generative Phrase Alignment Model

Process for generating a sentence pair:

Choose number of phrase pairs

Generate each phrase pair

Keep the English order

日本 冻结	提供 援助	向 俄
Japan to freeze	aid	to Russia

Japan to freeze aid to Russia

A Generative Phrase Alignment Model

Process for generating a sentence pair:

Choose number of phrase pairs

Generate each phrase pair

Keep the English order

Reorder the Chinese phrases

日本 冻结	提供 援助	向 俄
Japan to freeze	aid	to Russia

Japan to freeze aid to Russia

A Generative Phrase Alignment Model

Process for generating a sentence pair:

Choose number of phrase pairs

Generate each phrase pair

Keep the English order

Reorder the Chinese phrases

日本 冻结	提供 援助	向 俄
Japan to freeze	aid	to Russia

日本 冻结 向 俄 提供 援助

Japan to freeze aid to Russia

A Generative Phrase Alignment Model

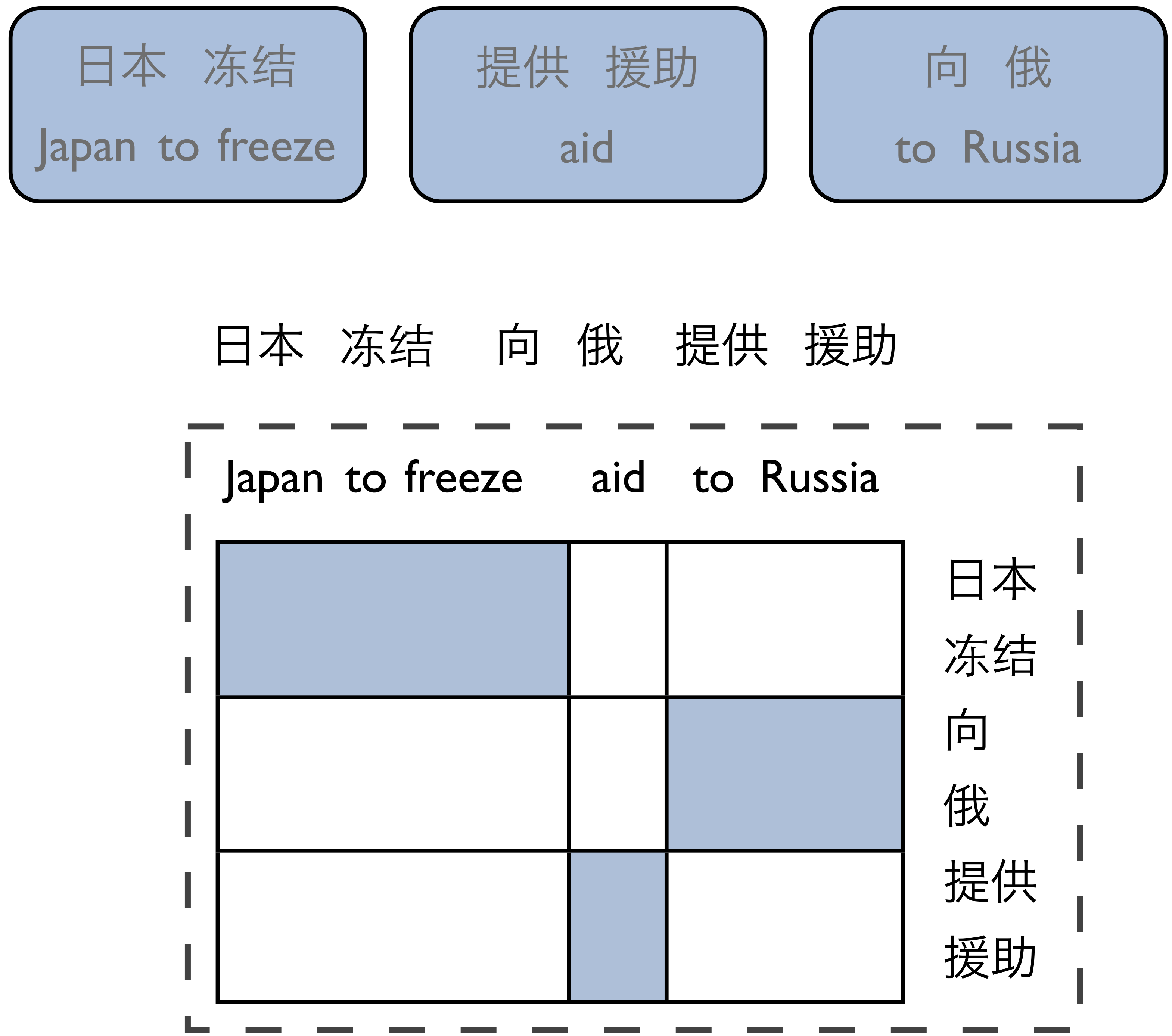
Process for generating a sentence pair:

Choose number of phrase pairs

Generate each phrase pair

Keep the English order

Reorder the Chinese phrases



A phrase-aligned sentence pair
(a list of phrase pairs and a permutation)

Estimating Model Parameters

A phrase-aligned sentence pair a *is*

a list of phrase pairs $\langle e, f \rangle$ *and a permutation* σ_a .

Estimating Model Parameters

A phrase-aligned sentence pair a *is*
a list of phrase pairs $\langle e, f \rangle$ *and a permutation* σ_a *.*

$$\mathrm{P}(a) = \prod_{\langle e,f \rangle \in a} \phi(\langle e, f \rangle) \cdot \delta(\sigma_a)$$

Estimating Model Parameters

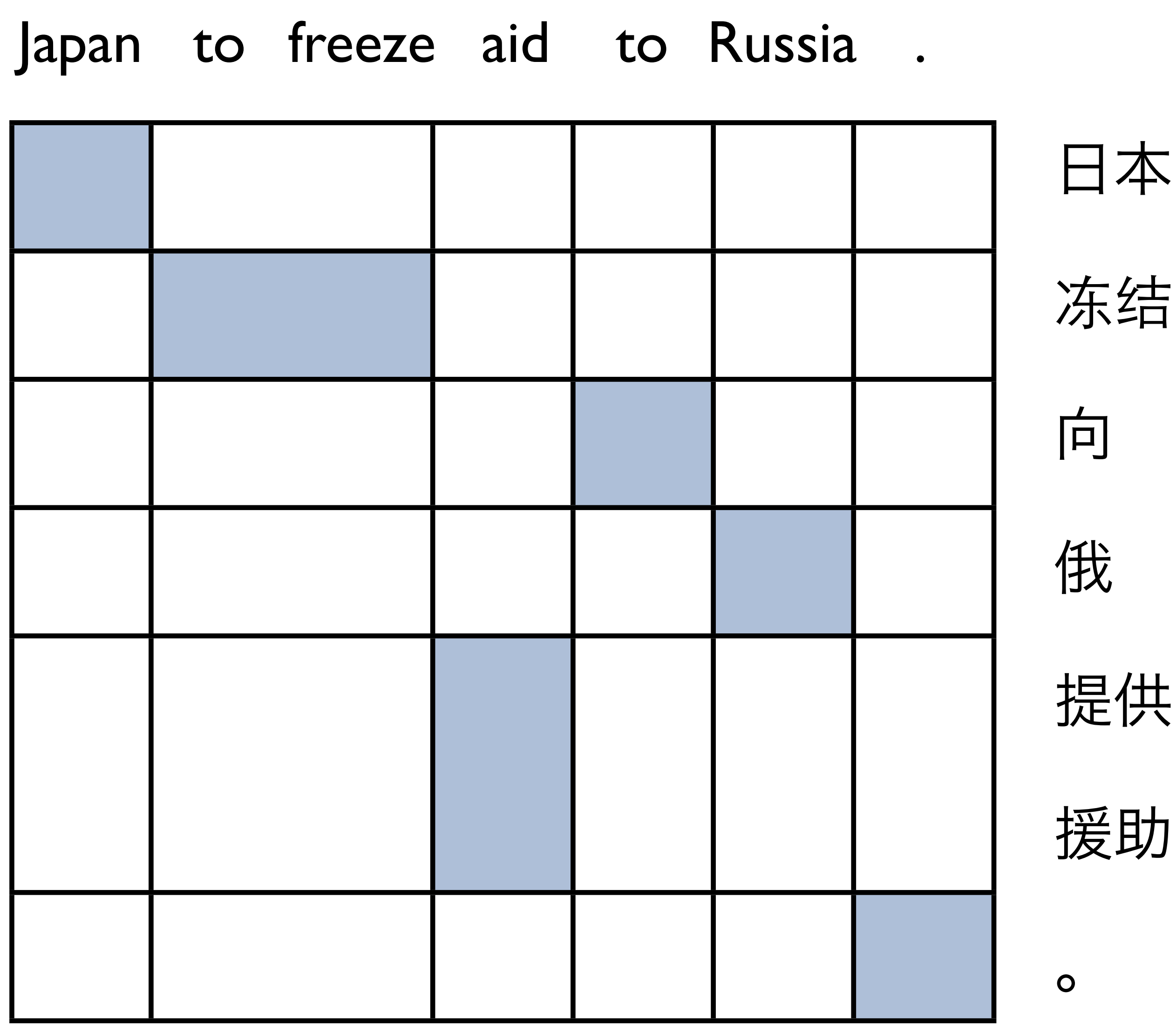
A phrase-aligned sentence pair a *is*
a list of phrase pairs $\langle e, f \rangle$ *and a permutation* σ_a *.*

$$\mathrm{P}(a) = \prod_{\langle e,f \rangle \in a} \phi(\langle e, f \rangle) \cdot \delta(\sigma_a)$$

Maximum likelihood:
choose ϕ *to maximize the probability of the training corpus:*

$$\max_{\phi} \prod_{(\mathbf{e},\mathbf{f})} \left[\sum_{a \text{ for } (\mathbf{e},\mathbf{f})} \prod_{\langle e,f \rangle \in a} \phi(\langle e, f \rangle) \cdot \delta(\sigma_a) \right]$$

We Need a Prior over Phrase Pairs



<u>Gloss</u>

Japan

freeze

to

Russia

supply

assistance

。

We Need a Prior over Phrase Pairs

Japan to freeze aid to Russia .

	Gloss
日本	*Japan*
冻结	*freeze*
向	*to*
俄	*Russia*
提供	*supply*
援助	*assistance*
。	。

We Need a Prior over Phrase Pairs

Japan to freeze aid to Russia .

	<u>Gloss</u>
日本	*Japan*
冻结	*freeze*
向	*to*
俄	*Russia*
提供	*supply*
援助	*assistance*
。	。

$$\max_{\phi} \prod_{(\mathbf{e},\mathbf{f})} \left[\sum_{a \text{ for } (\mathbf{e},\mathbf{f})} \prod_{\langle e,f \rangle \in a} \phi(\langle e, f \rangle) \cdot \delta(\sigma_a) \right]$$

We Need a Prior over Phrase Pairs

Japan to freeze aid to Russia .

	Gloss
日本	*Japan*
冻结	*freeze*
向	*to*
俄	*Russia*
提供	*supply*
援助	*assistance*
。	。

$$\max_{\phi} \prod_{(\mathbf{e},\mathbf{f})} \left[\sum_{a \text{ for } (\mathbf{e},\mathbf{f})} \boxed{\prod_{\langle e,f\rangle \in a} \phi(\langle e,f\rangle)} \cdot \delta(\sigma_a) \right]$$

We Need a Prior over Phrase Pairs

Japan to freeze aid to Russia .

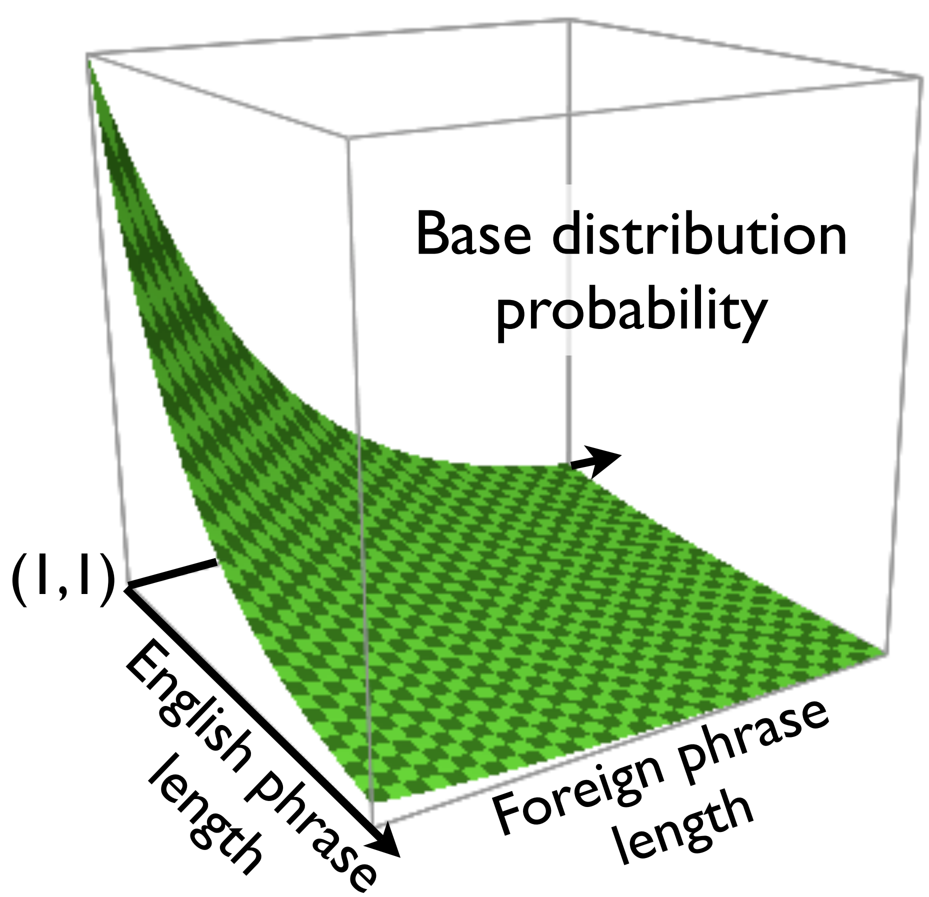
	Gloss
日本	*Japan*
冻结	*freeze*
向	*to*
俄	*Russia*
提供	*supply*
援助	*assistance*
。	。

$$\max_{\phi} \boxed{\mathrm{P}(\phi)} \cdot \prod_{(\mathbf{e},\mathbf{f})} \left[\sum_{a \text{ for } (\mathbf{e},\mathbf{f})} \prod_{\langle e,f\rangle \in a} \phi(\langle e, f\rangle) \cdot \delta(\sigma_a) \right]$$

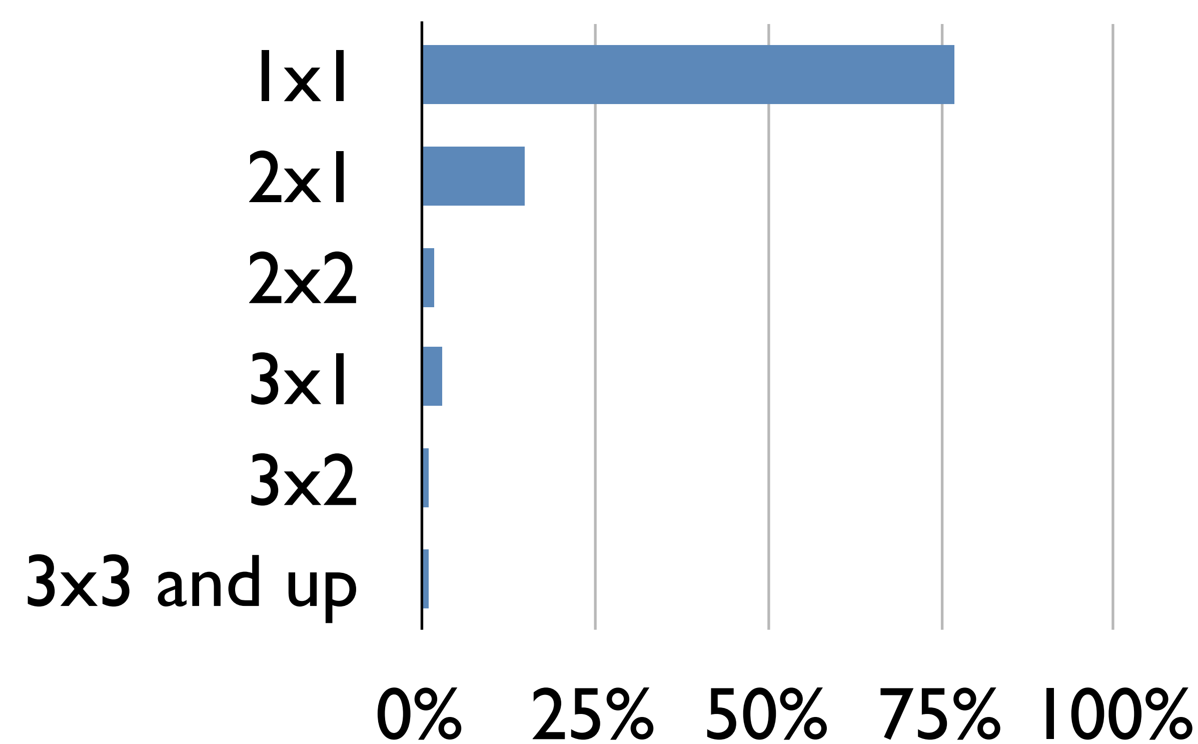
Our Prior, in Brief

A Dirichlet Process Prior that

- Strongly prefers shorter phrases (base distribution)
- Strongly prefers to reuse phrases (concentration)
- Plays nicely with our sampler (collapsed Gibbs)



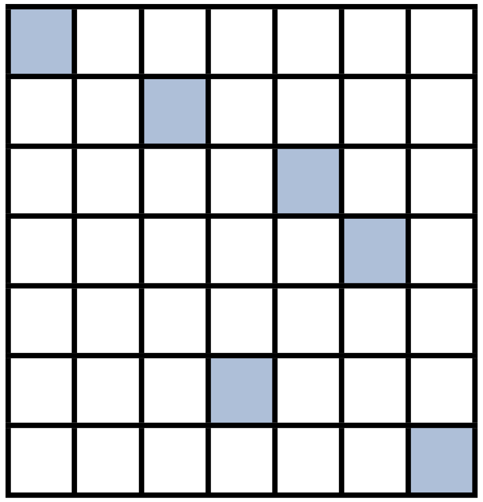
Observed phrase pair sizes



Sampling Phrase Alignments

Sampling Phrase Alignments

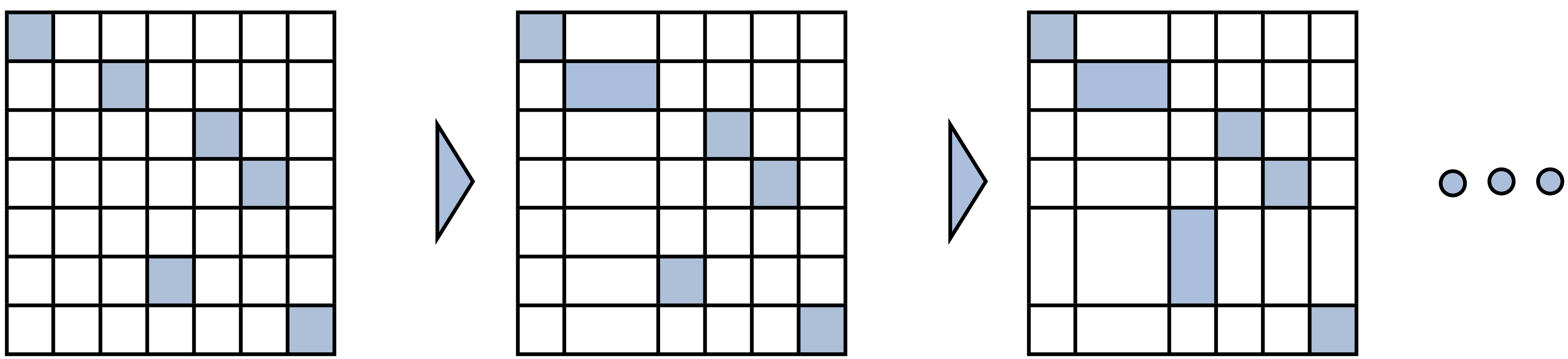
Initial phrase alignment



Sampling Phrase Alignments

Initial phrase alignment

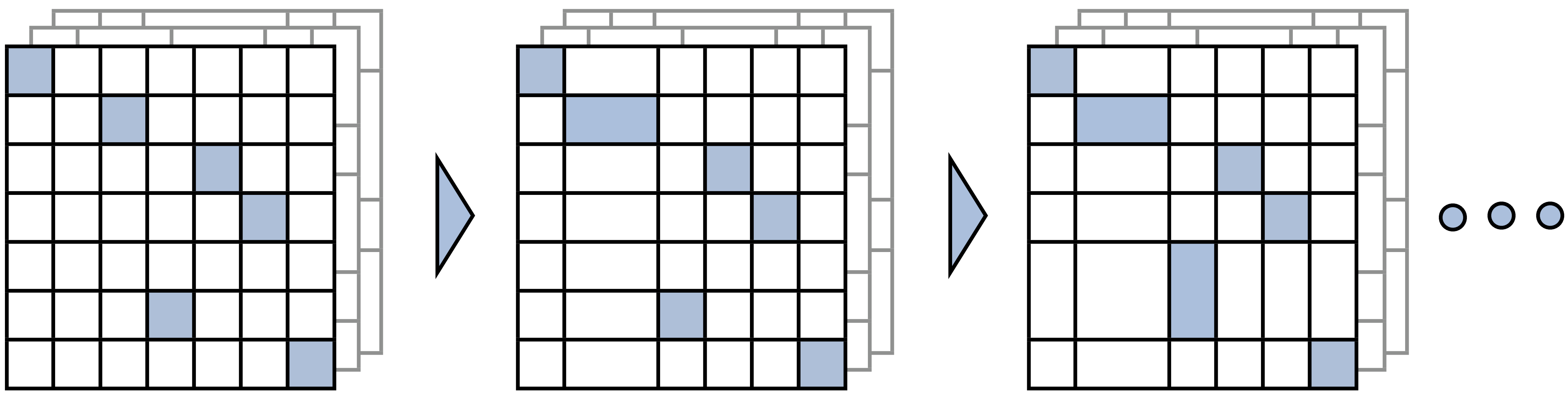
Stochastically apply local edit operators



Sampling Phrase Alignments

Initial phrase alignment

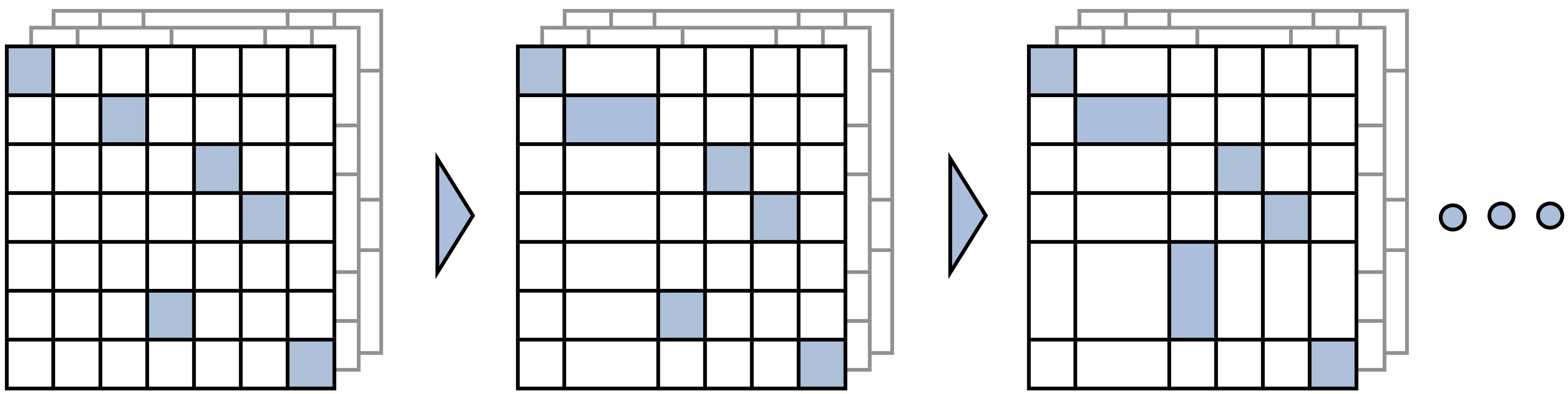
Stochastically apply local edit operators



Sampling Phrase Alignments

Initial phrase alignment

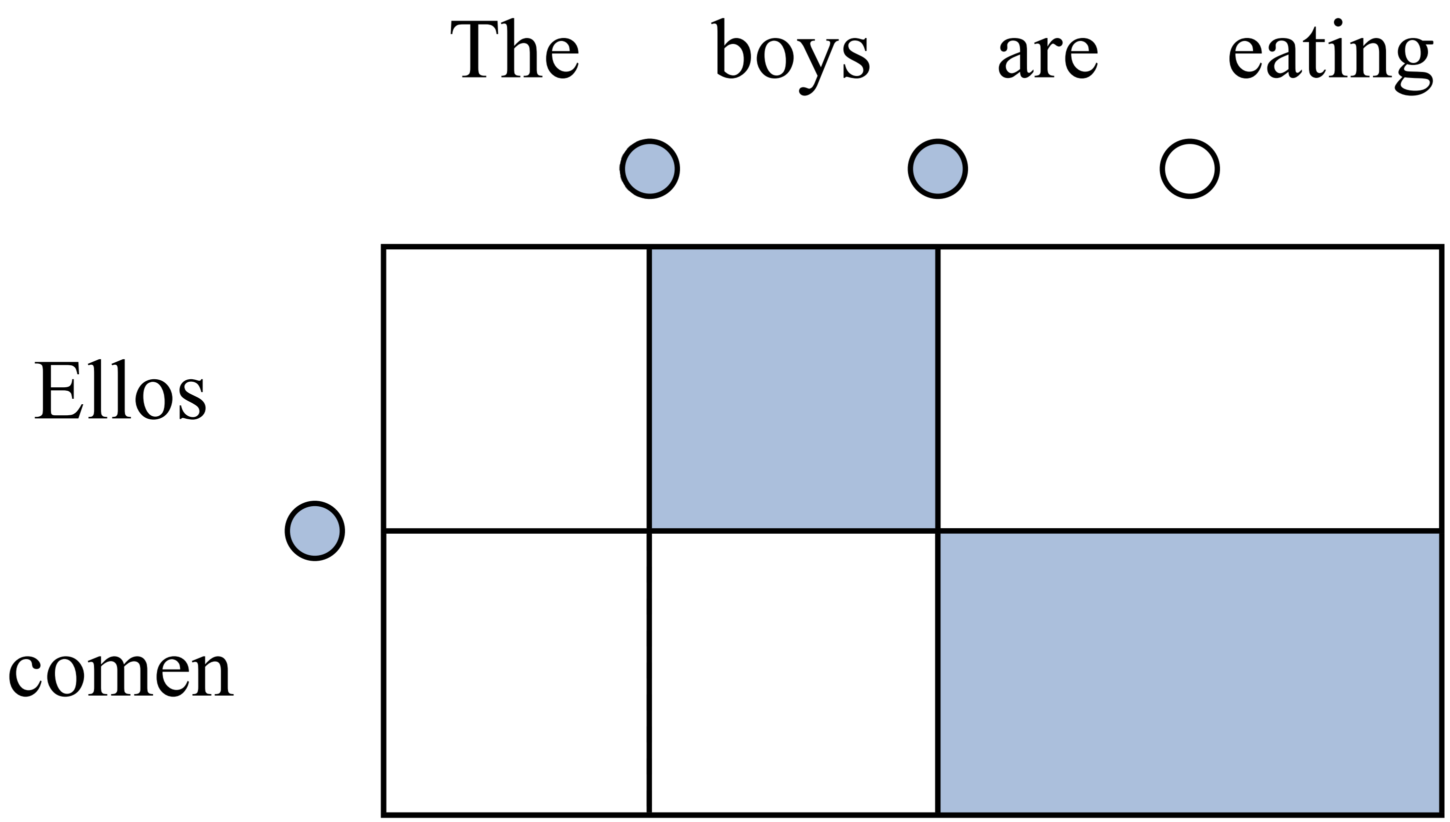
Stochastically apply local edit operators



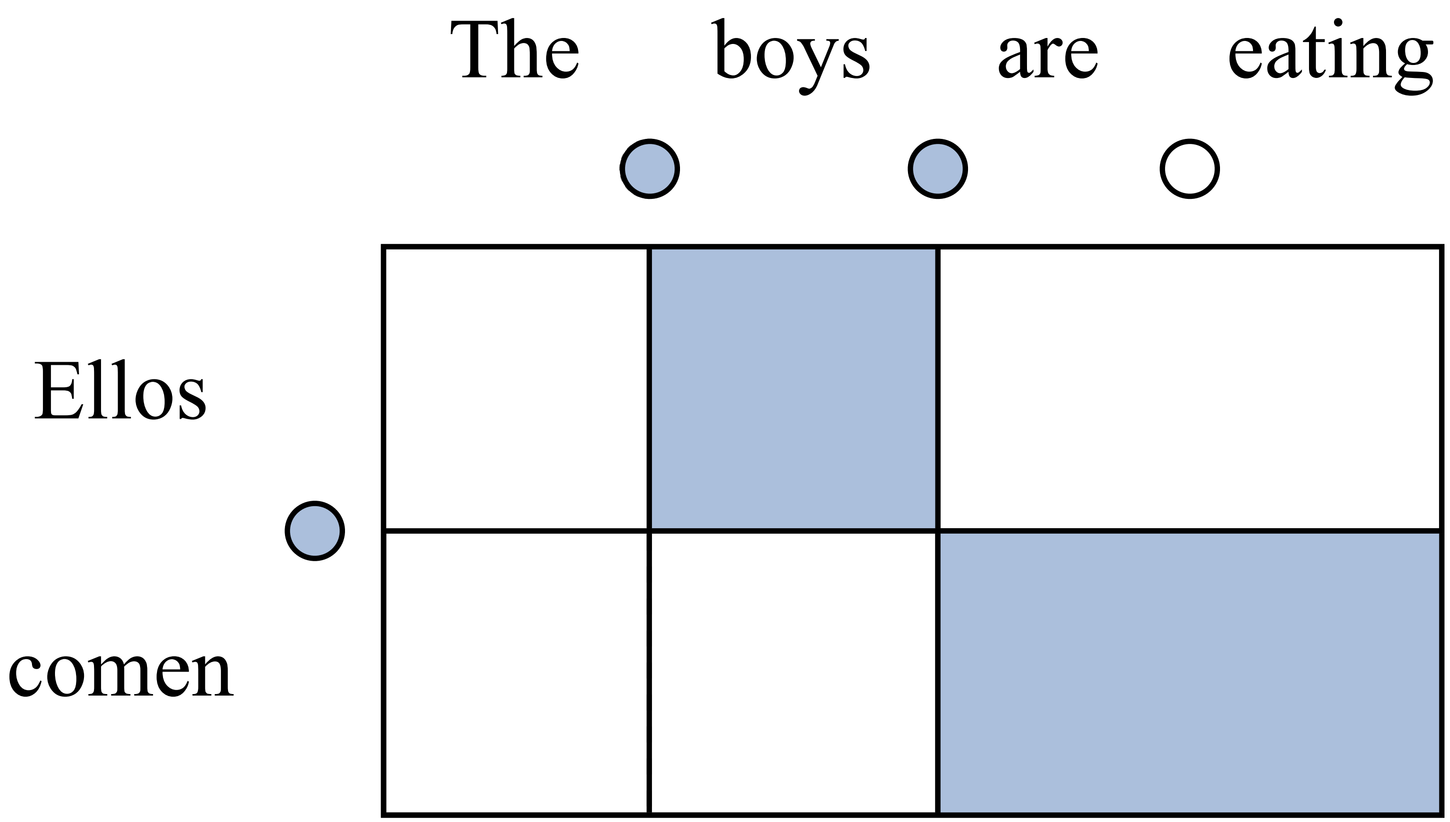
As samples are generated:

- Track phrase pair counts for current sample
- Average phrase alignment counts over all samples

Gibbs Sampling Example: Flip Operator

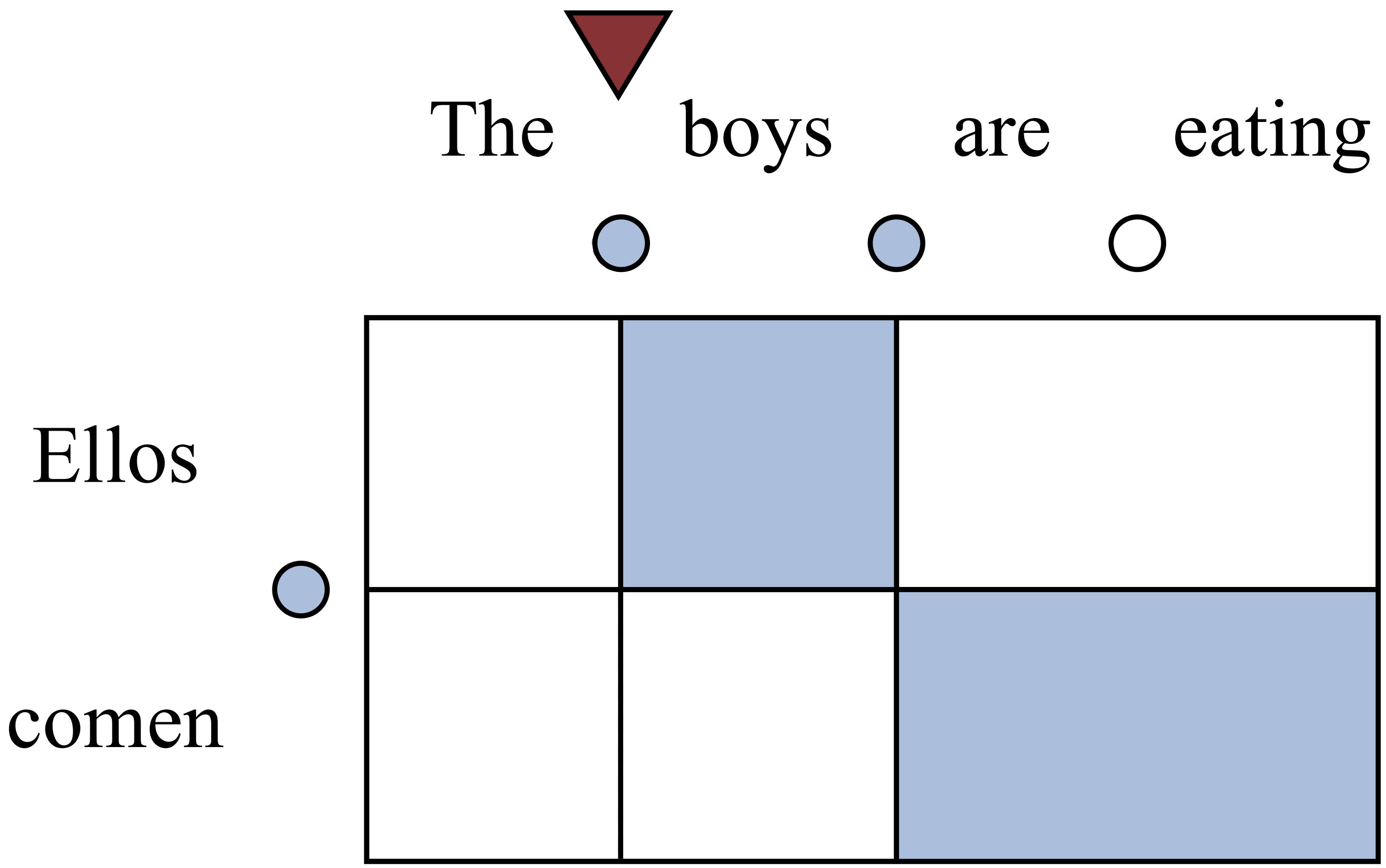


Gibbs Sampling Example: Flip Operator



Procedure:

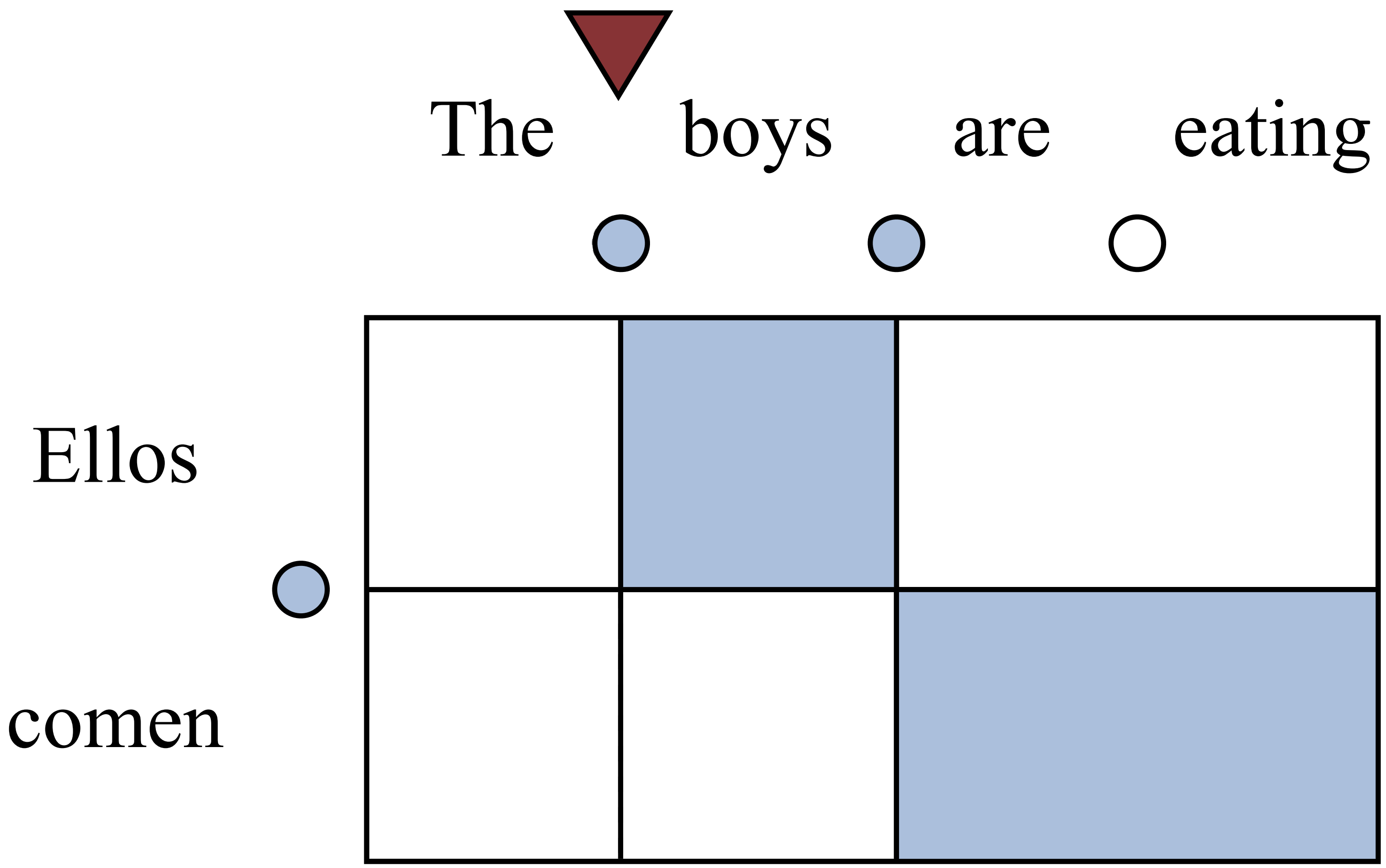
Gibbs Sampling Example: Flip Operator



Procedure:

- Choose a position

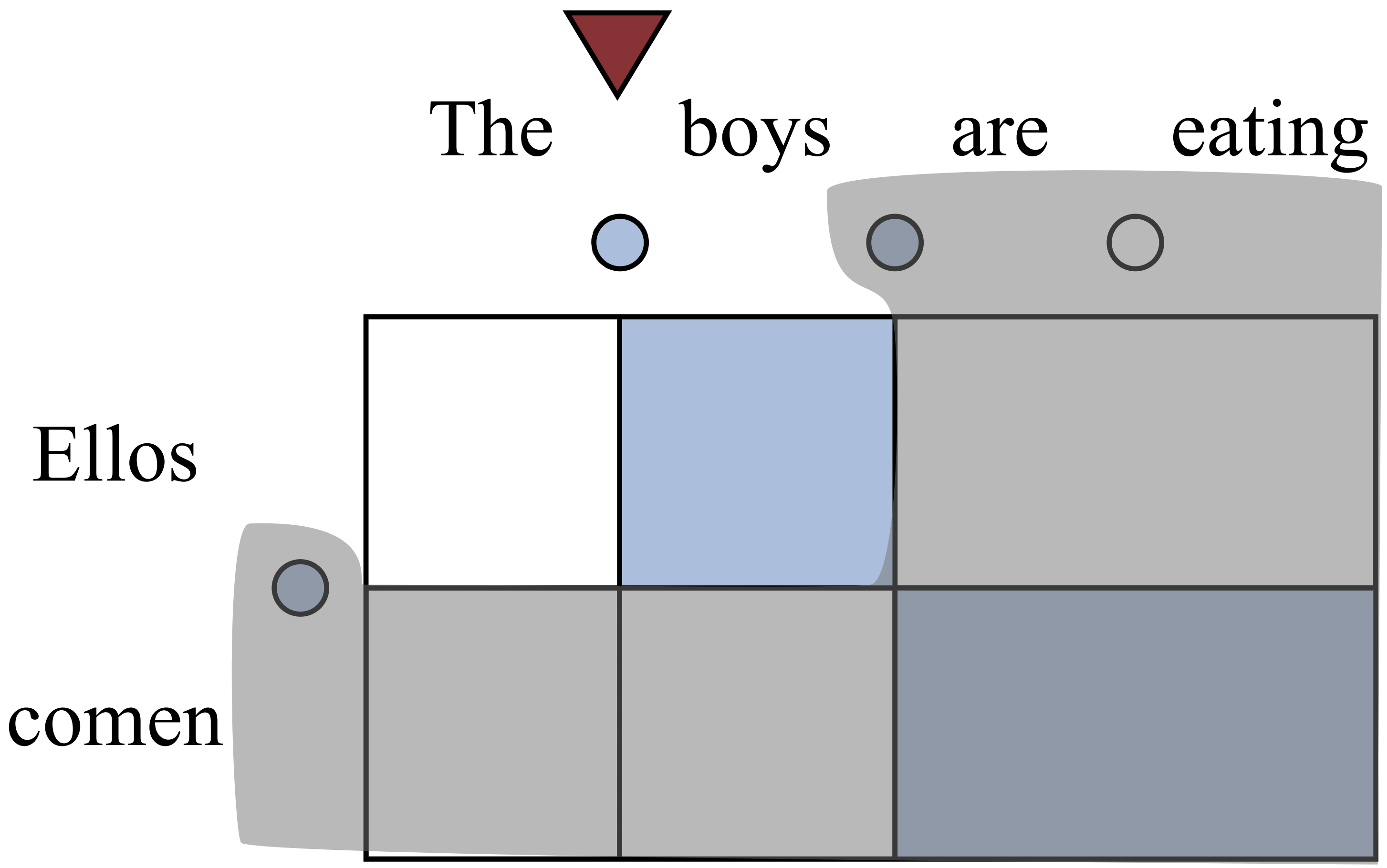
Gibbs Sampling Example: Flip Operator



Procedure:

- Choose a position
- List all outcomes

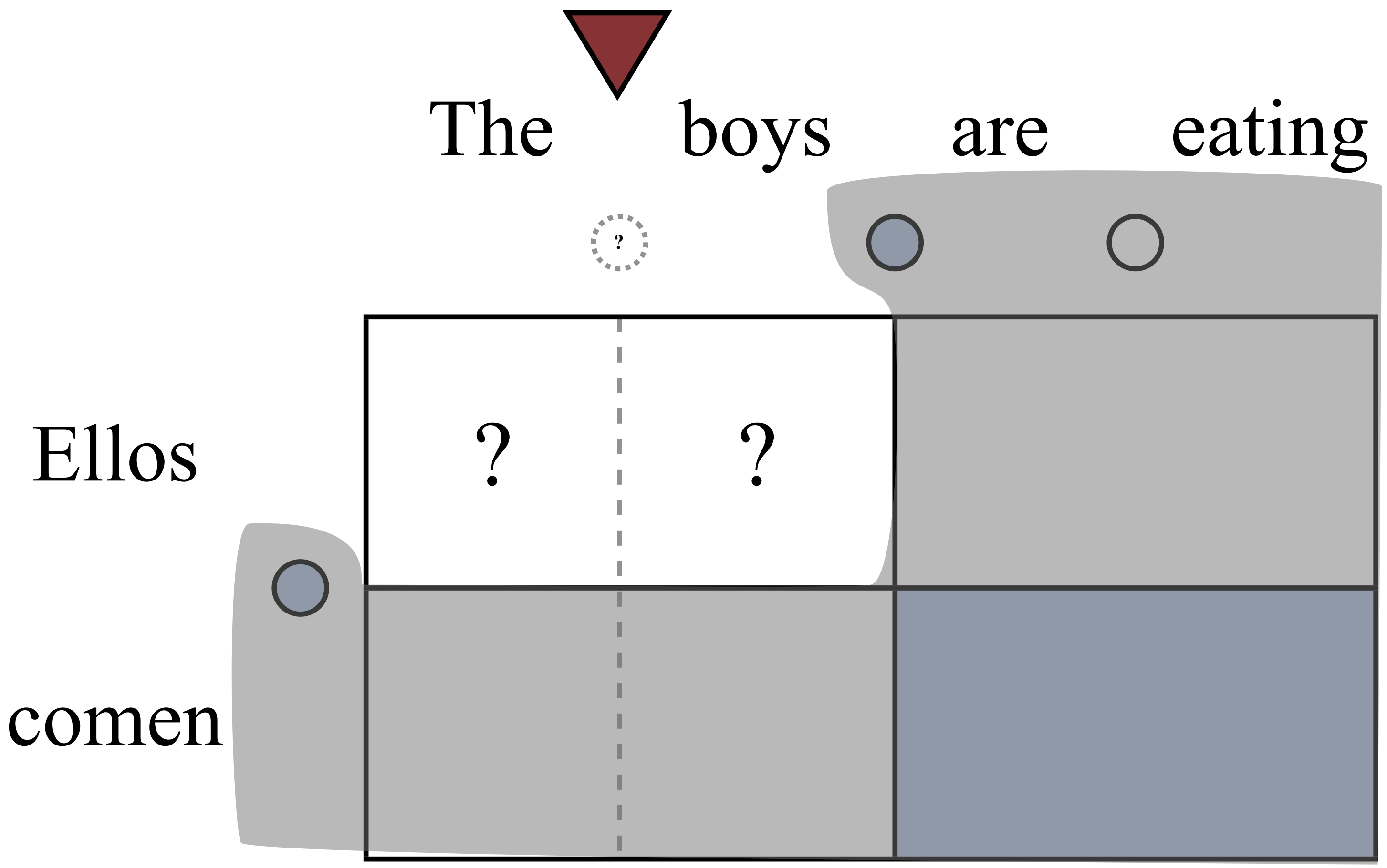
Gibbs Sampling Example: Flip Operator



Procedure:

- Choose a position
- List all outcomes

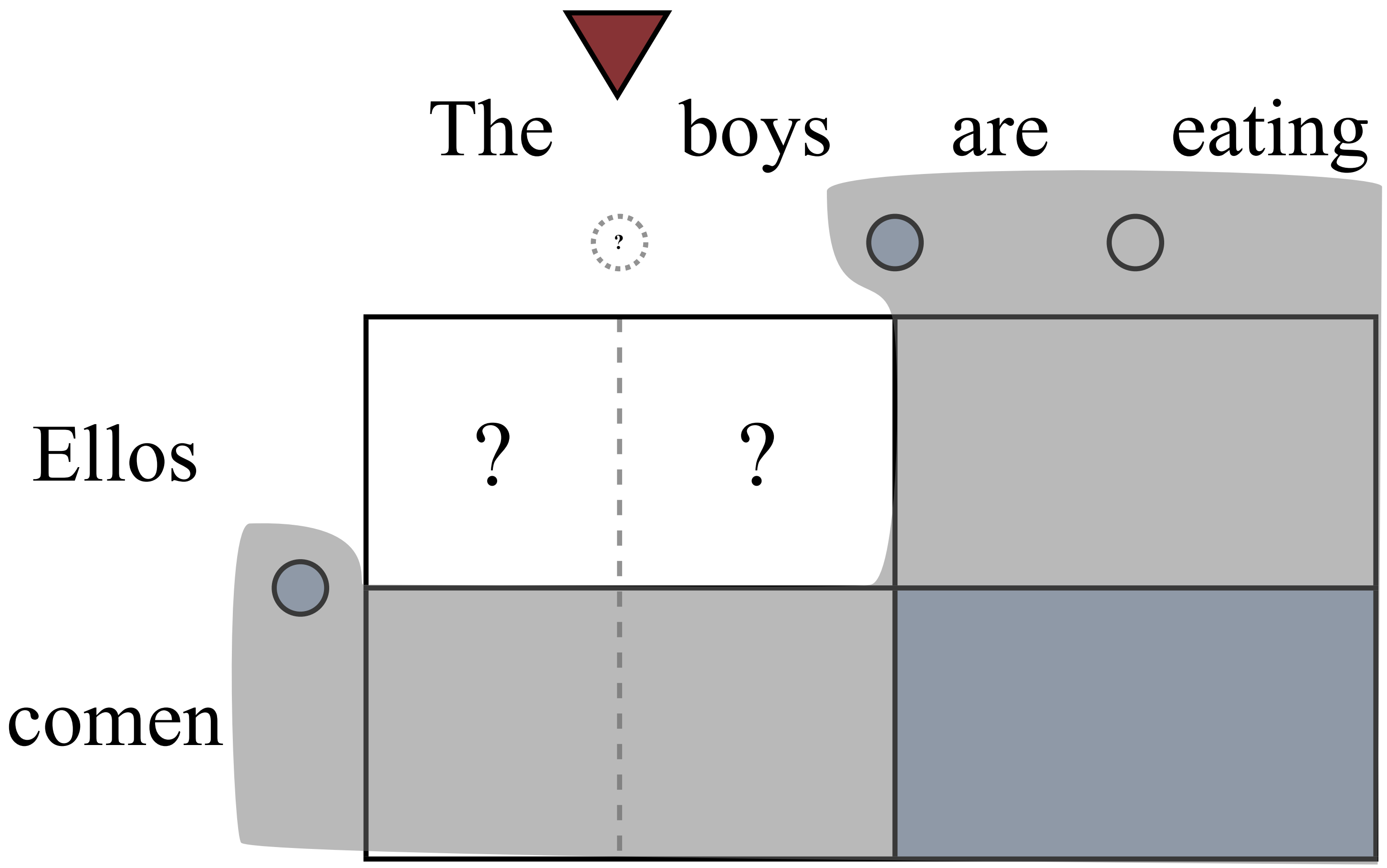
Gibbs Sampling Example: Flip Operator



Procedure:

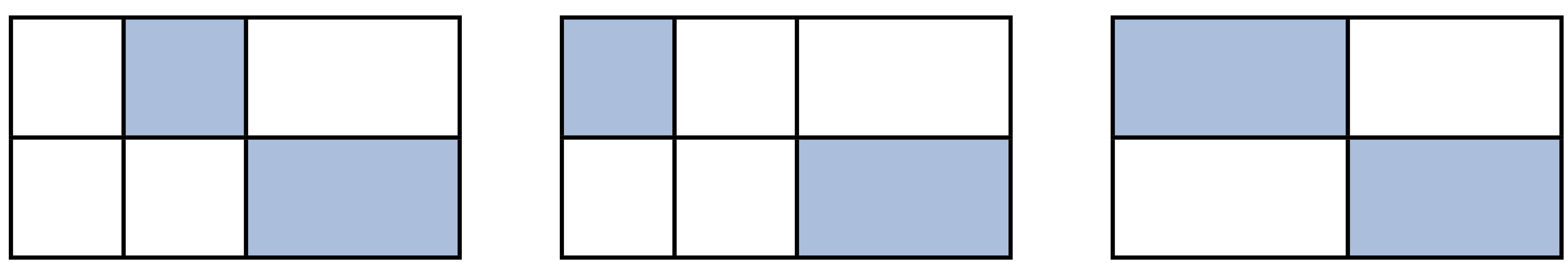
- Choose a position
- List all outcomes

Gibbs Sampling Example: Flip Operator

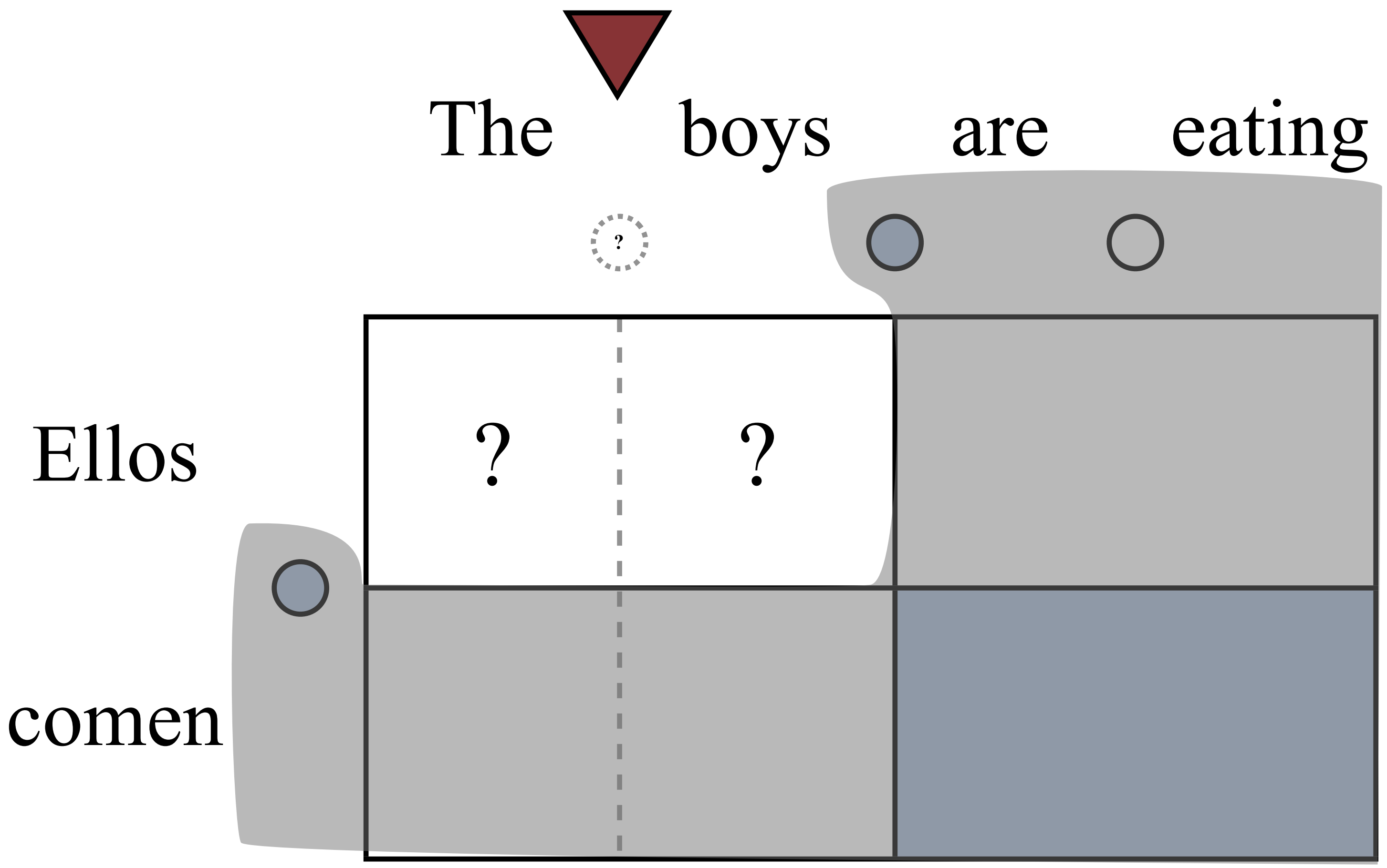


Procedure:

- Choose a position
- List all outcomes

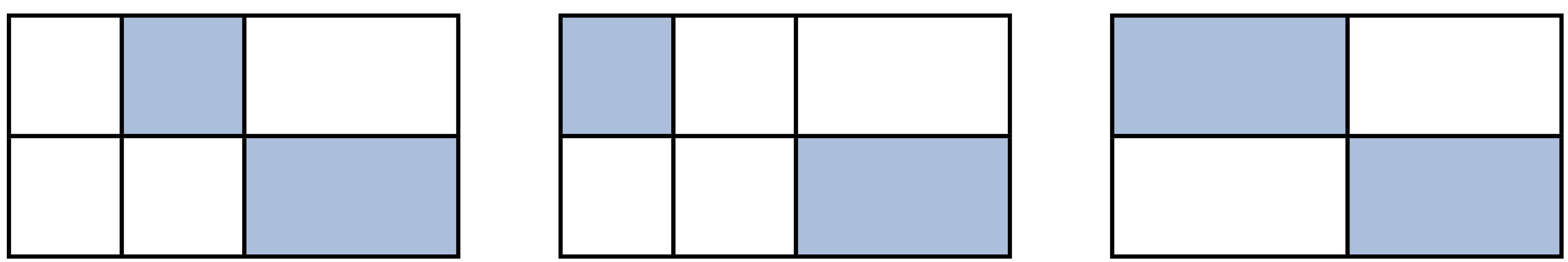


Gibbs Sampling Example: Flip Operator



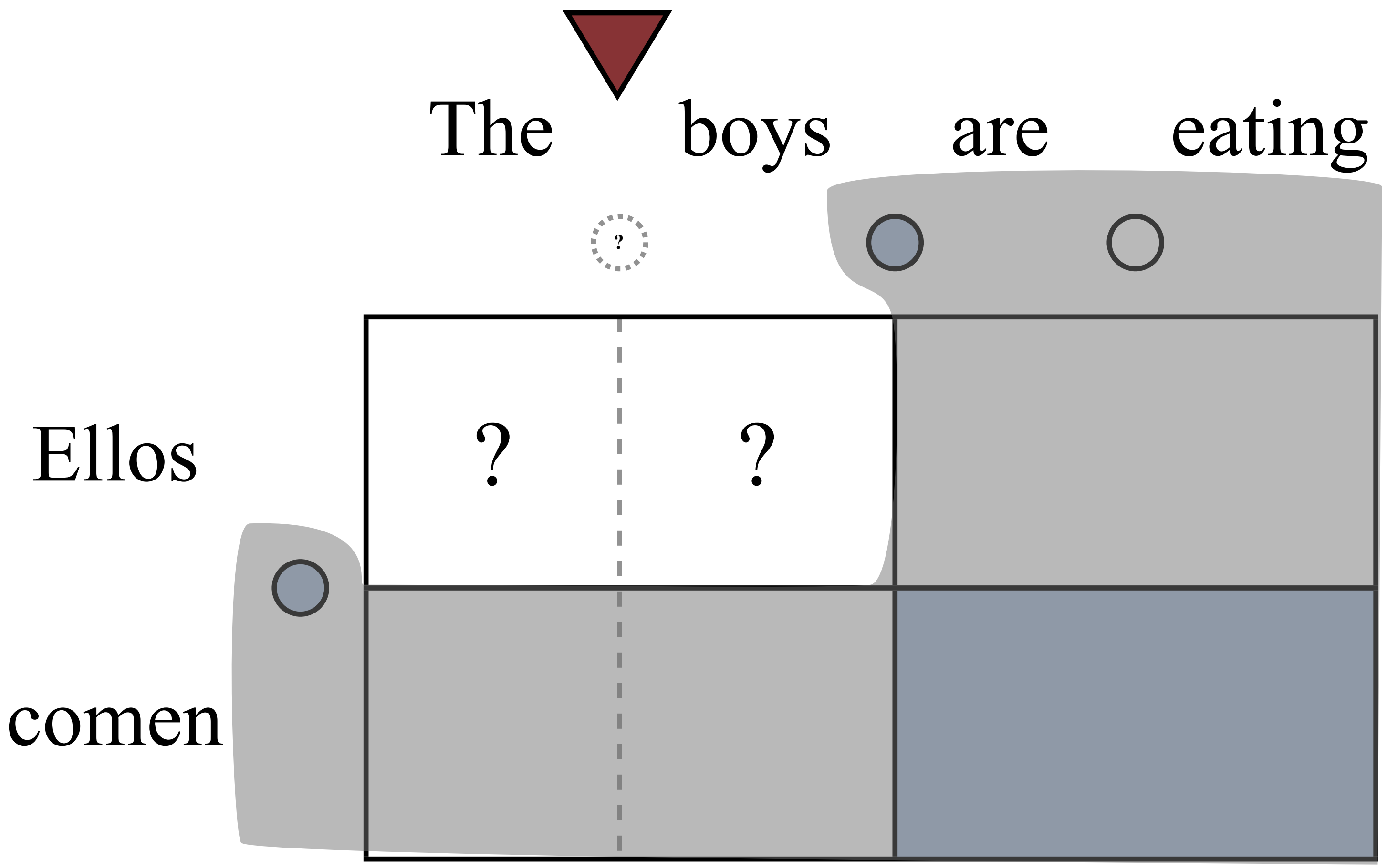
Procedure:

- Choose a position
- List all outcomes
- Compute posteriors



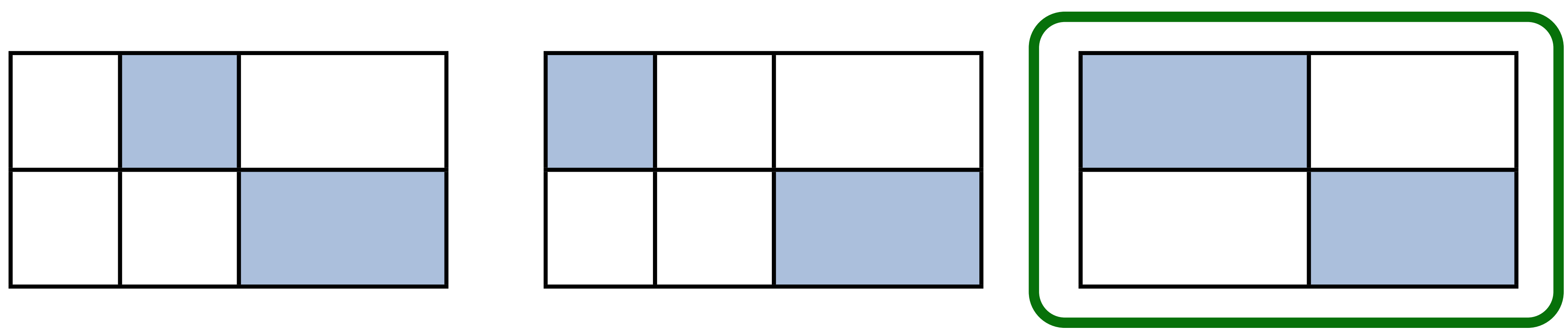
$$P(\cdot|state) \propto \quad P(boys, ellos)\cdot P(the, -) \quad\quad P(boys, -)\cdot P(the, ellos) \quad\quad P(the\ boys, ellos)$$

Gibbs Sampling Example: Flip Operator



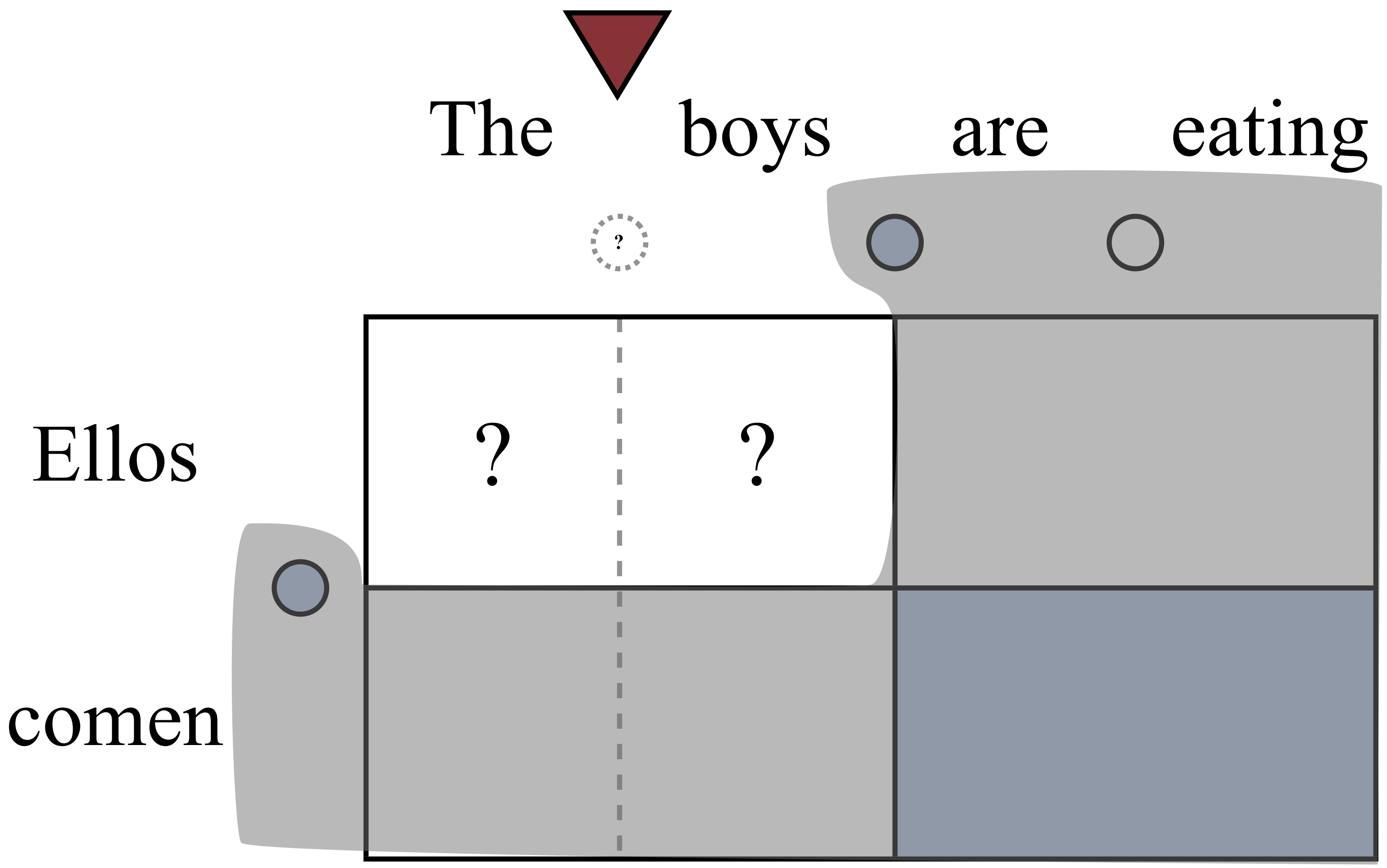
Procedure:

- Choose a position
- List all outcomes
- Compute posteriors
- Choose an outcome



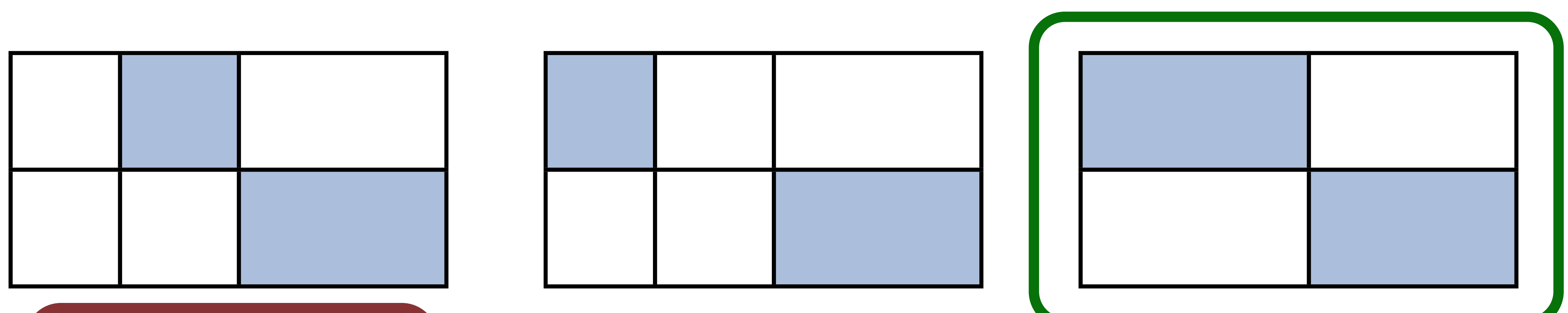
$$P(\cdot|state) \propto \quad P(boys, ellos) \cdot P(the, -) \qquad P(boys, -) \cdot P(the, ellos) \qquad P(the\ boys, ellos)$$

Gibbs Sampling Example: Flip Operator



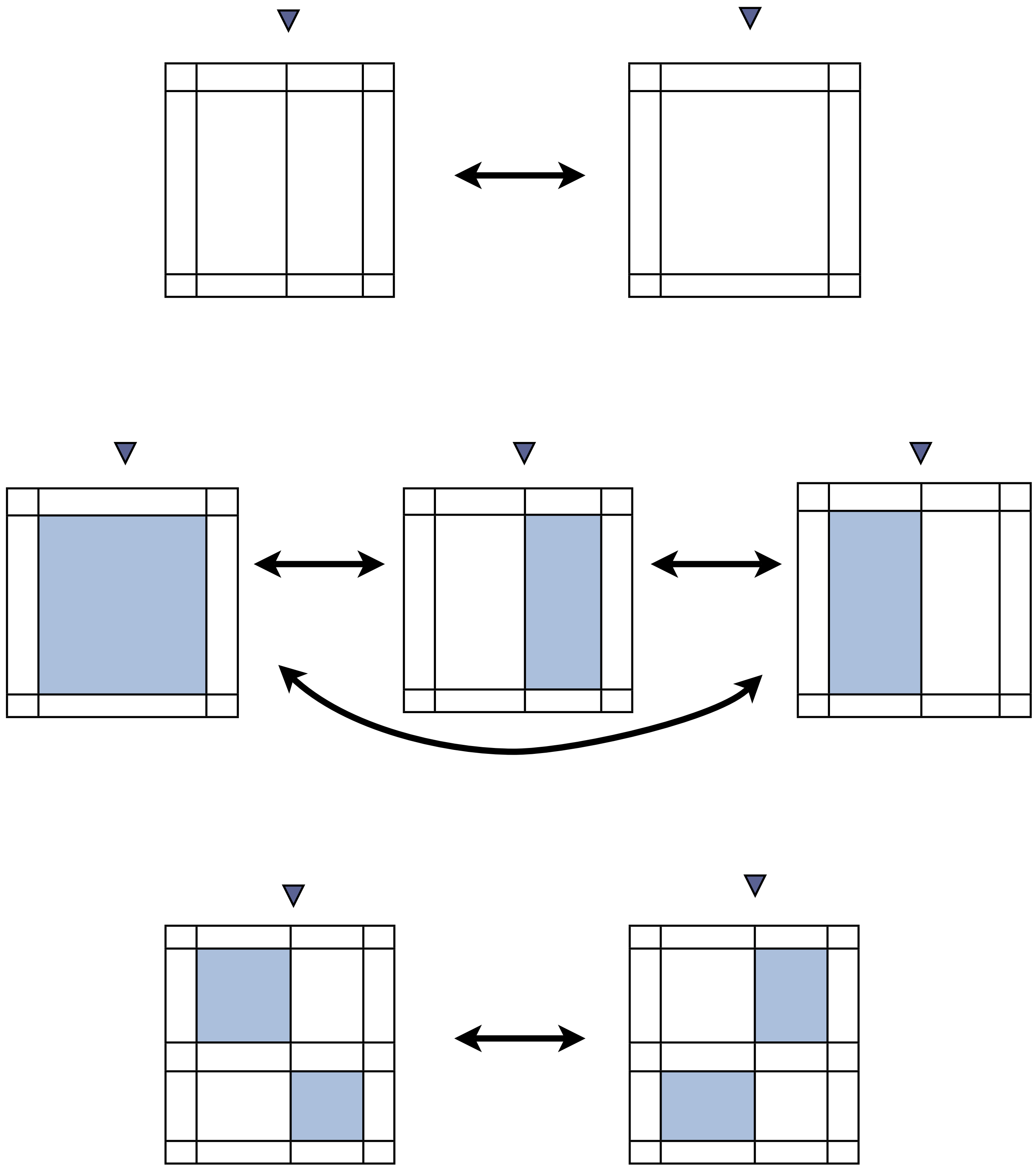
Procedure:

- Choose a position
- List all outcomes
- Compute posteriors
- Choose an outcome
- Update statistics



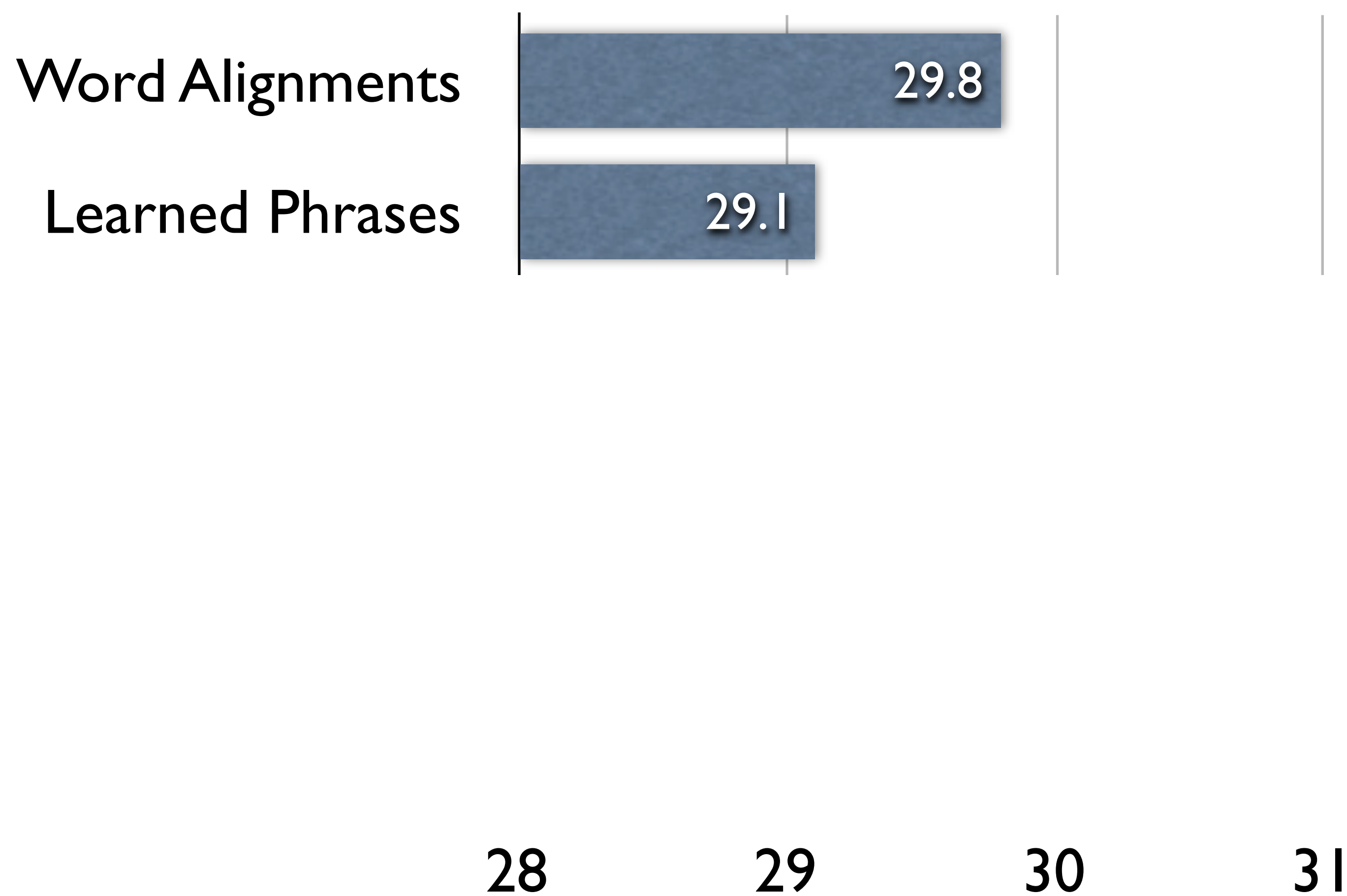
$$P(\cdot|state) \propto \quad \boxed{\begin{matrix} P(boys, ellos)\cdot \\ P(the, -) \end{matrix}} \quad \begin{matrix} P(boys, -)\cdot \\ P(the, ellos) \end{matrix} \quad \boxed{P(the\ boys, ellos)}$$

The Flip Operator Configurations



Performance Results

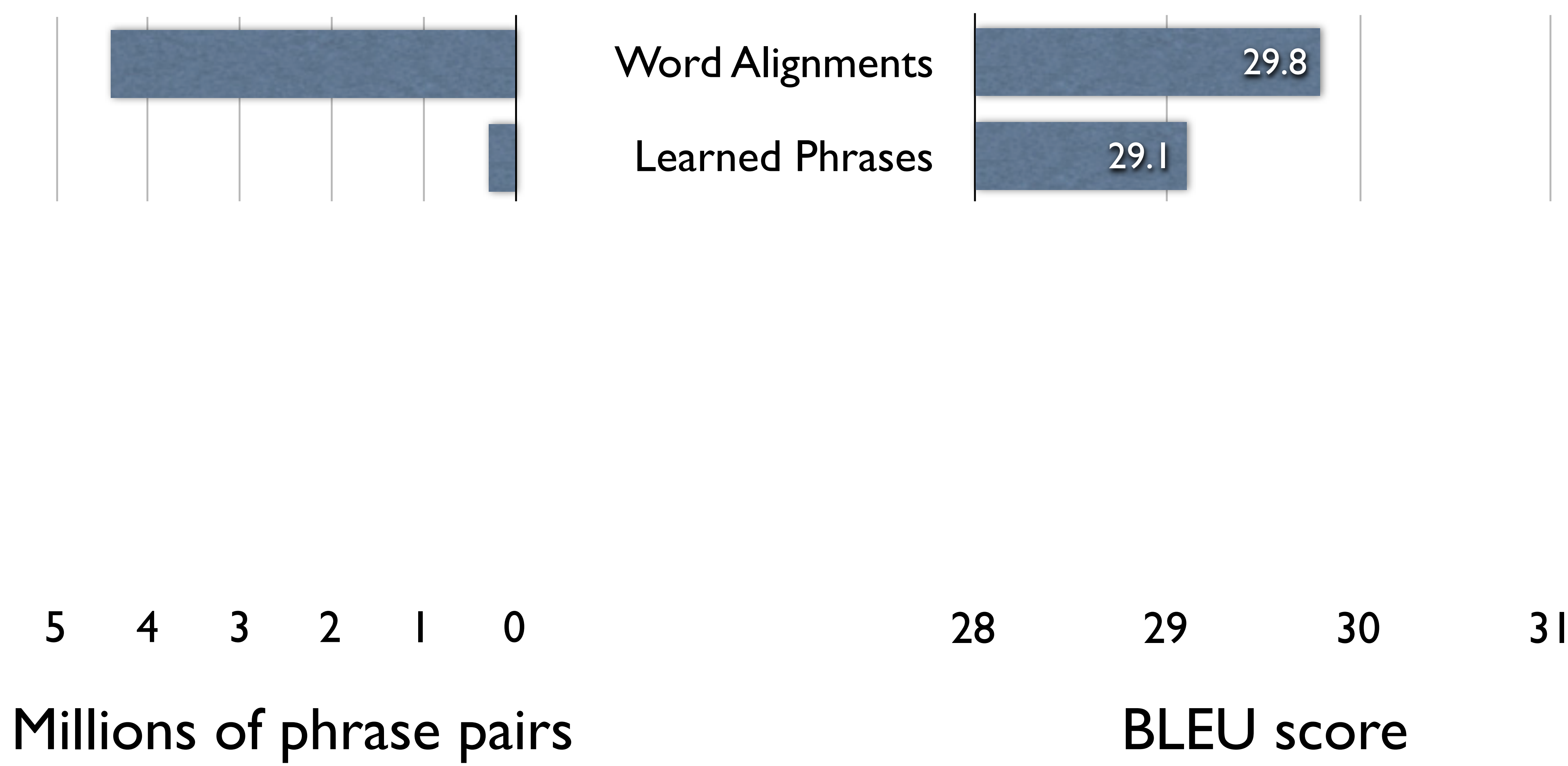
Translation performance in a phrase-based system (Moses) for English-Spanish parliamentary proceedings (Europarl)



* Includes additional lexical features to discourage undesirable phrase pairs

Performance Results

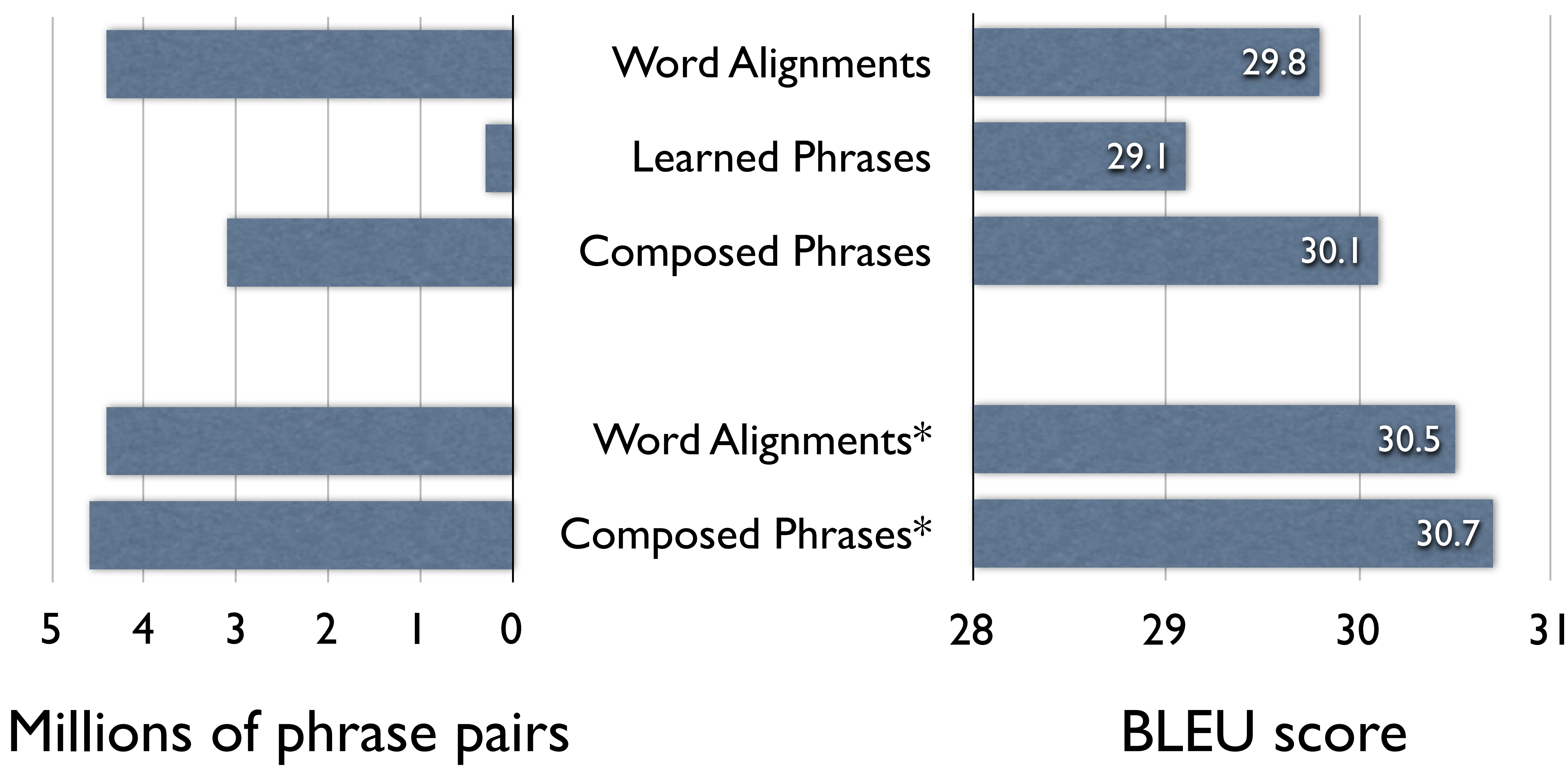
Translation performance in a phrase-based system (Moses)
for English-Spanish parliamentary proceedings (Europarl)



* Includes additional lexical features to discourage undesirable phrase pairs

Performance Results

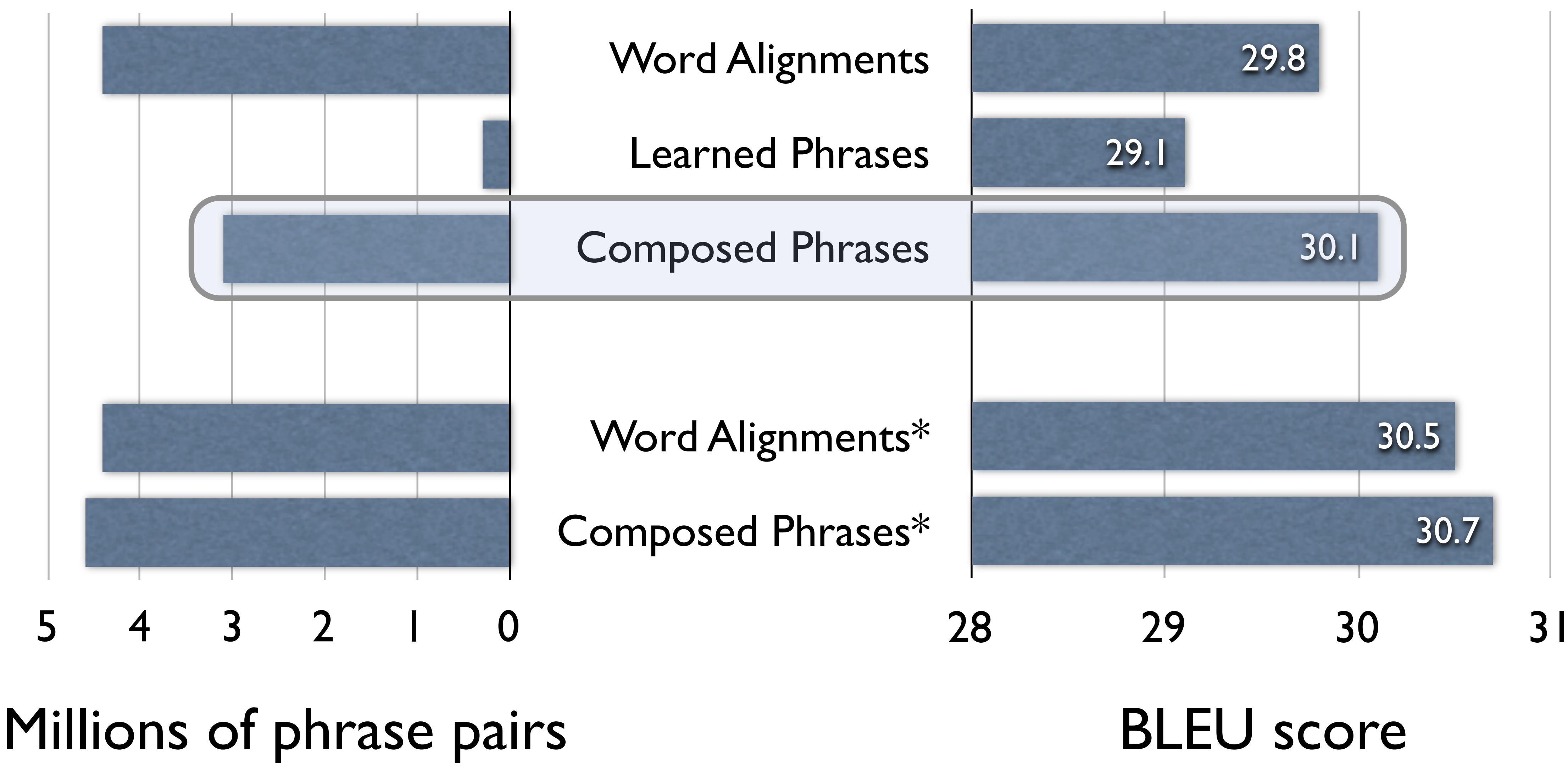
Translation performance in a phrase-based system (Moses) for English-Spanish parliamentary proceedings (Europarl)



* Includes additional lexical features to discourage undesirable phrase pairs

Performance Results

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



* Includes additional lexical features to discourage undesirable phrase pairs