Model-Based Aligner Combination Using Dual Decomposition

John DeNero and Klaus Macherey
Google Research

Overview

Task: Combine predictions of two directional alignment models

Approach: Search for maximal assignment in a graphical model

Method:

Motivation:

Result:

Outline

Task: Combine predictions of two directional alignment models

Approach: Search for maximal assignment in a graphical model

Method: Dual decomposition inference

Motivation: Globally optimal upon convergence

Result:

Outline

Task: Combine predictions of two directional alignment models

Approach: Search for maximal assignment in a graphical model

Method: Dual decomposition inference

Motivation: Globally optimal upon convergence

Result: Convergence is rare, but method yields empirical benefit

Word Alignment Models

$\mathbf{e}$:	How	are	you	?
$\mathbf{f}$:	你 [you]	好 [good]	吗 [*interrogative particle*]	?

Word Alignment Models

$\mathbf{e}$:	How	are	you	?
$\mathbf{f}$:	你 [you]	好 [good]	吗 [*interrogative particle*]	?

$$\mathrm{P}(\mathbf{f}, \mathbf{a} | \mathbf{e})$$

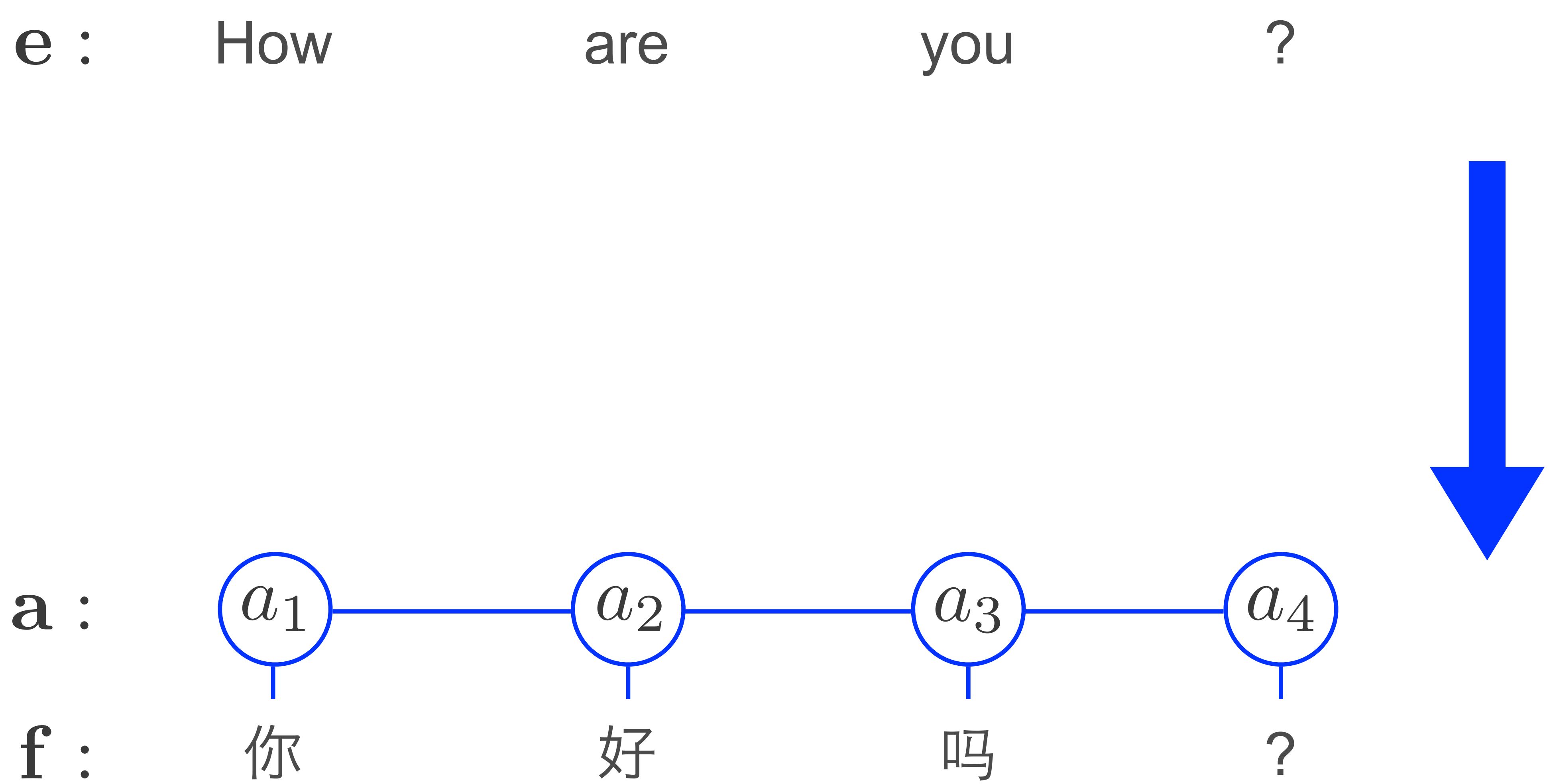
HMM-Based Generative Alignment Model

$\mathbf{e}$: How are you ?

$\mathbf{f}$: 你 好 吗 ?

$$\mathrm{P}(\mathbf{f}, \mathbf{a} | \mathbf{e})$$

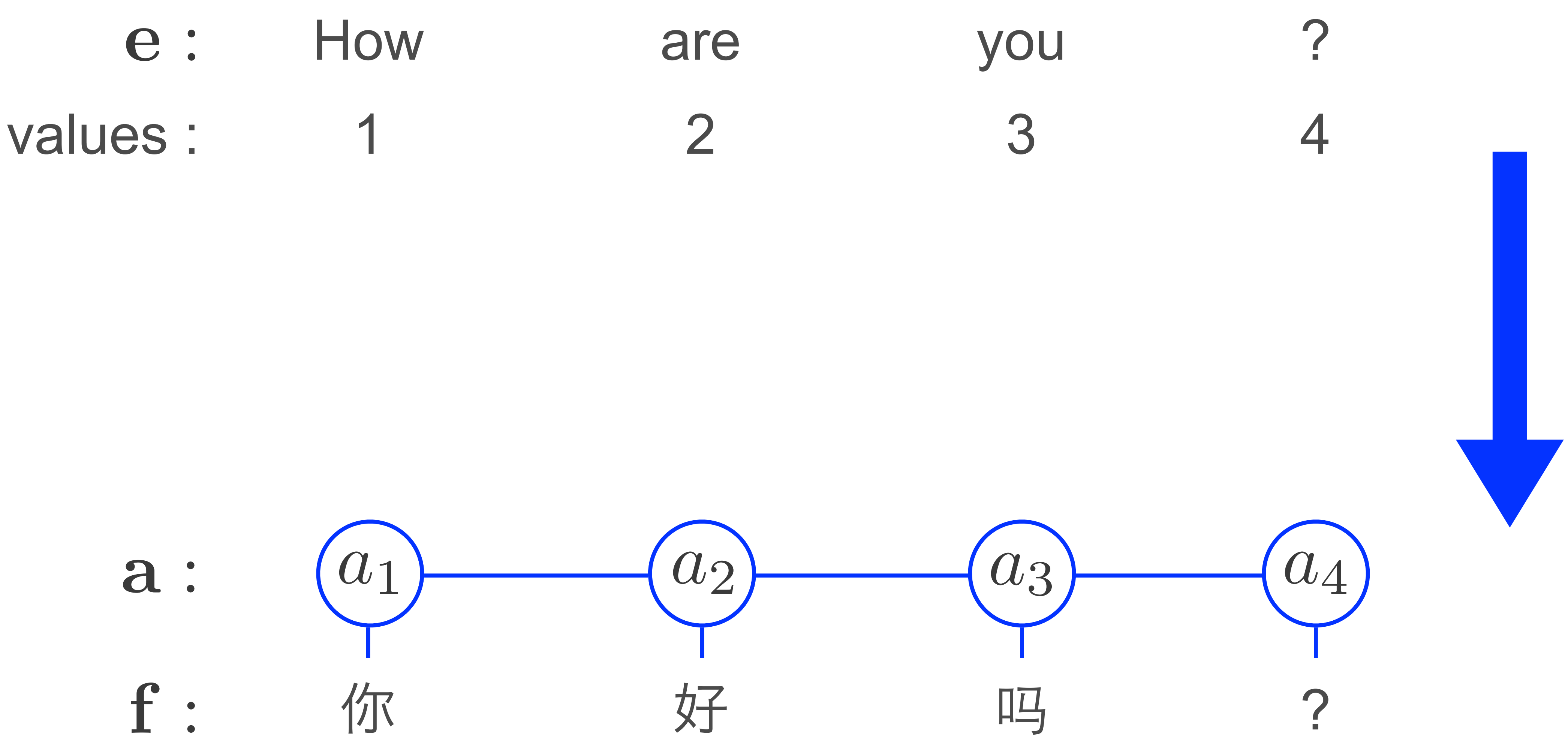
HMM-Based Generative Alignment Model



$$\mathrm{P}(\mathbf{f}, \mathbf{a}|\mathbf{e}) = \prod_{j=1}^{|\mathbf{f}|} \mathrm{P}(a_j|a_{j-1}) \cdot \mathrm{P}(f_j|e_{a_j})$$

(Vogel et al., 1996)

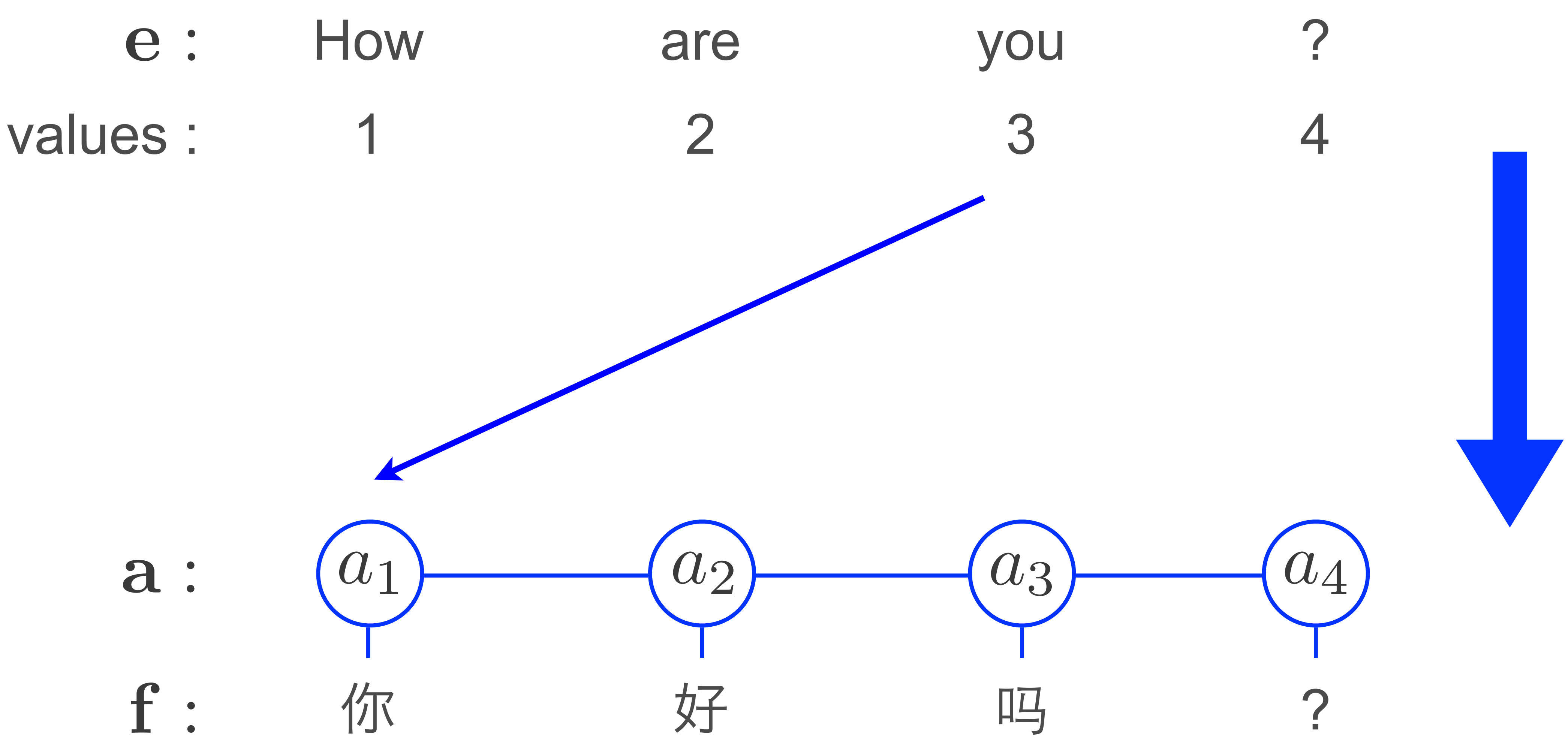
HMM-Based Generative Alignment Model



$$\mathrm{P}(\mathbf{f}, \mathbf{a} | \mathbf{e}) = \prod_{j=1}^{|\mathbf{f}|} \mathrm{P}(a_j | a_{j-1}) \cdot \mathrm{P}(f_j | e_{a_j})$$

(Vogel et al., 1996)

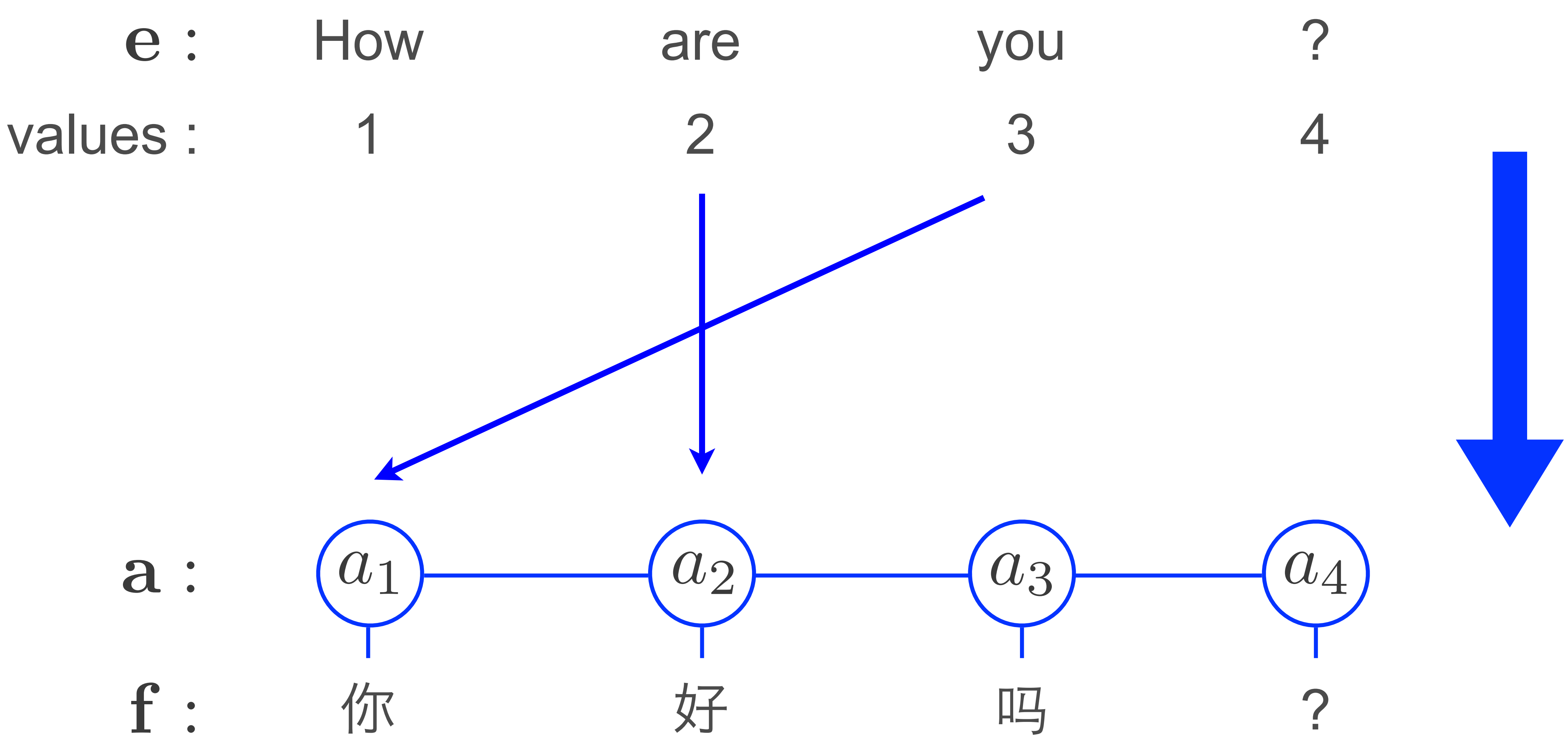
HMM-Based Generative Alignment Model



$$\mathrm{P}(\mathbf{f}, \mathbf{a} | \mathbf{e}) = \prod_{j=1}^{|\mathbf{f}|} \mathrm{P}(a_j | a_{j-1}) \cdot \mathrm{P}(f_j | e_{a_j})$$

(Vogel et al., 1996)

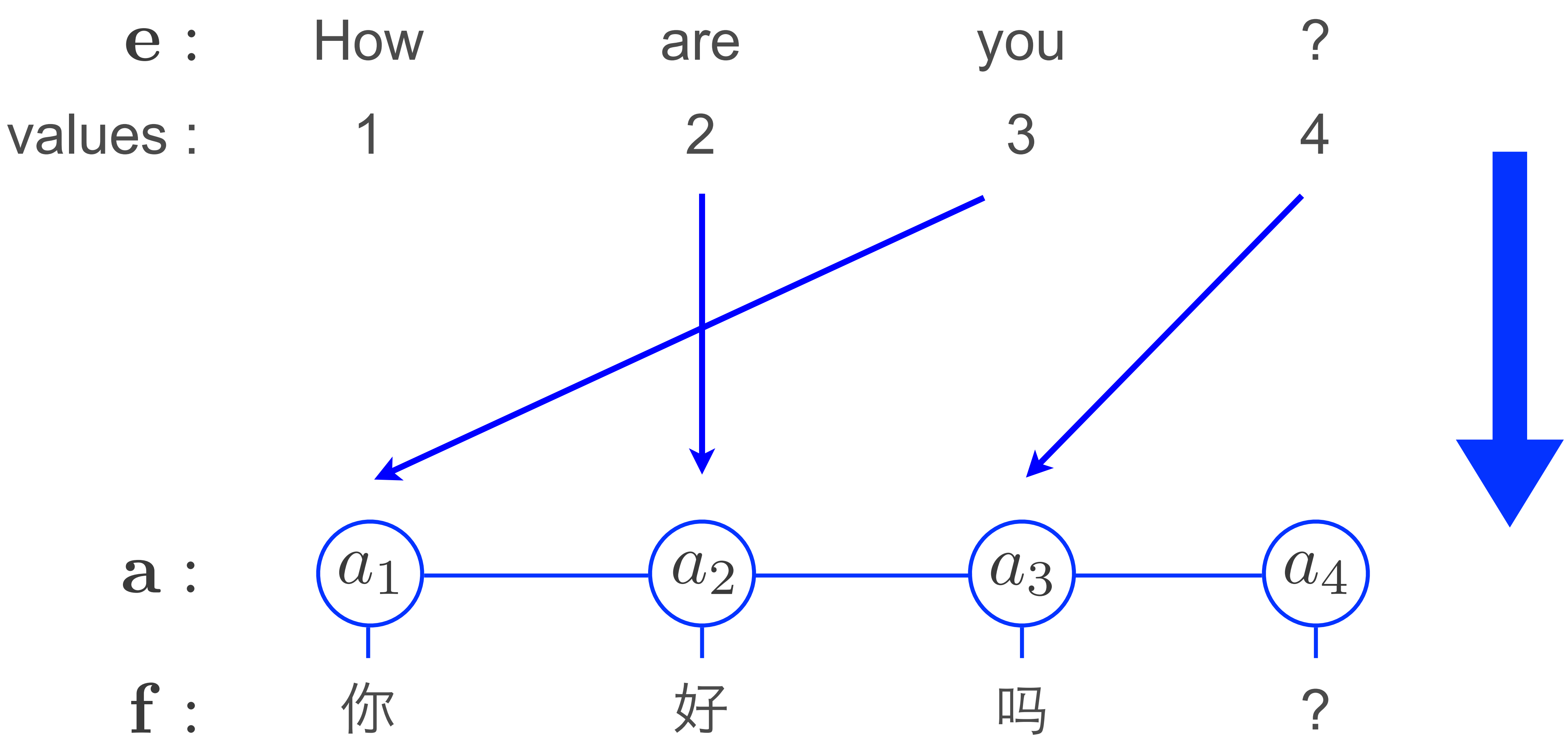
HMM-Based Generative Alignment Model



$$P(\mathbf{f}, \mathbf{a} | \mathbf{e}) = \prod_{j=1}^{|\mathbf{f}|} P(a_j | a_{j-1}) \cdot P(f_j | e_{a_j})$$

(Vogel et al., 1996)

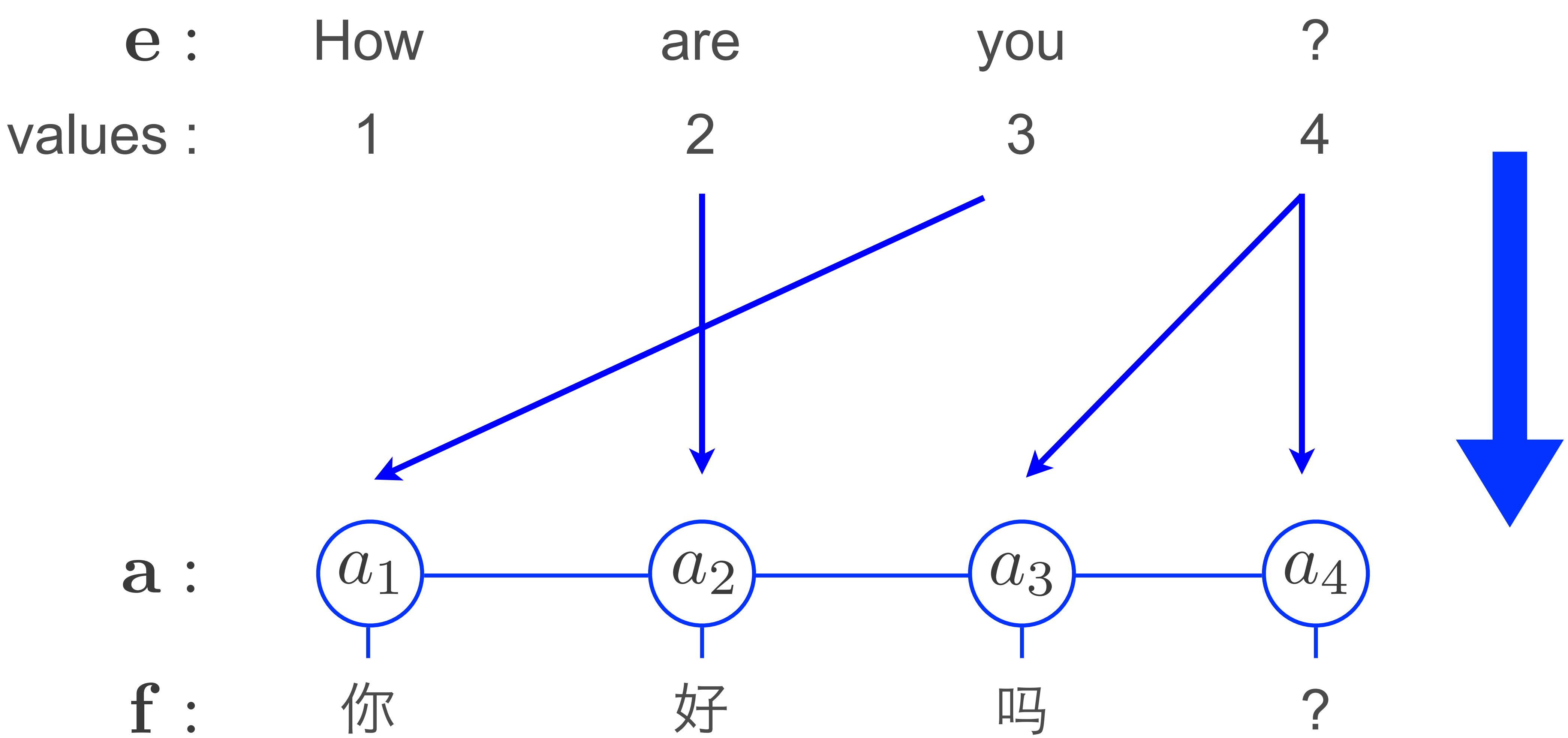
HMM-Based Generative Alignment Model



$$P(\mathbf{f}, \mathbf{a}|\mathbf{e}) = \prod_{j=1}^{|\mathbf{f}|} P(a_j|a_{j-1}) \cdot P(f_j|e_{a_j})$$

(Vogel et al., 1996)

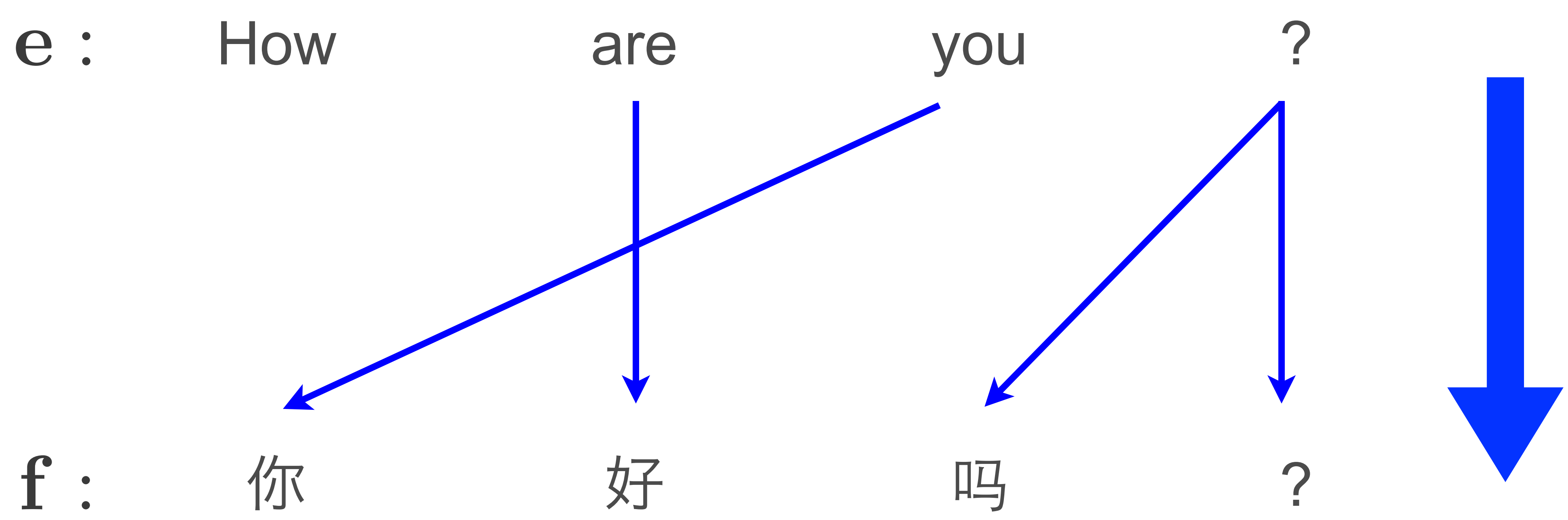
HMM-Based Generative Alignment Model



$$\mathrm{P}(\mathbf{f}, \mathbf{a} | \mathbf{e}) = \prod_{j=1}^{|\mathbf{f}|} \mathrm{P}(a_j | a_{j-1}) \cdot \mathrm{P}(f_j | e_{a_j})$$

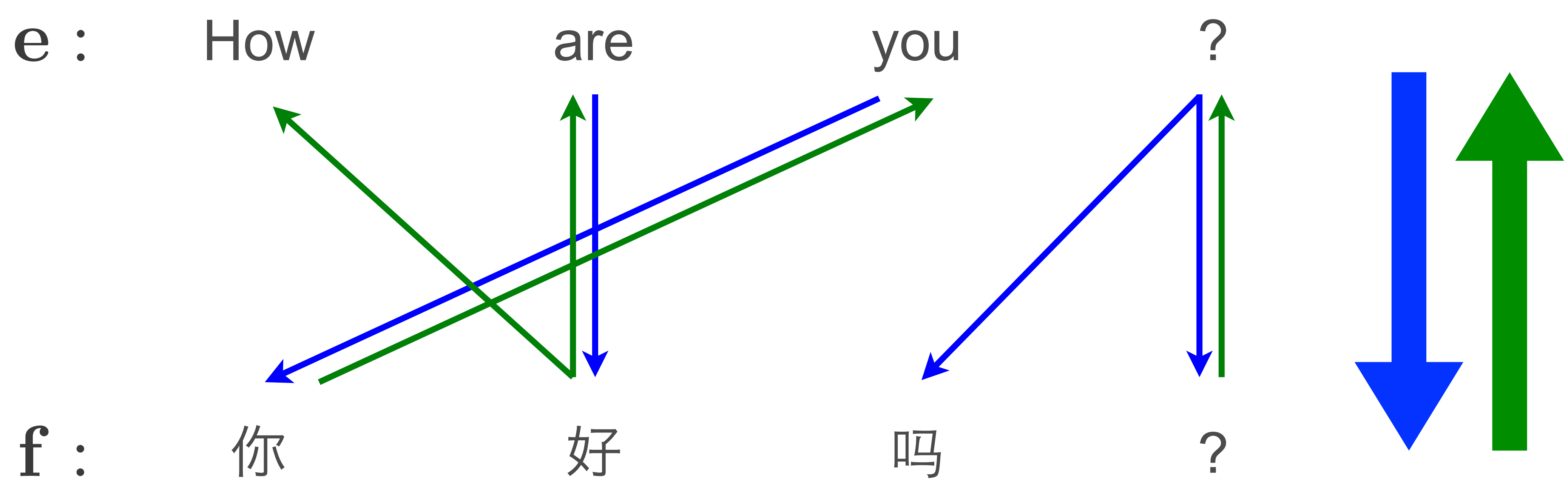
(Vogel et al., 1996)

Heuristic Approaches to Aligner Combination



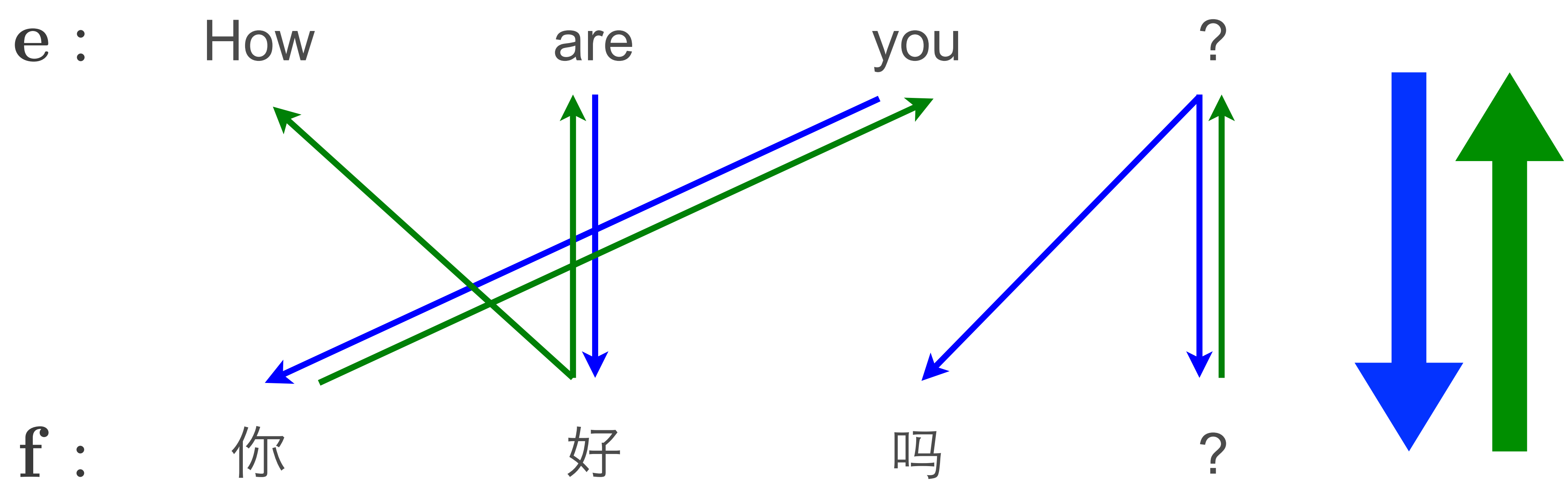
$$P(\mathbf{f}, \mathbf{a} | \mathbf{e})$$

Heuristic Approaches to Aligner Combination



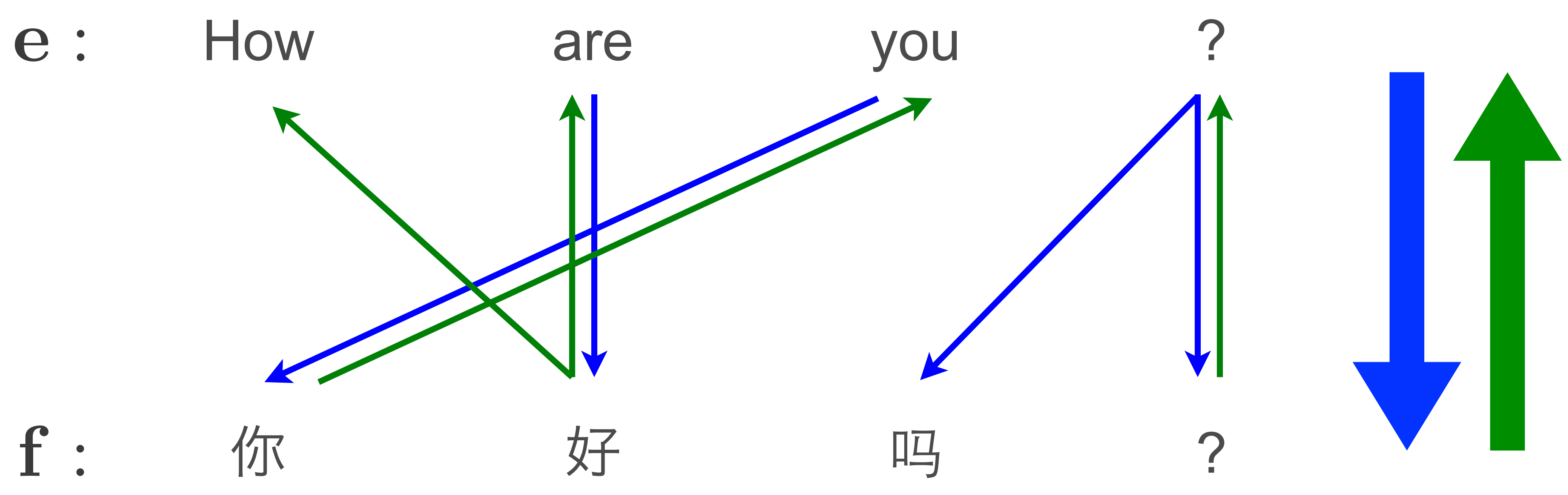
$P(\mathbf{f}, \mathbf{a} | \mathbf{e})$ $P(\mathbf{e}, \mathbf{b} | \mathbf{f})$

Heuristic Approaches to Aligner Combination



$$\hat{\mathbf{a}} = \max_{\mathbf{a}} \mathrm{P}(\mathbf{f}, \mathbf{a} | \mathbf{e}) \qquad \hat{\mathbf{b}} = \max_{\mathbf{b}} \mathrm{P}(\mathbf{e}, \mathbf{b} | \mathbf{f})$$

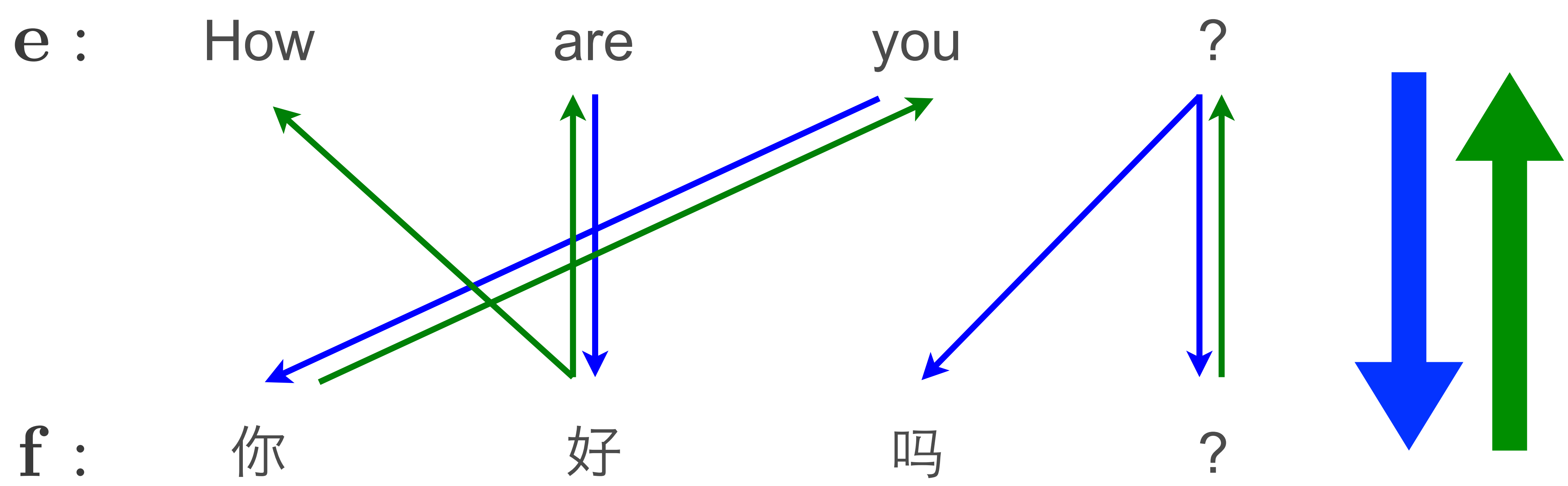
Heuristic Approaches to Aligner Combination



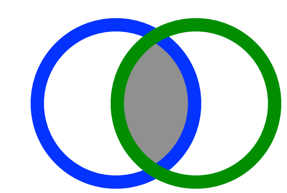
$\hat{\mathbf{a}} = \max_{\mathbf{a}} \mathrm{P}(\mathbf{f}, \mathbf{a} | \mathbf{e})$ $\hat{\mathbf{b}} = \max_{\mathbf{b}} \mathrm{P}(\mathbf{e}, \mathbf{b} | \mathbf{f})$ $\mathrm{symm}(\hat{\mathbf{a}}, \hat{\mathbf{b}})$

(Och et al., 1999)

Heuristic Approaches to Aligner Combination



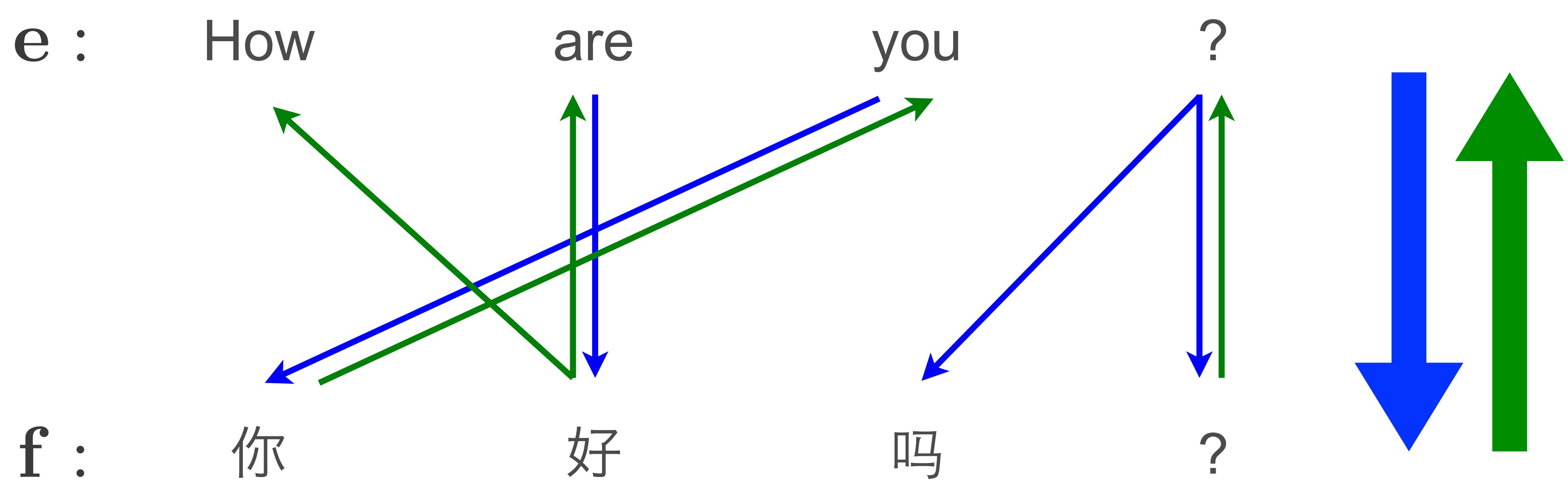
Heuristic symmetrization:



$$\hat{\mathbf{a}} = \max_{\mathbf{a}} \mathrm{P}(\mathbf{f}, \mathbf{a} | \mathbf{e}) \qquad \hat{\mathbf{b}} = \max_{\mathbf{b}} \mathrm{P}(\mathbf{e}, \mathbf{b} | \mathbf{f}) \qquad \mathrm{symm}(\hat{\mathbf{a}}, \hat{\mathbf{b}})$$

(Och et al., 1999)

Heuristic Approaches to Aligner Combination

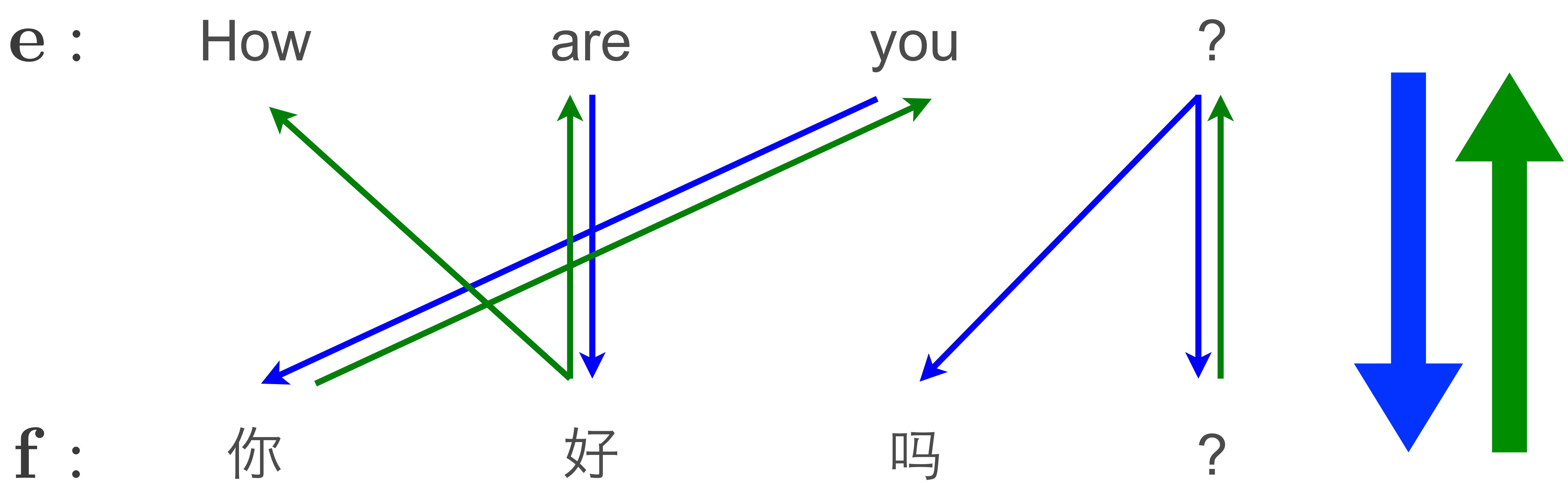


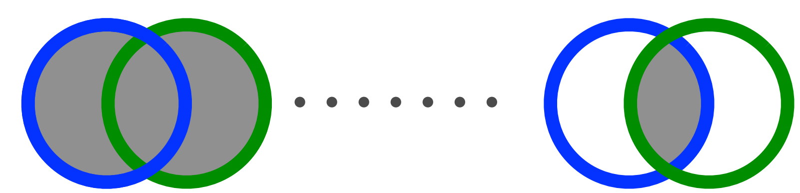
Heuristic symmetrization:

$$\hat{\mathbf{a}} = \max_{\mathbf{a}} \mathrm{P}(\mathbf{f}, \mathbf{a} | \mathbf{e}) \qquad \hat{\mathbf{b}} = \max_{\mathbf{b}} \mathrm{P}(\mathbf{e}, \mathbf{b} | \mathbf{f}) \qquad \mathrm{symm}(\hat{\mathbf{a}}, \hat{\mathbf{b}})$$

(Och et al., 1999)

Heuristic Approaches to Aligner Combination



Heuristic symmetrization: 

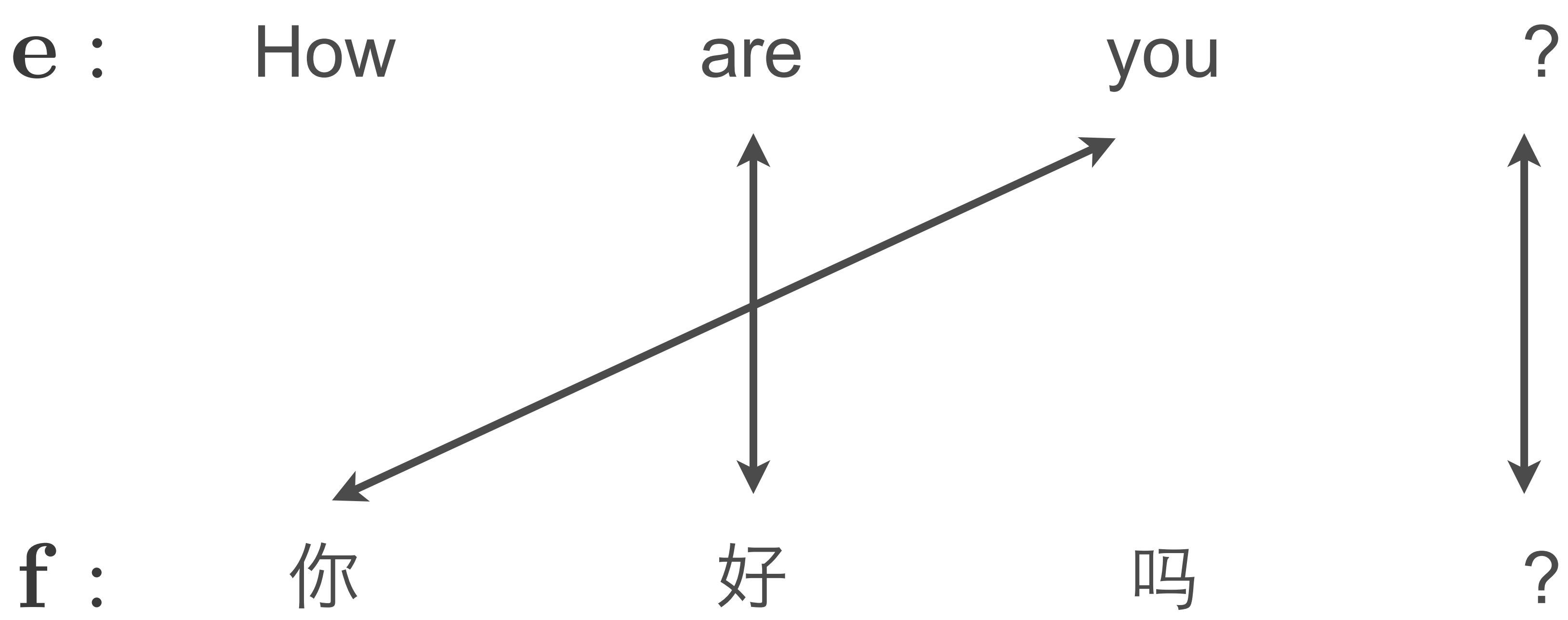
$\hat{\mathbf{a}} = \max_{\mathbf{a}} \mathrm{P}(\mathbf{f}, \mathbf{a} | \mathbf{e})$ $\hat{\mathbf{b}} = \max_{\mathbf{b}} \mathrm{P}(\mathbf{e}, \mathbf{b} | \mathbf{f})$ $\mathrm{symm}(\hat{\mathbf{a}}, \hat{\mathbf{b}})$

(Och et al., 1999)

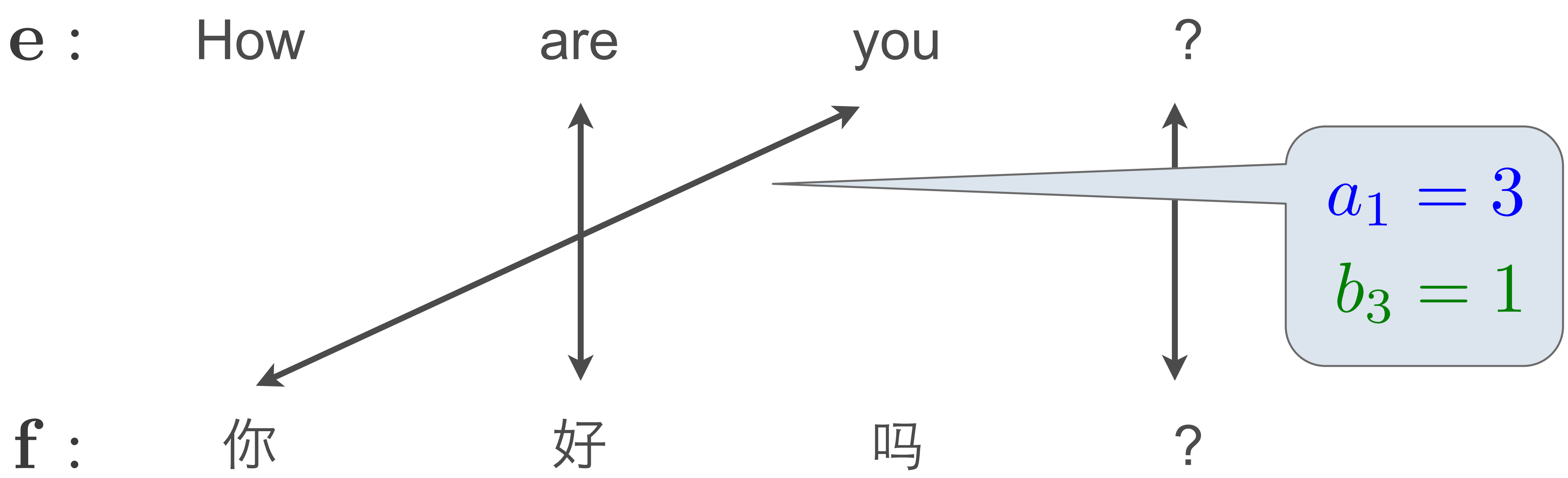
A Joint Objective for Combination

$\mathbf{e}$:	How	are	you	?
$\mathbf{f}$:	你	好	吗	?

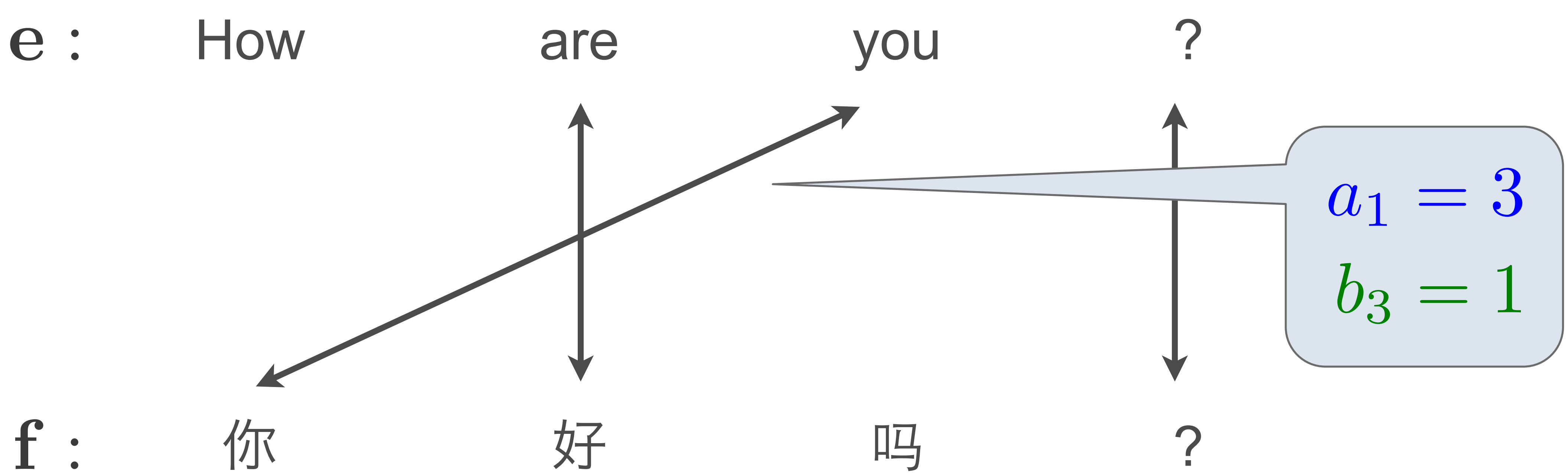
A Joint Objective for Combination



A Joint Objective for Combination

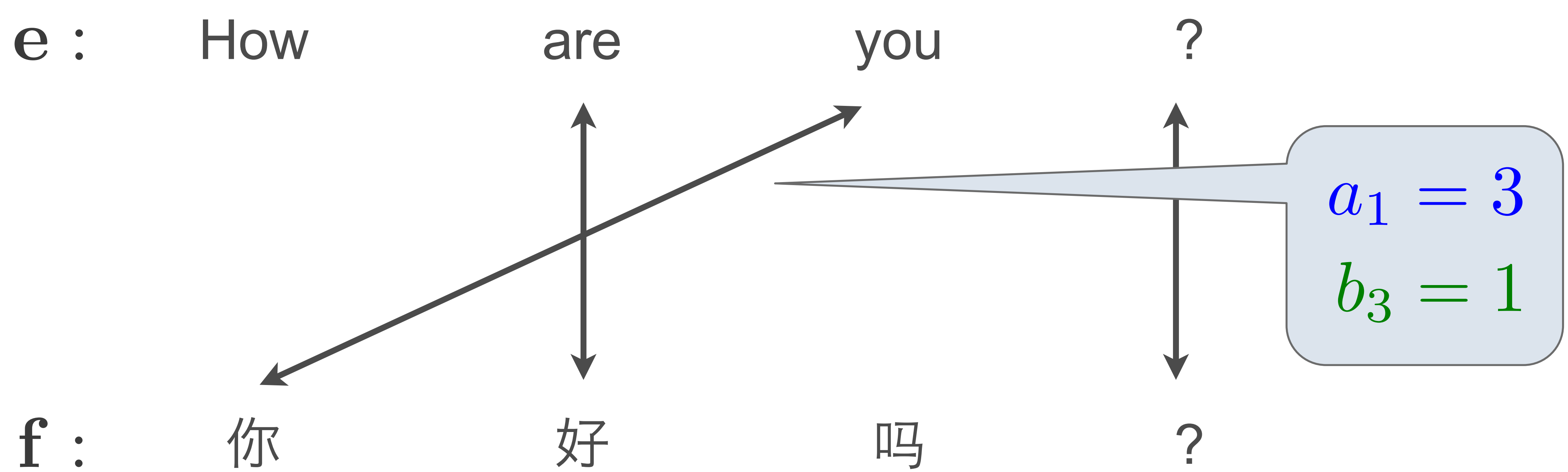


A Joint Objective for Combination



Proposal : $$\hat{\mathbf{a}} = \max_{\mathbf{a}} \left[\underset{\mathbf{e}\to\mathbf{f}}{\mathrm{P}}(\mathbf{f}, \mathbf{a}|\mathbf{e}) \cdot \underset{\mathbf{f}\to\mathbf{e}}{\mathrm{P}}(\mathbf{e}, \mathrm{inv}(\mathbf{a})|\mathbf{f}) \right]$$

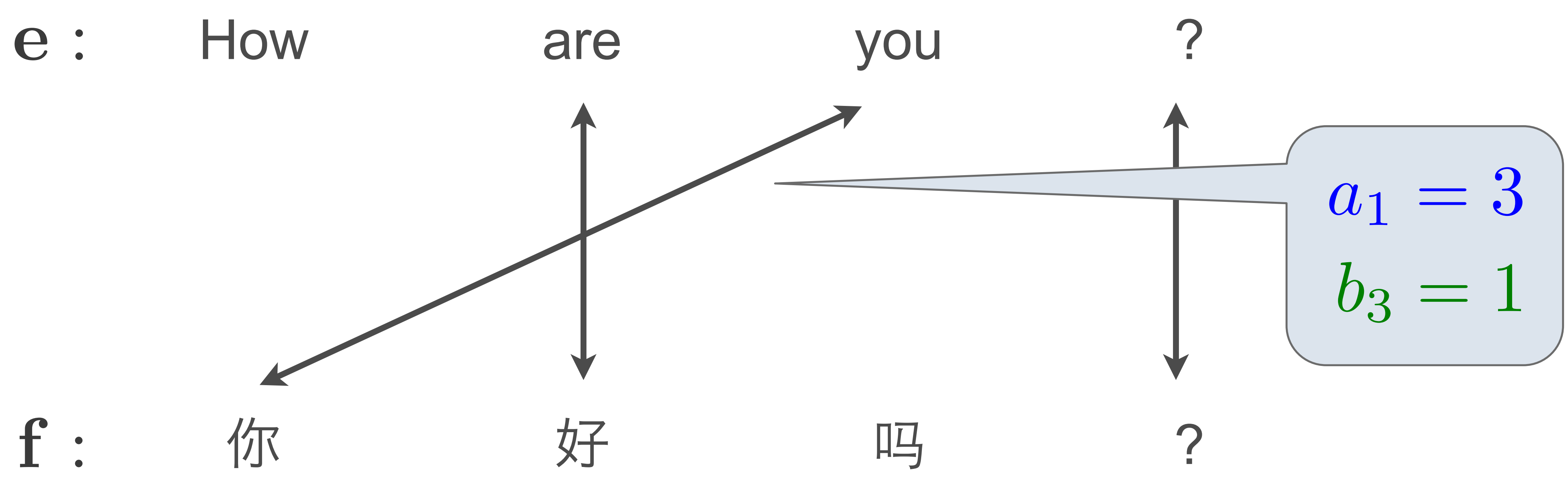
A Joint Objective for Combination



Proposal : $\hat{\mathbf{a}} = \max_{\mathbf{a}} \left[\underset{\mathbf{e}\to\mathbf{f}}{\mathrm{P}}(\mathbf{f}, \mathbf{a} | \mathbf{e}) \cdot \underset{\mathbf{f}\to\mathbf{e}}{\mathrm{P}}(\mathbf{e}, \mathrm{inv}(\mathbf{a}) | \mathbf{f}) \right]$

Issues :

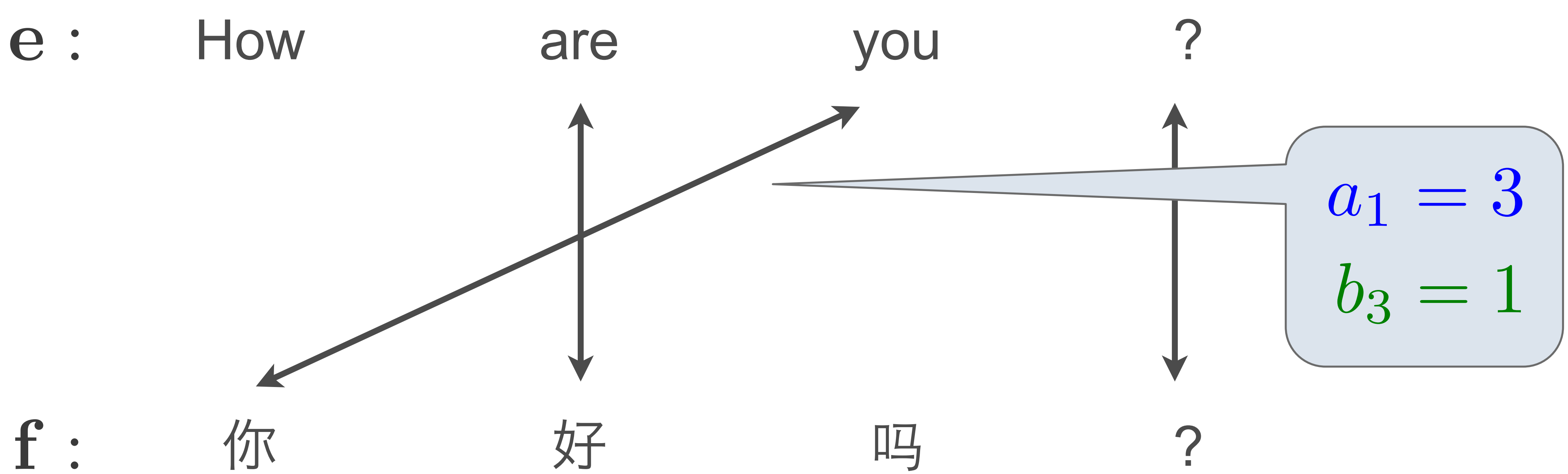
A Joint Objective for Combination



Proposal : $\hat{\mathbf{a}} = \max_{\mathbf{a}} \left[\underset{\mathbf{e}\to\mathbf{f}}{\mathrm{P}}(\mathbf{f}, \mathbf{a} | \mathbf{e}) \cdot \underset{\mathbf{f}\to\mathbf{e}}{\mathrm{P}}(\mathbf{e}, \mathrm{inv}(\mathbf{a}) | \mathbf{f}) \right]$

Issues : One-to-one

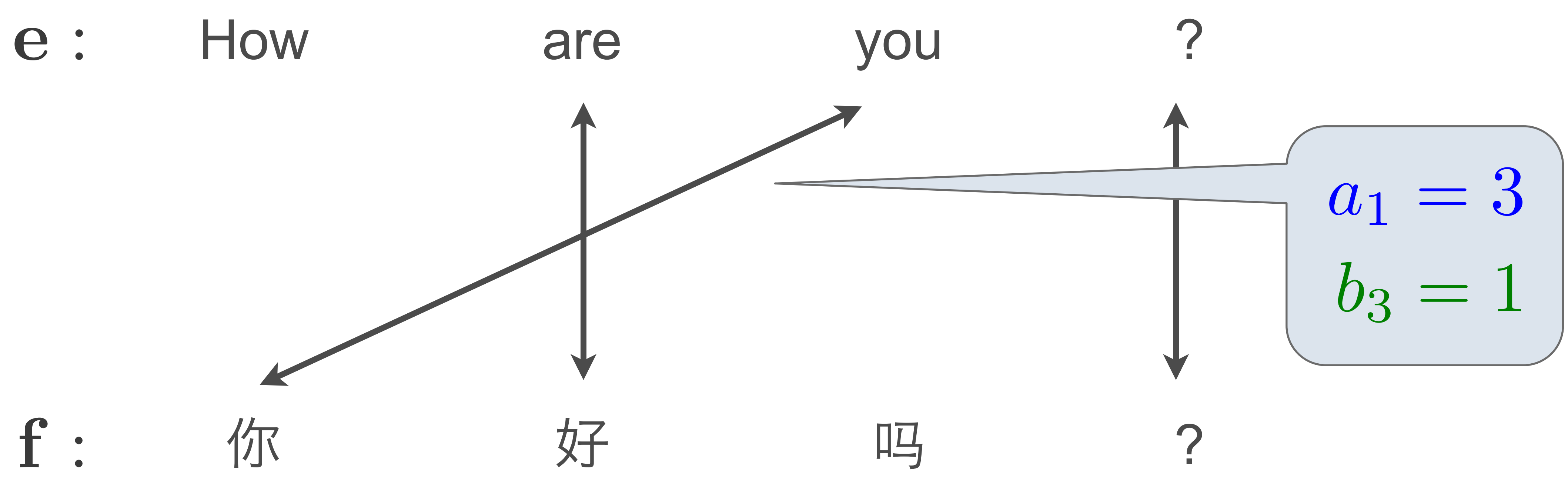
A Joint Objective for Combination



Proposal : $$\hat{\mathbf{a}} = \max_{\mathbf{a}} \left[\underset{\mathbf{e}\to\mathbf{f}}{\mathrm{P}}(\mathbf{f}, \mathbf{a} | \mathbf{e}) \cdot \underset{\mathbf{f}\to\mathbf{e}}{\mathrm{P}}(\mathbf{e}, \mathrm{inv}(\mathbf{a}) | \mathbf{f}) \right]$$

Issues : One-to-one

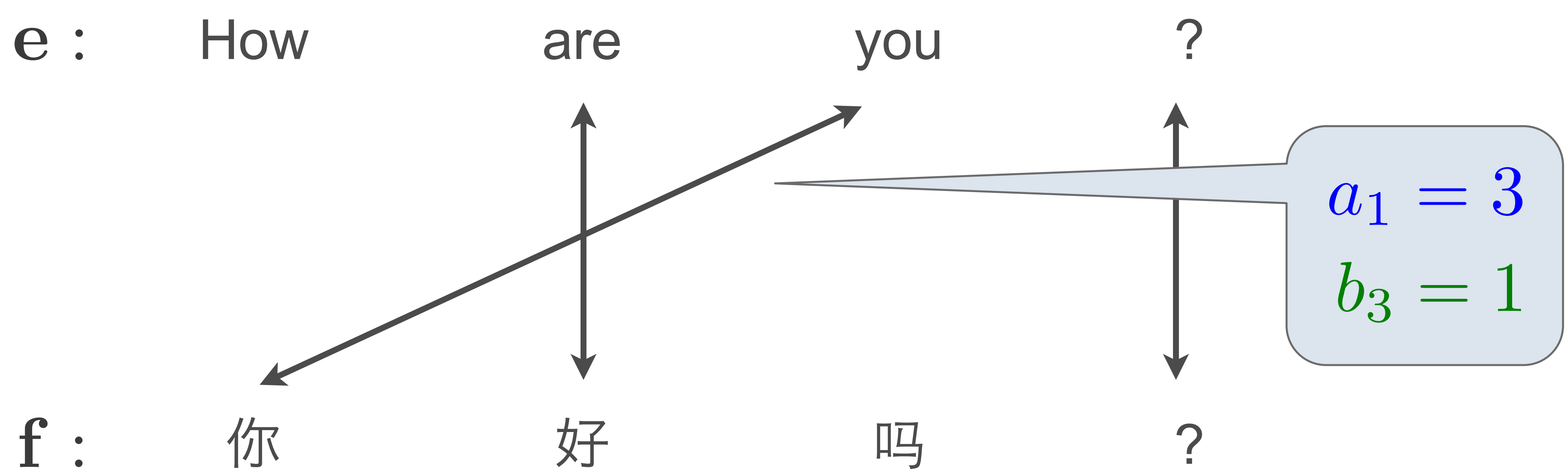
A Joint Objective for Combination



Proposal : $$\hat{\mathbf{a}} = \max_{\mathbf{a}} \left[\underset{\mathbf{e}\to\mathbf{f}}{\mathrm{P}}(\mathbf{f}, \mathbf{a}|\mathbf{e}) \cdot \underset{\mathbf{f}\to\mathbf{e}}{\mathrm{P}}(\mathbf{e}, \mathrm{inv}(\mathbf{a})|\mathbf{f}) \right]$$

Issues : One-to-one Hard inference

A Joint Objective for Combination



Proposal : $$\hat{\mathbf{a}} = \max_{\mathbf{a}} \left[\underset{\mathbf{e}\rightarrow\mathbf{f}}{\mathrm{P}}(\mathbf{f}, \mathbf{a}|\mathbf{e}) \cdot \underset{\mathbf{f}\rightarrow\mathbf{e}}{\mathrm{P}}(\mathbf{e}, \mathrm{inv}(\mathbf{a})|\mathbf{f}) \right]$$

Issues : One-to-one Hard inference

Combination as a Graphical Model

Probability of complete assignment ∝ product of node and edge potentials

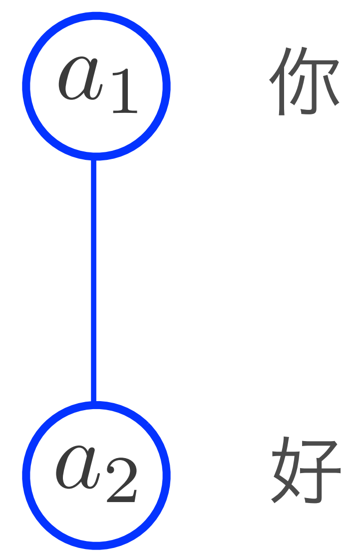
你

好

How are you

Combination as a Graphical Model

Probability of complete assignment ∝ product of node and edge potentials



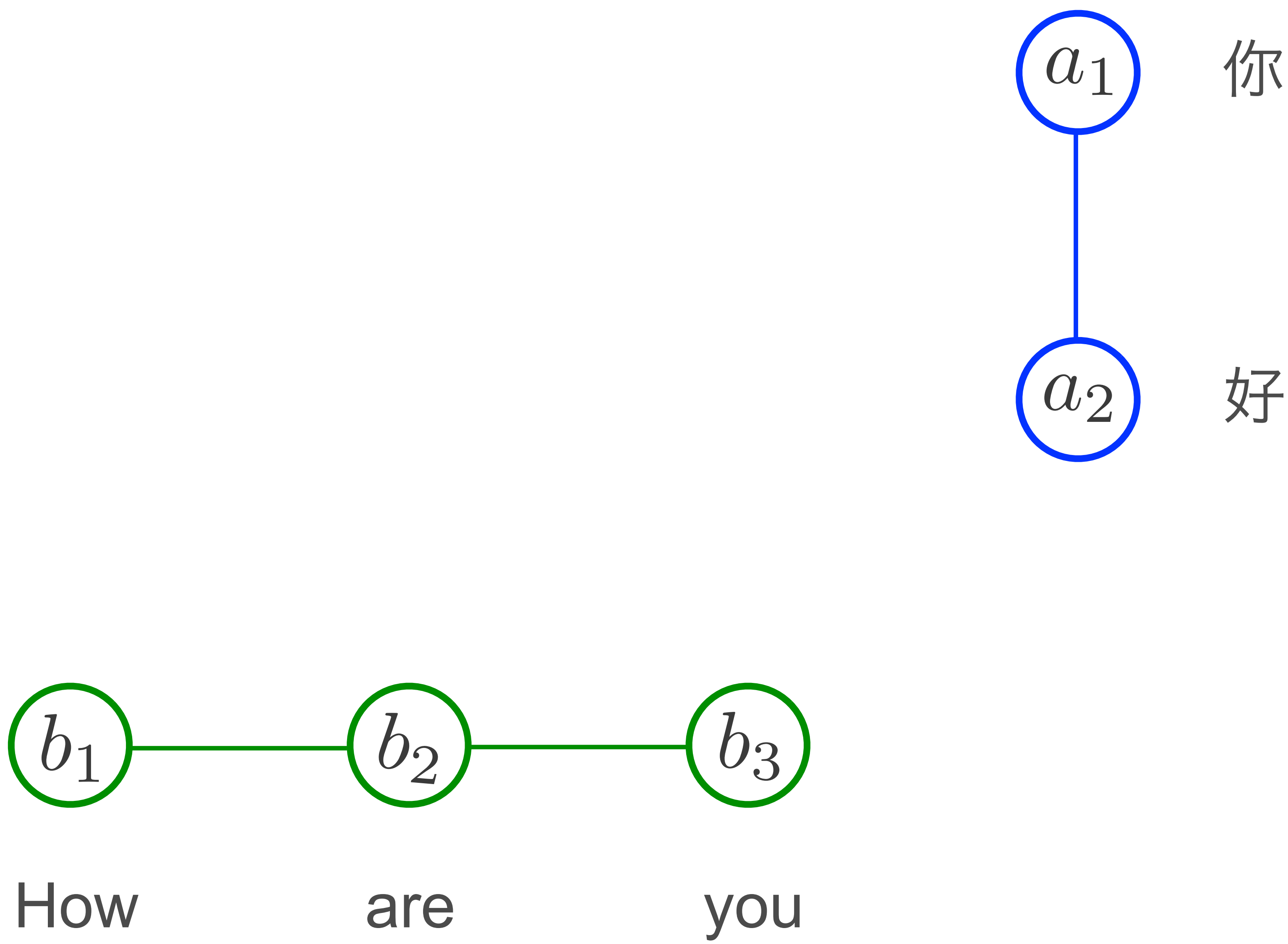
$a_j \in \{0, 1, 2, 3\}$

E→F HMM Variables

How are you

Combination as a Graphical Model

Probability of complete assignment ∝ product of node and edge potentials



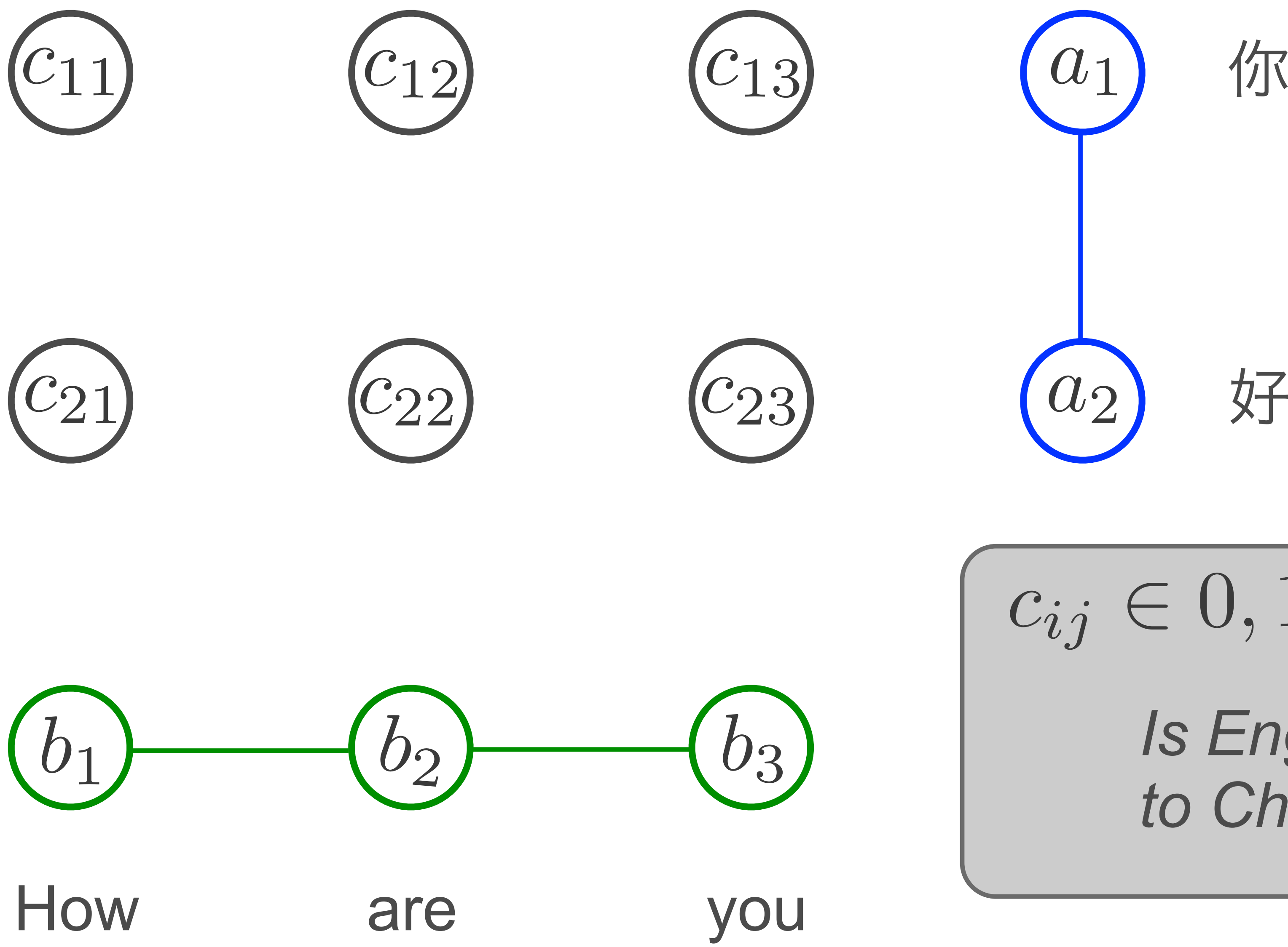
$a_j \in \{0, 1, 2, 3\}$

E→F HMM Variables

$b_i \in \{0, 1, 2\}$ *F→E HMM Variables*

Combination as a Graphical Model

Probability of complete assignment ∝ product of node and edge potentials



$a_j \in \{0, 1, 2, 3\}$

E→F HMM Variables

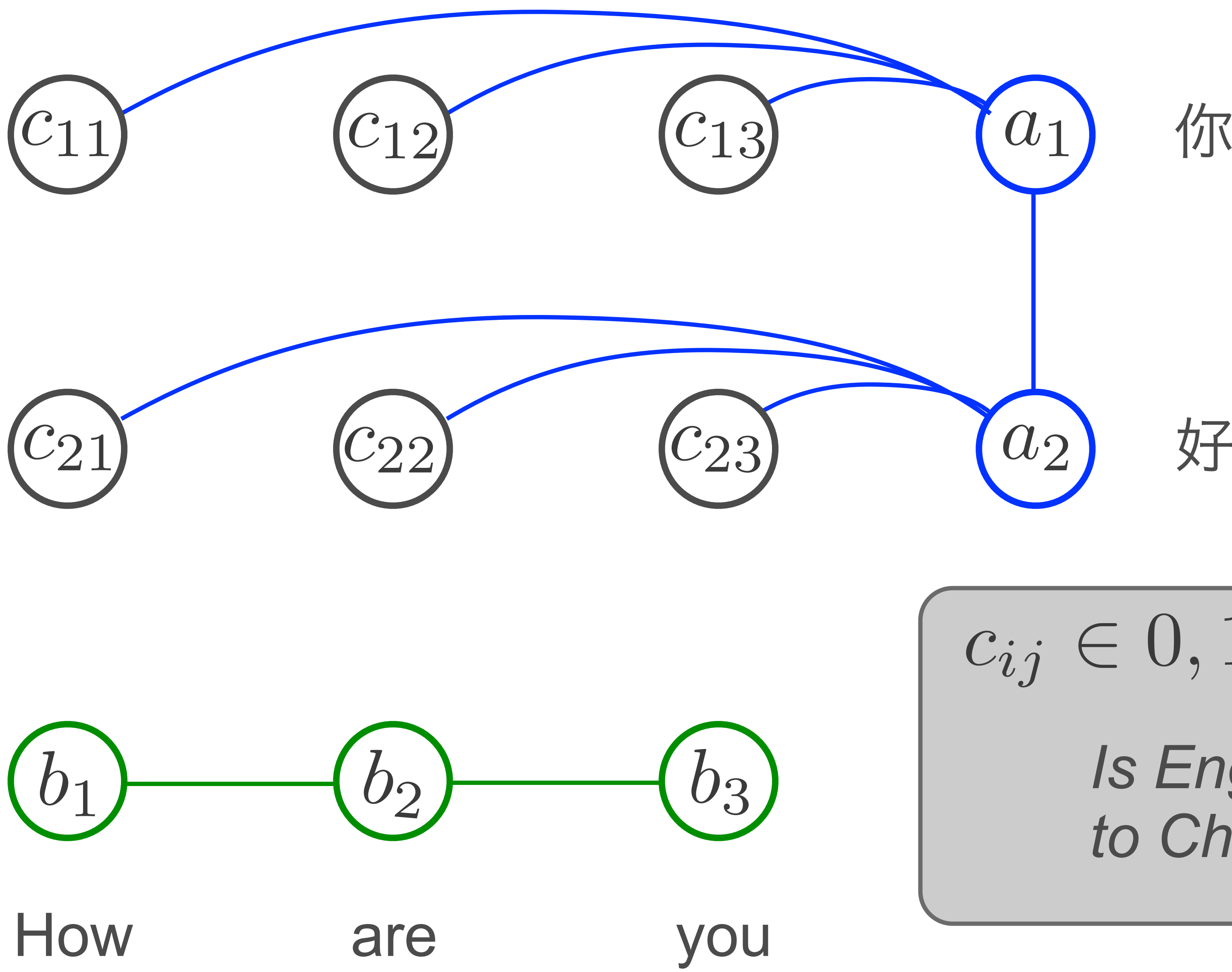
$c_{ij} \in 0, 1$

Is English word i aligned to Chinese word j

$b_i \in \{0, 1, 2\}$ *F→E HMM Variables*

Combination as a Graphical Model

Probability of complete assignment ∝ product of node and edge potentials



$a_j \in \{0, 1, 2, 3\}$

E→F HMM Variables

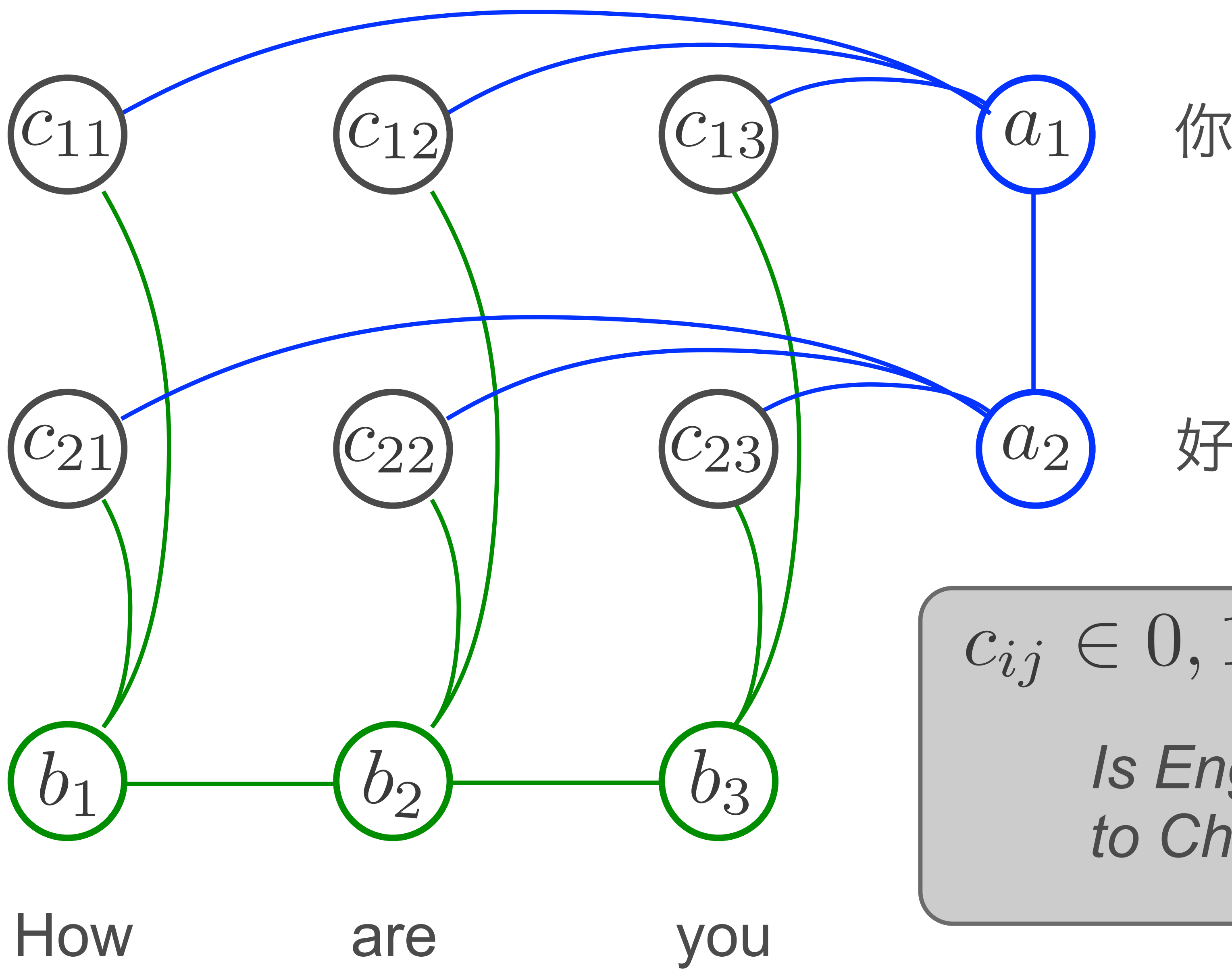
$c_{ij} \in 0, 1$

Is English word i aligned to Chinese word j

$b_i \in \{0, 1, 2\}$ *F→E HMM Variables*

Combination as a Graphical Model

Probability of complete assignment ∝ product of node and edge potentials



$a_j \in \{0, 1, 2, 3\}$

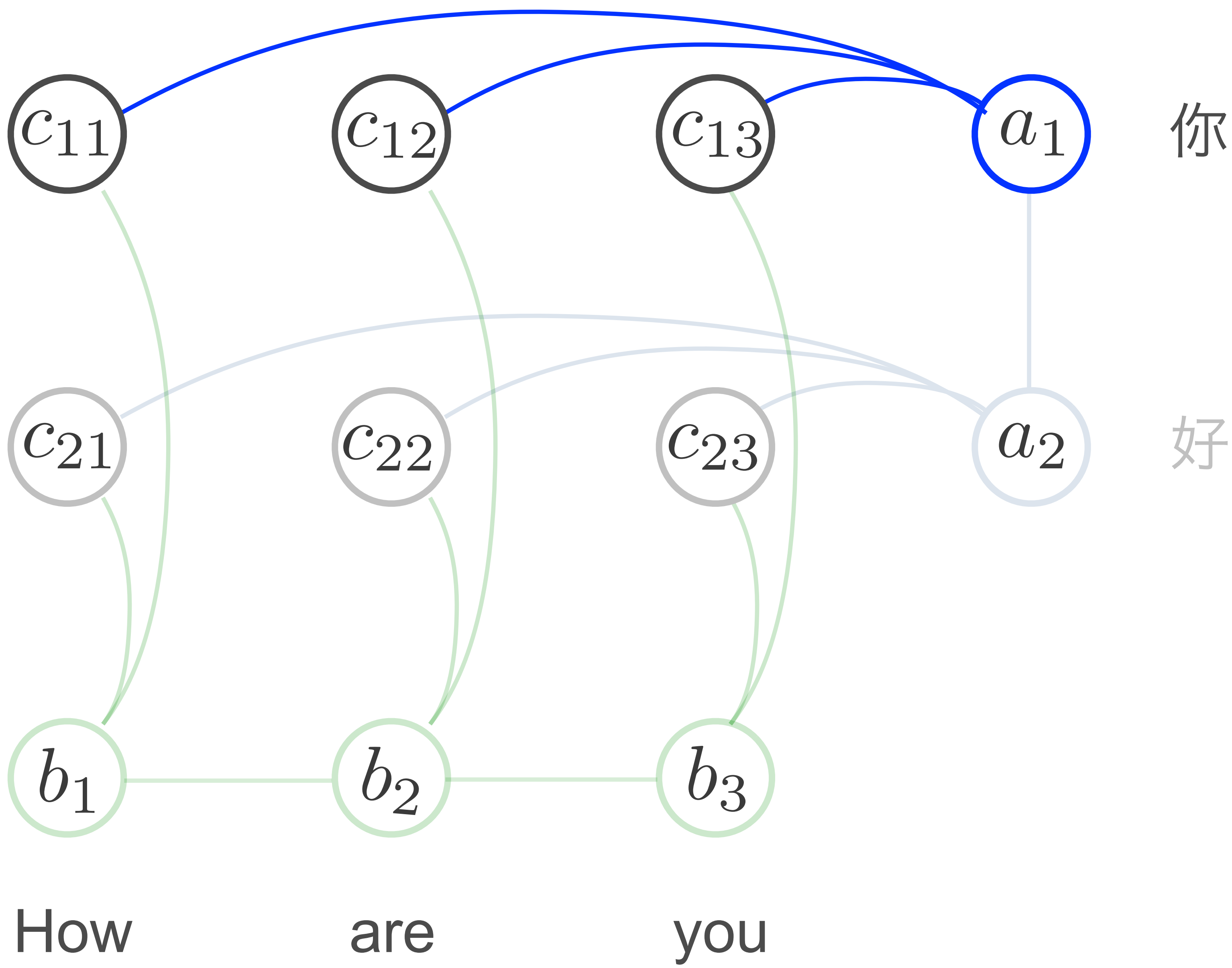
E→F HMM Variables

$c_{ij} \in 0, 1$

Is English word i aligned to Chinese word j

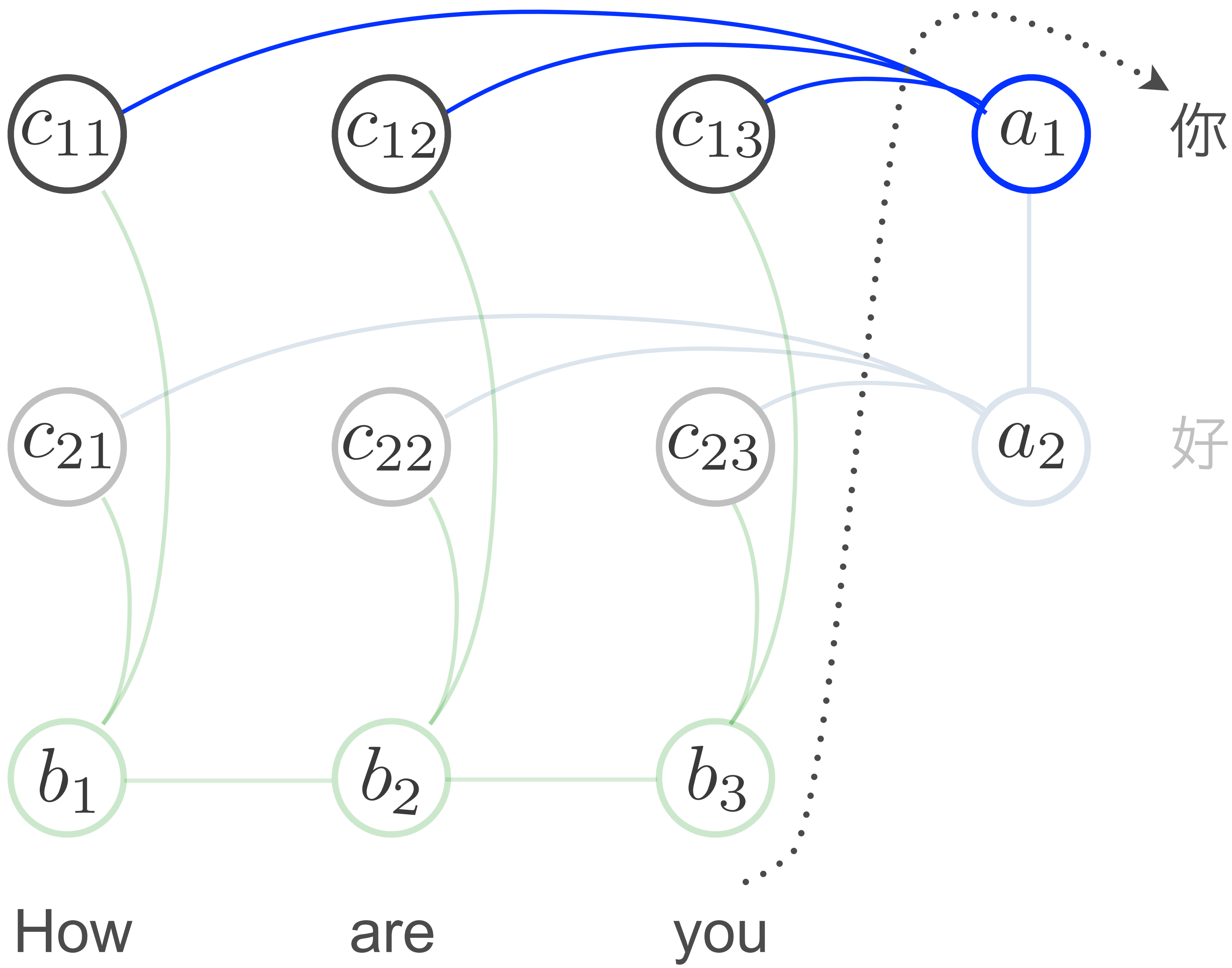
$b_i \in \{0, 1, 2\}$ *F→E HMM Variables*

Allowing Phrasal Alignments

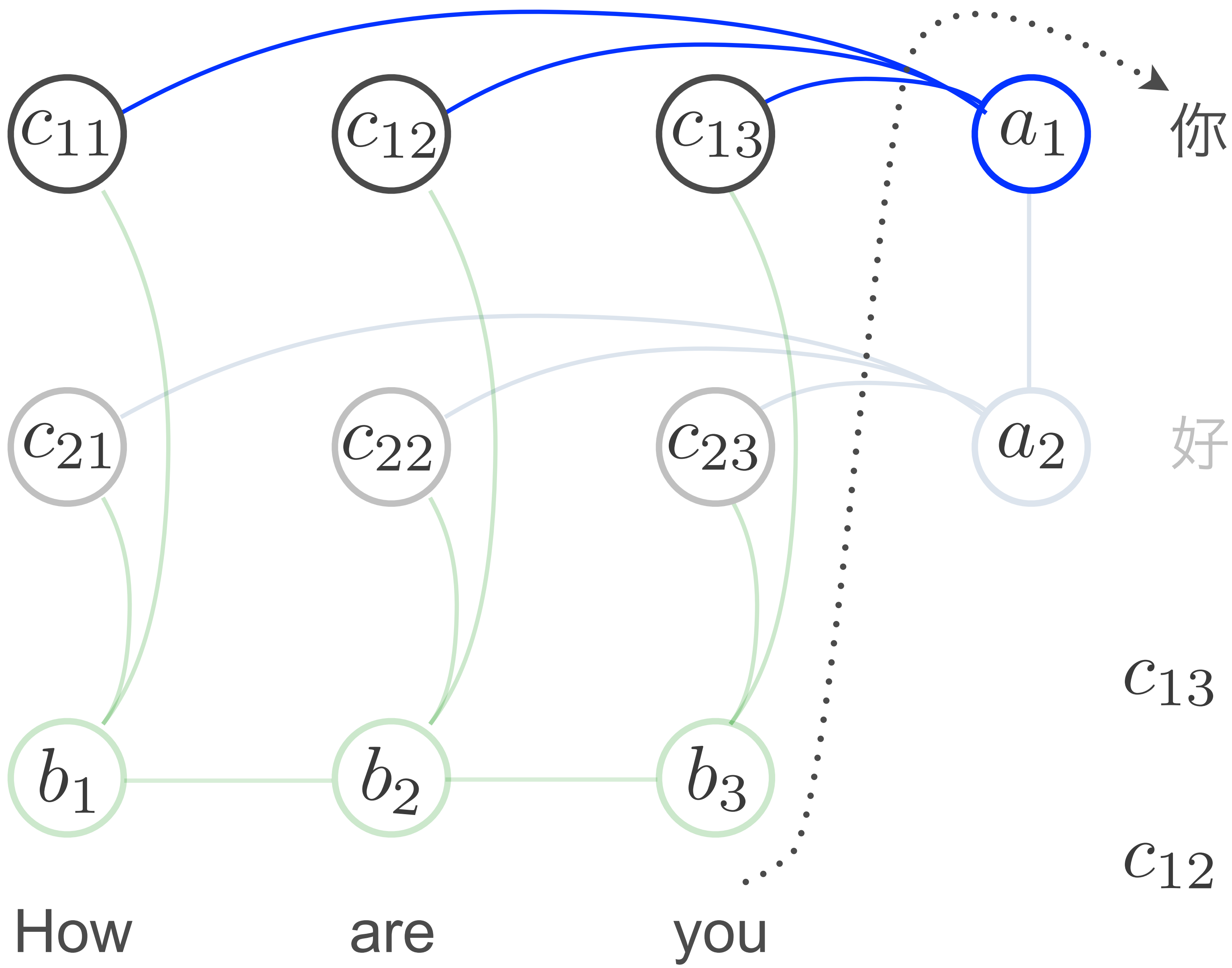


Allowing Phrasal Alignments

Google



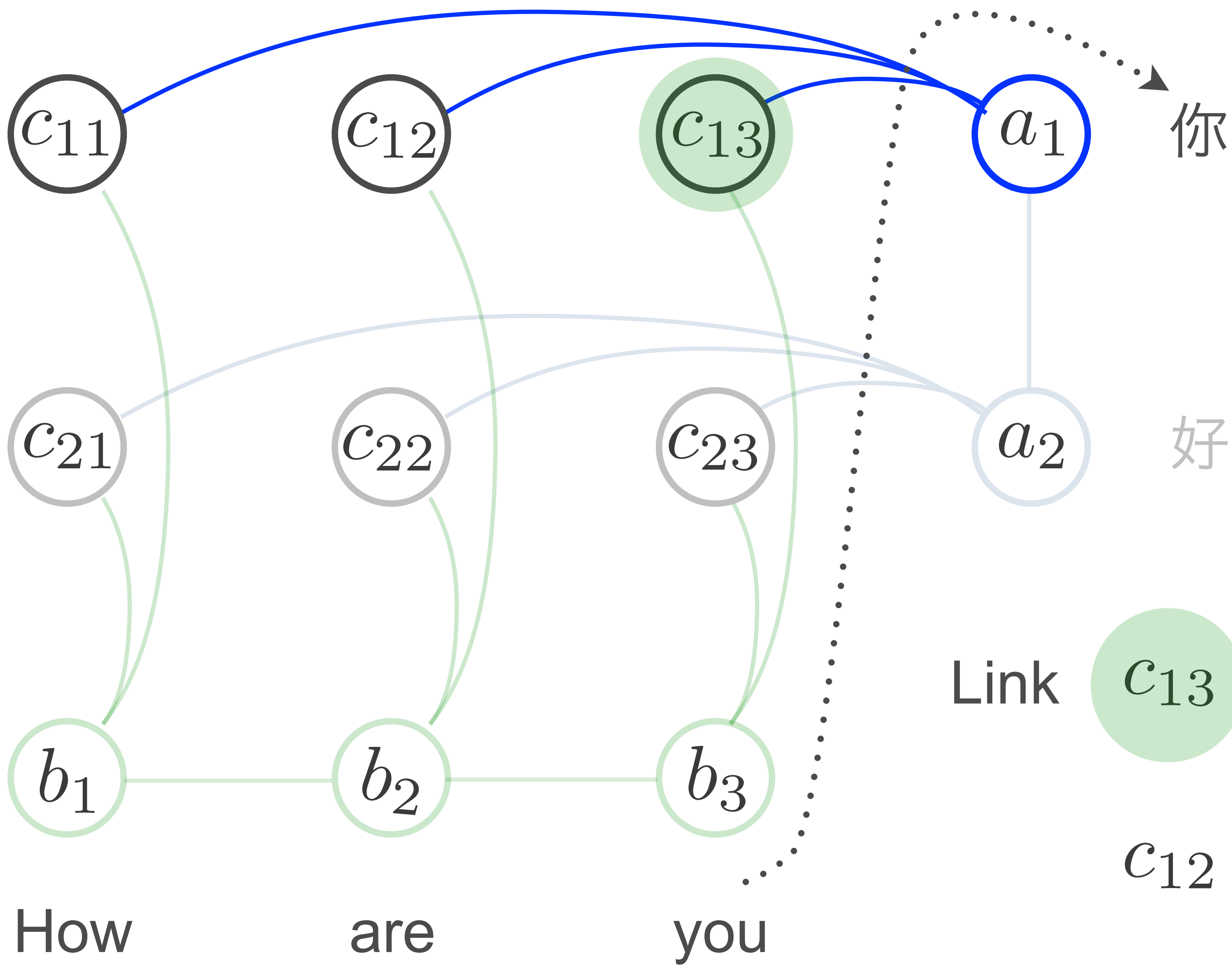
Allowing Phrasal Alignments



$$a_1 = 3; c =$$

	0	1
c_{13}		
c_{12}		
c_{11}		

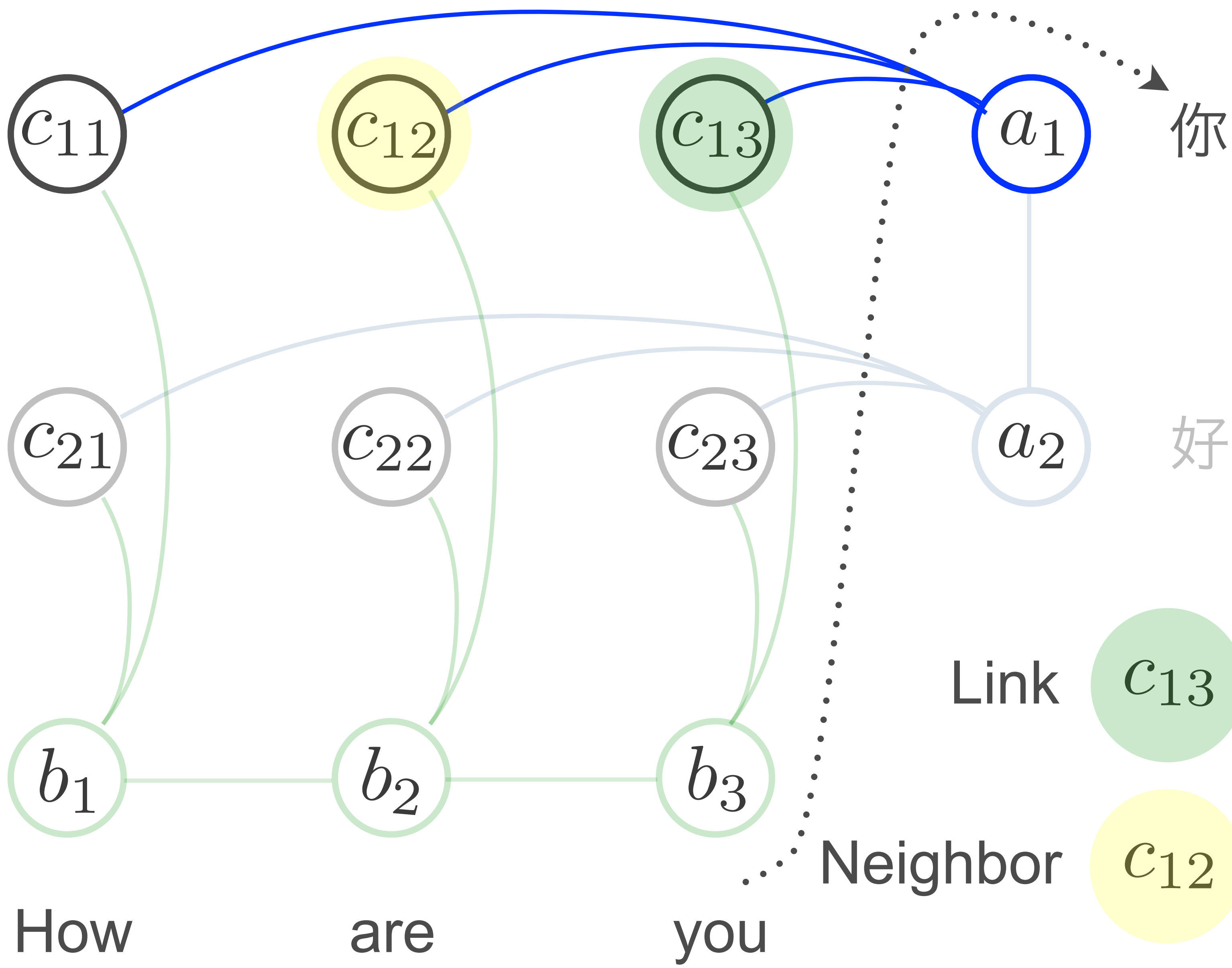
Allowing Phrasal Alignments



$$a_1 = 3; c =$$

	0	1
Link c_{13}	0	1
c_{12}		
c_{11}		

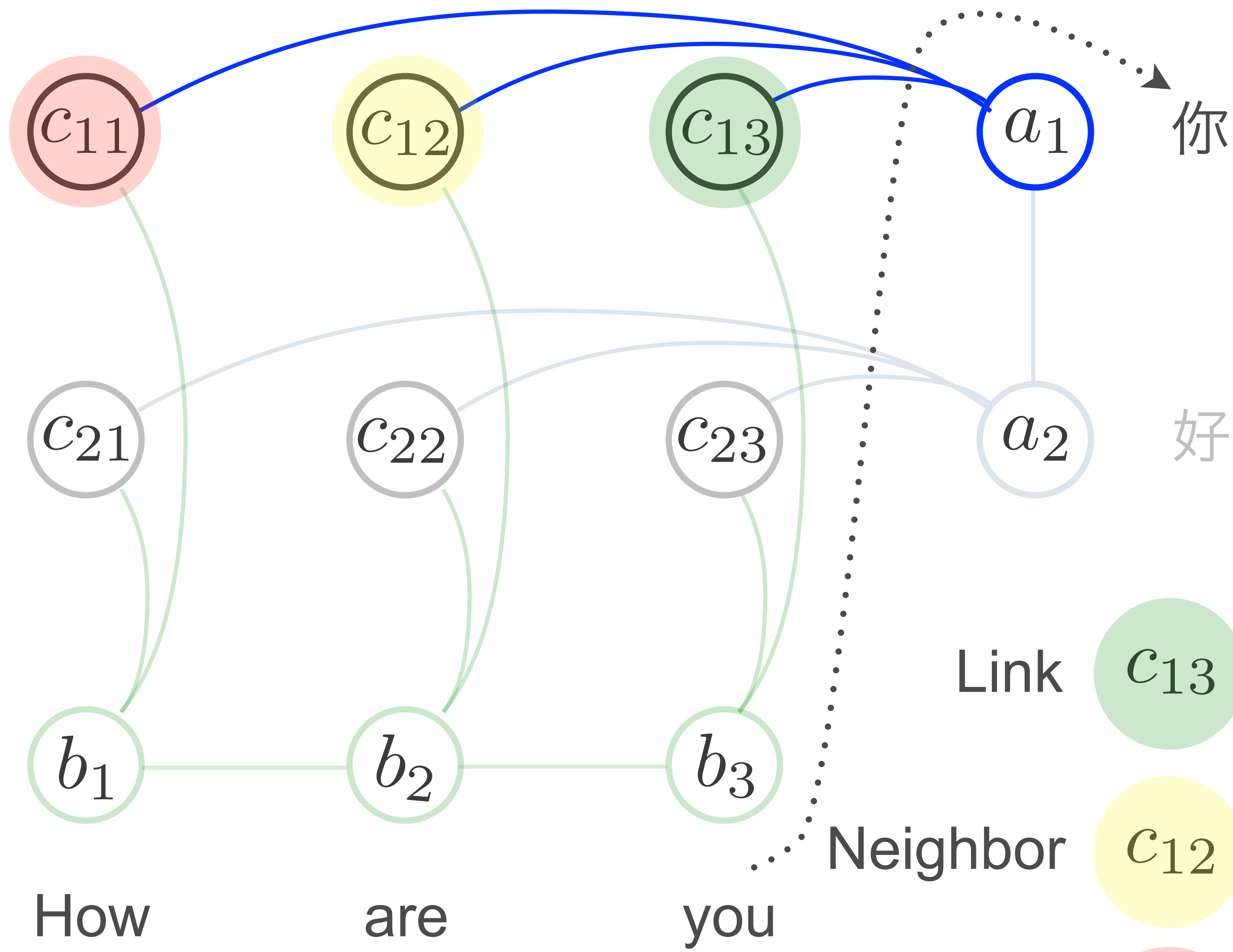
Allowing Phrasal Alignments



$$a_1 = 3; c =$$

	0	1
Link c_{13}	0	1
Neighbor c_{12}	1	$e^{-\alpha}$
c_{11}		

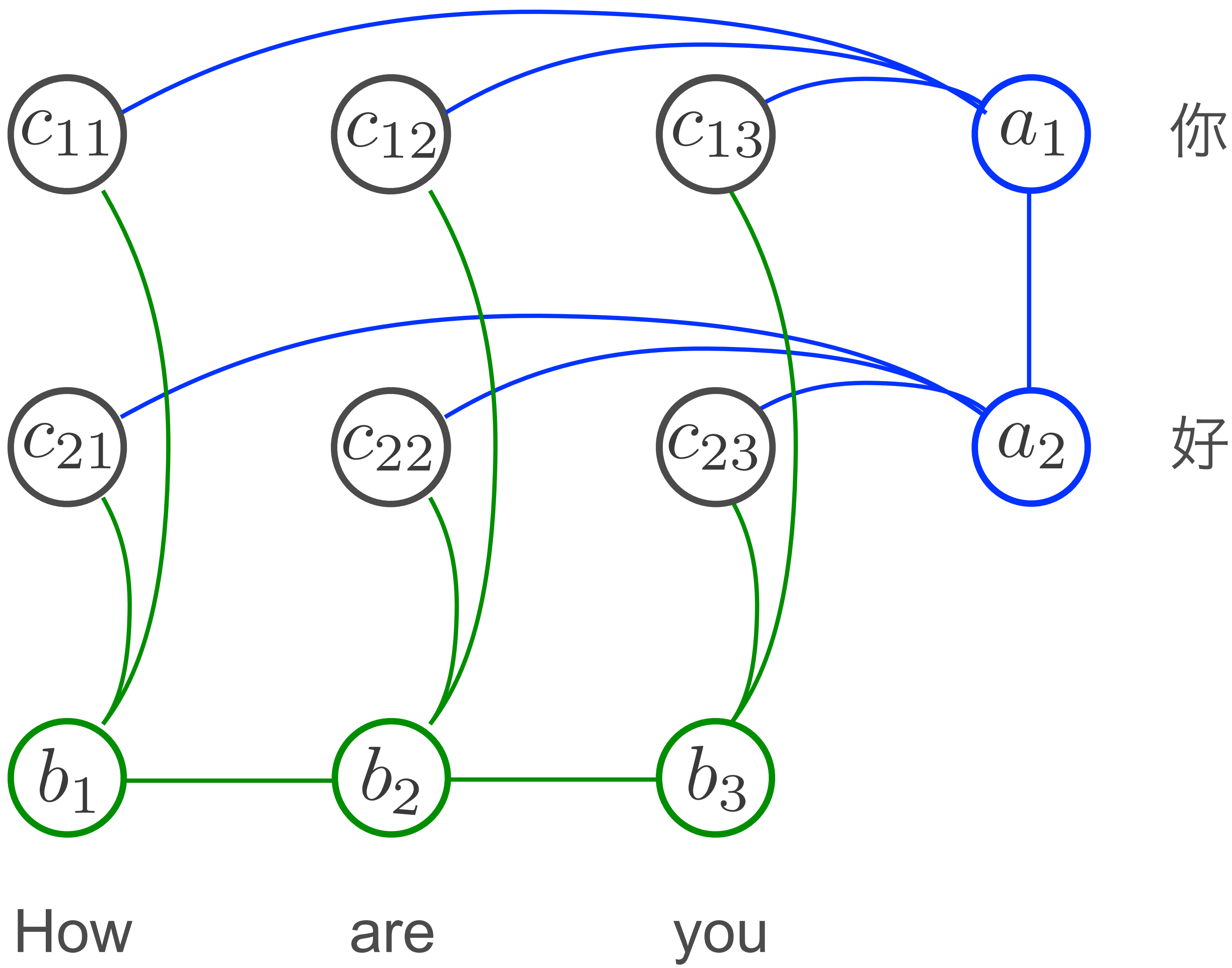
Allowing Phrasal Alignments



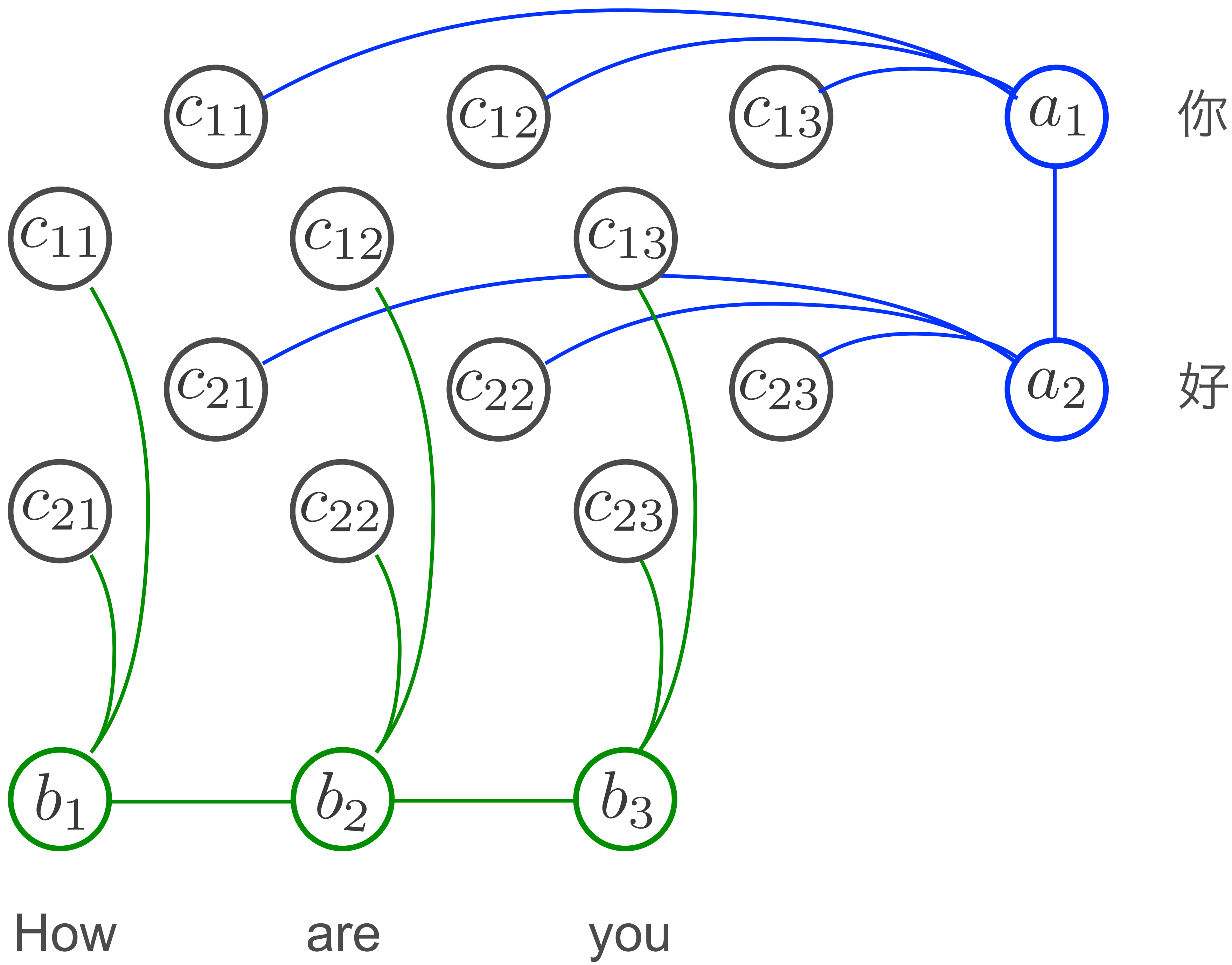
$$a_1 = 3; c =$$

	0	1
Link c_{13}	0	1
Neighbor c_{12}	1	$e^{-\alpha}$
Other c_{11}	1	0

Subproblem Decomposition

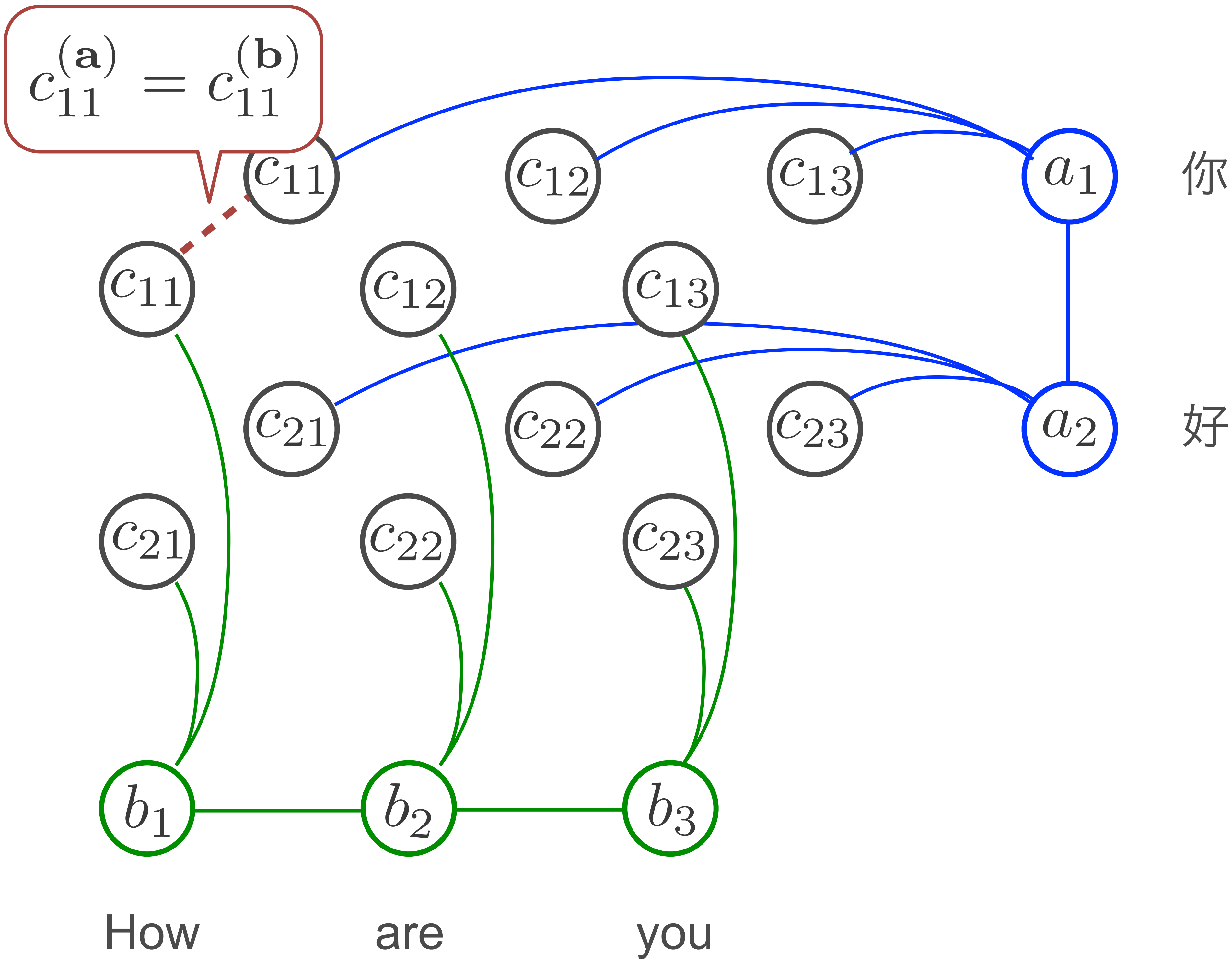


Subproblem Decomposition



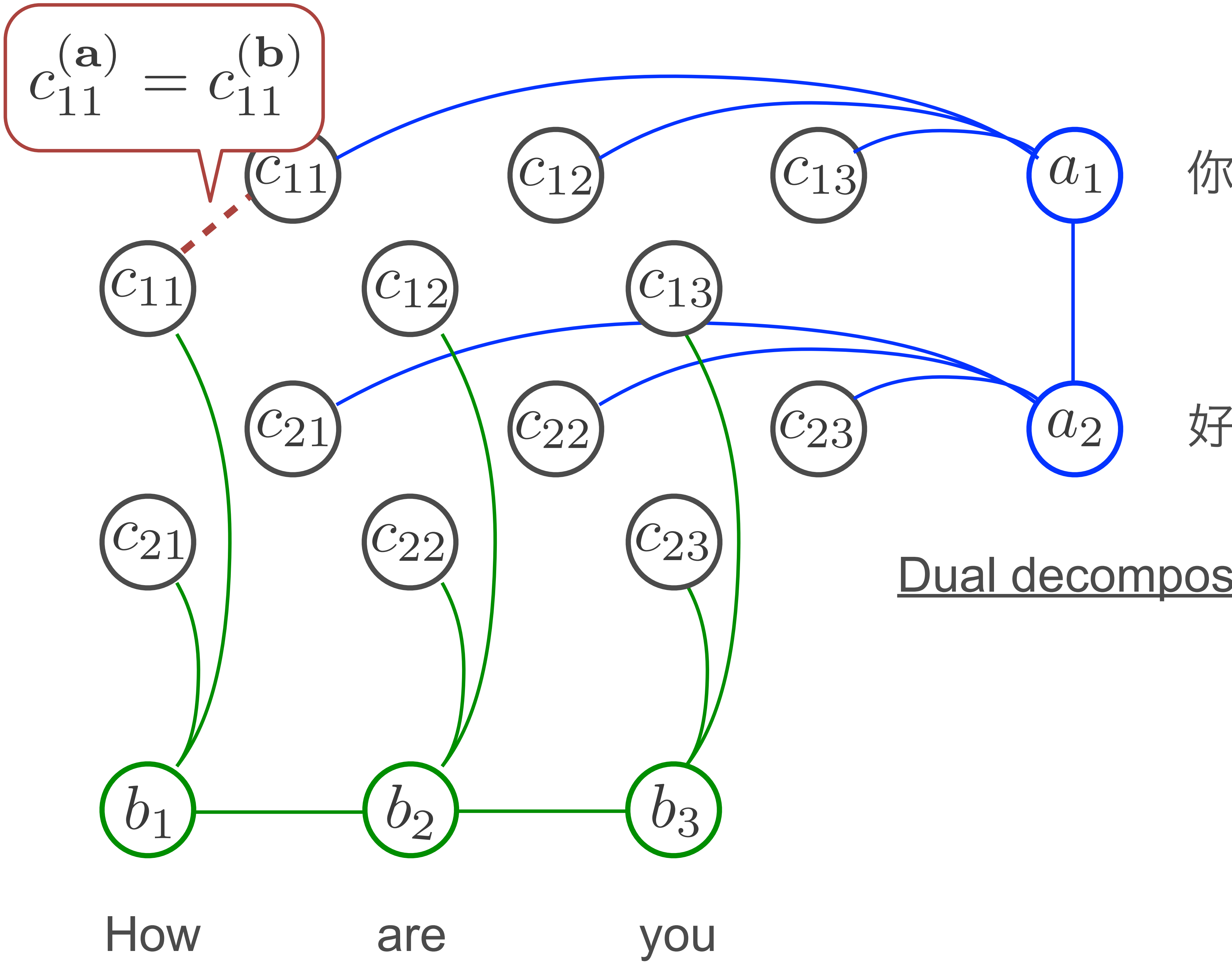
(Rush et al., 2010)

Subproblem Decomposition



(Rush et al., 2010)

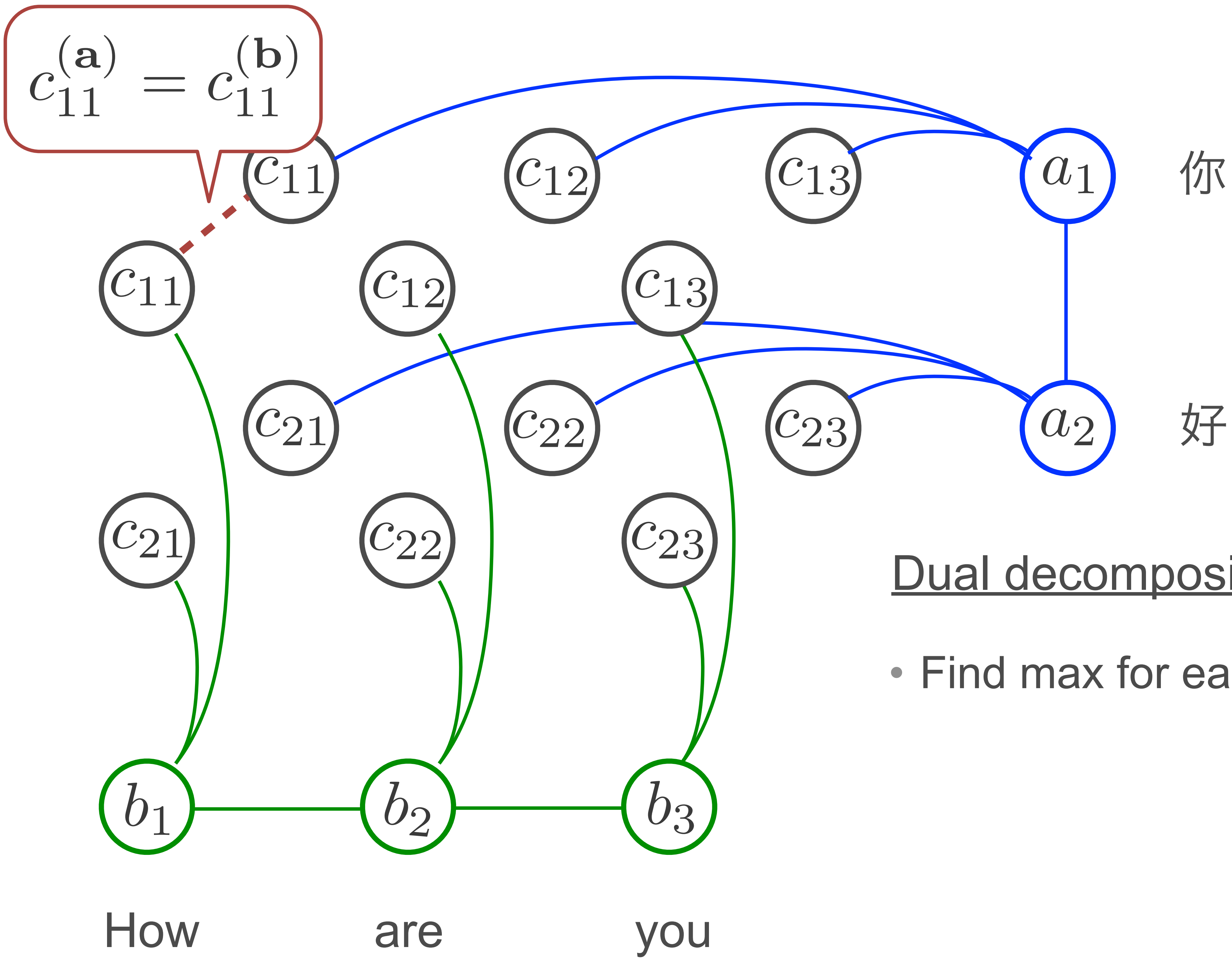
Subproblem Decomposition



Dual decomposition inference

(Rush et al., 2010)

Subproblem Decomposition

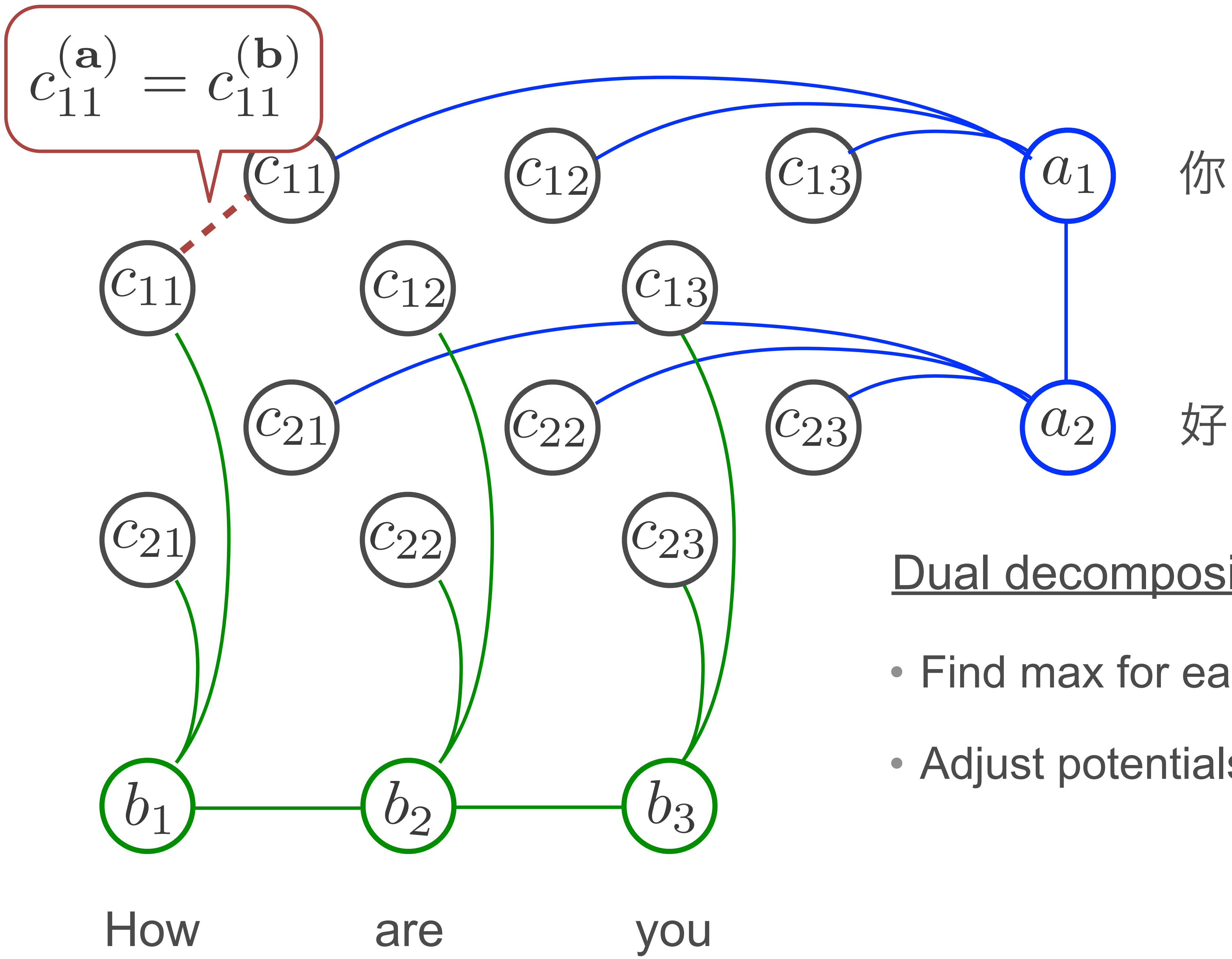


Dual decomposition inference

- Find max for each subproblem

(Rush et al., 2010)

Subproblem Decomposition

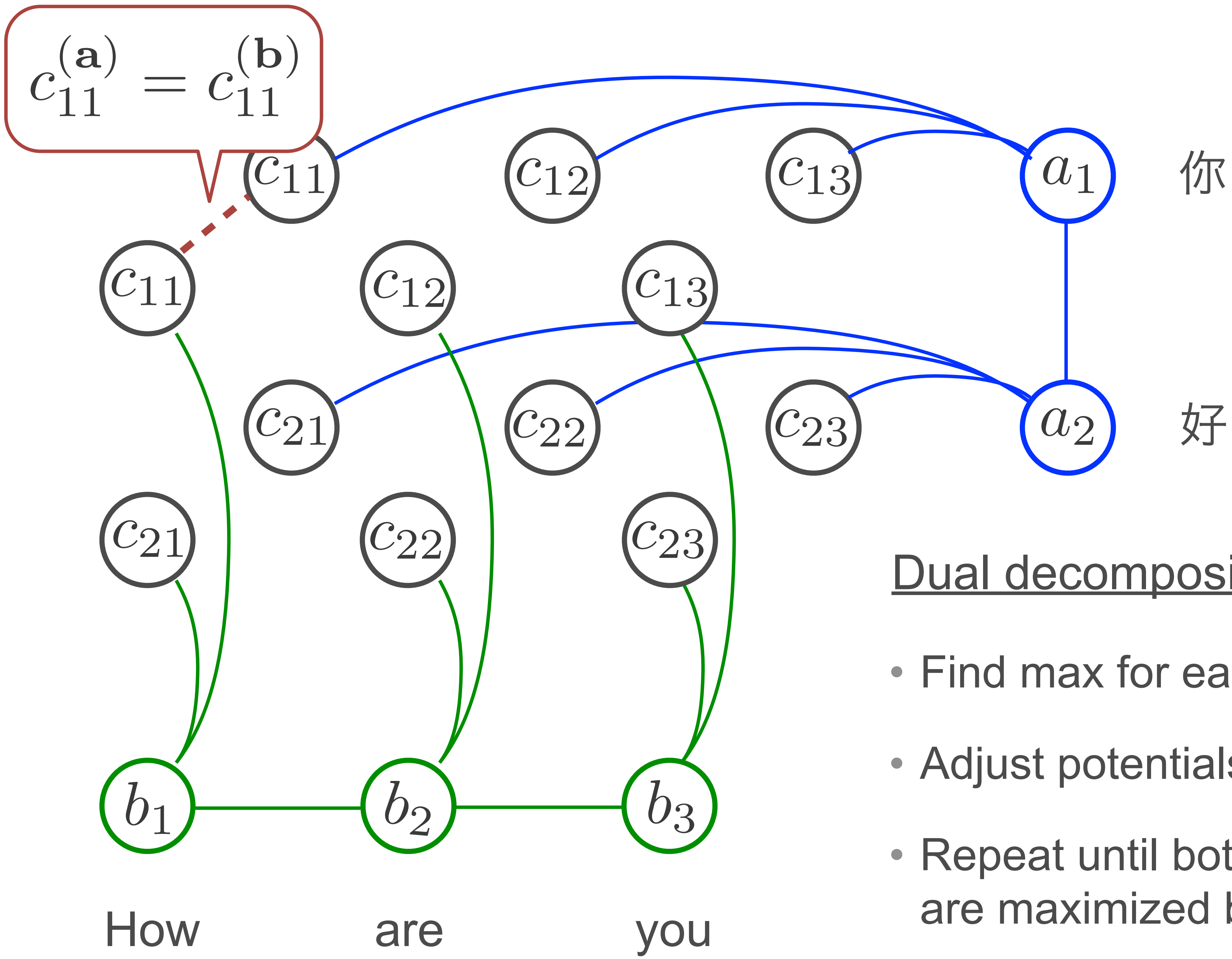


<u>Dual decomposition inference</u>

- Find max for each subproblem
- Adjust potentials for unequal c_{ij}

(Rush et al., 2010)

Subproblem Decomposition

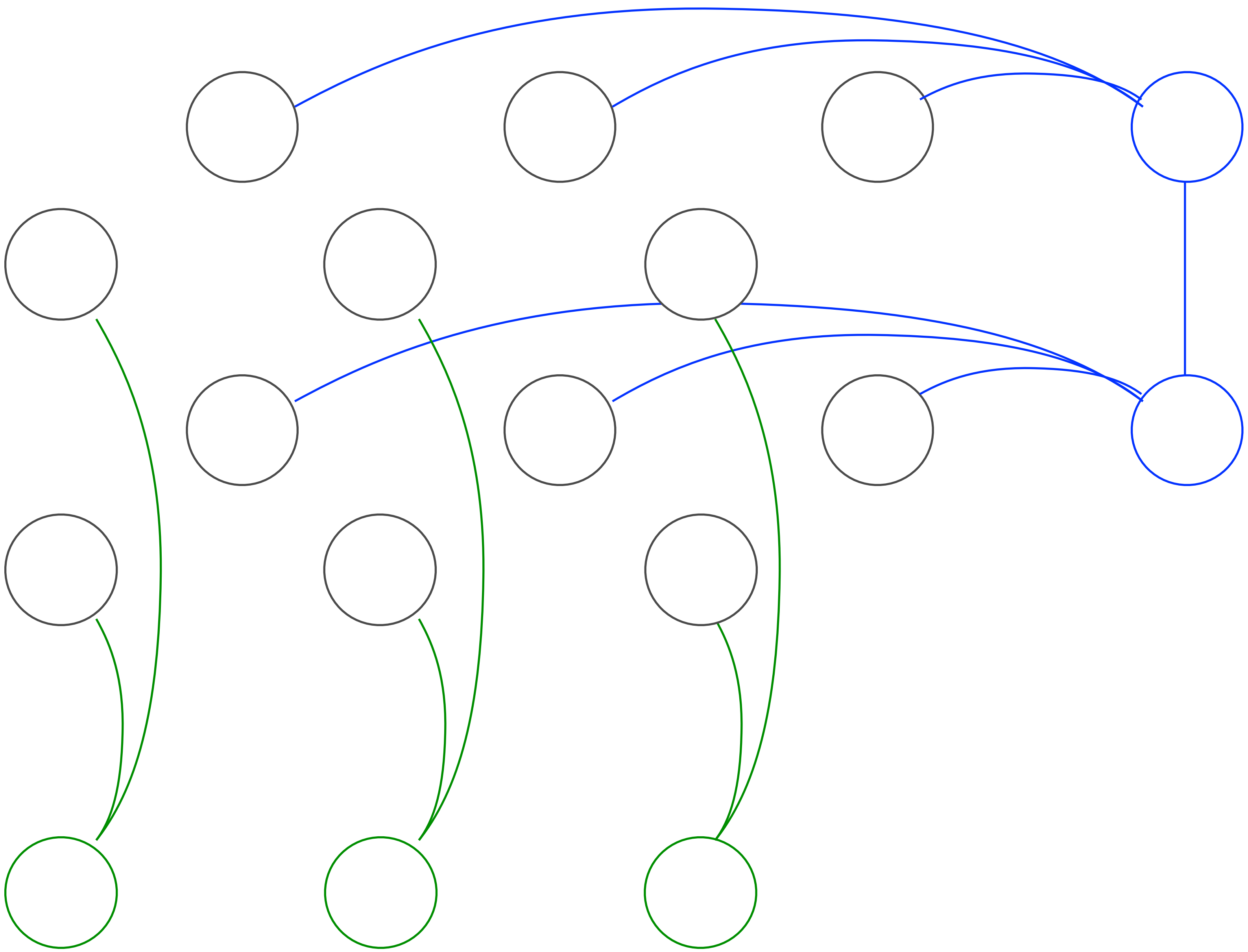


<u>Dual decomposition inference</u>

- Find max for each subproblem
- Adjust potentials for unequal c_{ij}
- Repeat until both subproblems are maximized by the same c

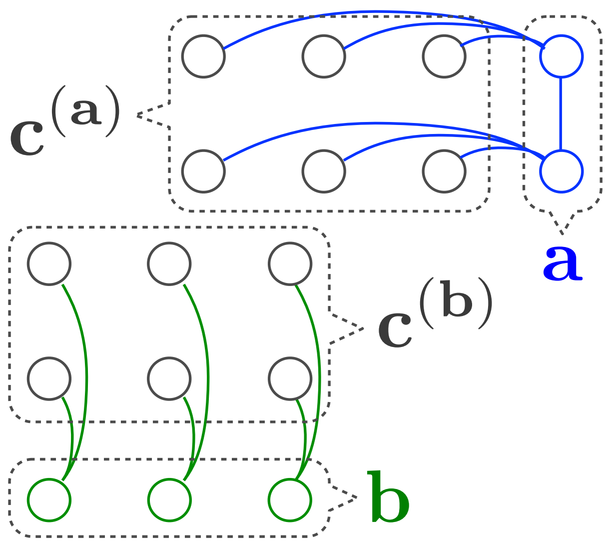
(Rush et al., 2010)

Subproblem Decomposition



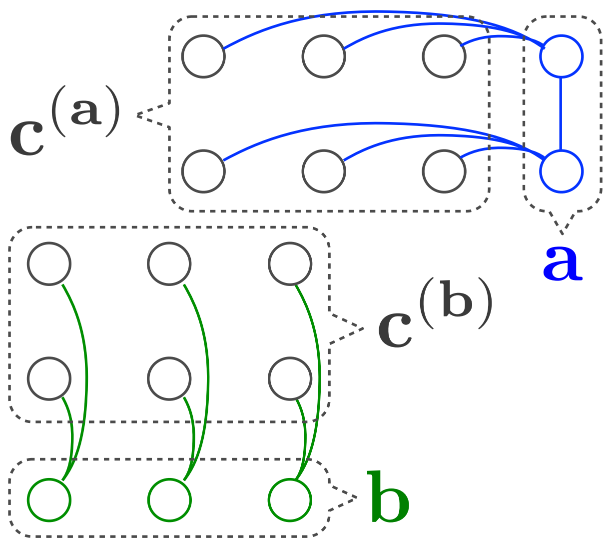
Dual Problem Derivation

Primal problem:



Dual Problem Derivation

Primal problem:

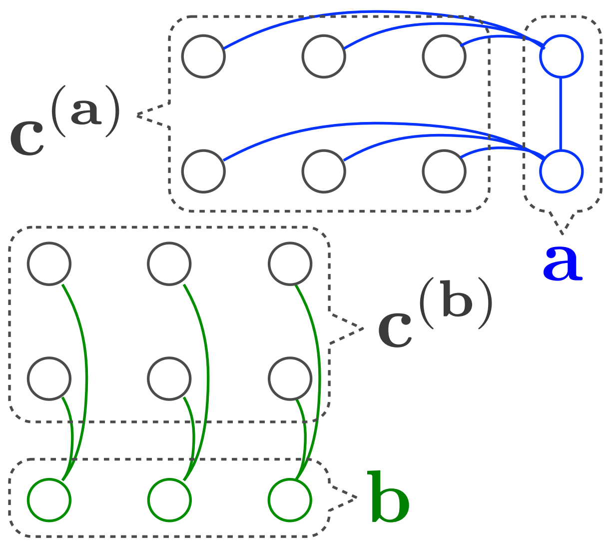


$$\max_{\mathbf{a},\mathbf{b},\mathbf{c}^{(\mathbf{a})},\mathbf{c}^{(\mathbf{b})}} f(\mathbf{a},\mathbf{c}^{(\mathbf{a})}) + g(\mathbf{b},\mathbf{c}^{(\mathbf{b})})$$

$$\textit{such that: } c_{ij}^{(\mathbf{a})} = c_{ij}^{(\mathbf{b})} \quad \forall\ (i,j) \in \mathcal{I}$$

Dual Problem Derivation

Primal problem:



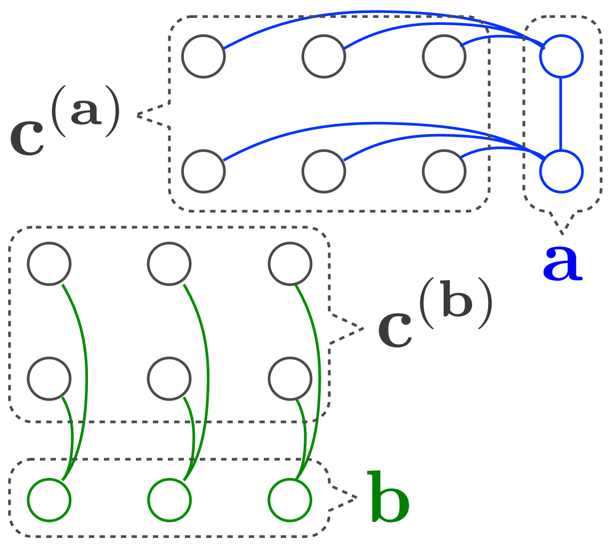
$\log \mathrm{P}(\mathbf{f}, \mathbf{a} | \mathbf{e})$

$$\max_{\mathbf{a},\mathbf{b},\mathbf{c}^{(\mathbf{a})},\mathbf{c}^{(\mathbf{b})}} f(\mathbf{a}, \mathbf{c}^{(\mathbf{a})}) + g(\mathbf{b}, \mathbf{c}^{(\mathbf{b})})$$

$$\textit{such that: } c_{ij}^{(\mathbf{a})} = c_{ij}^{(\mathbf{b})} \quad \forall\ (i,j) \in \mathcal{I}$$

Dual Problem Derivation

Primal problem:



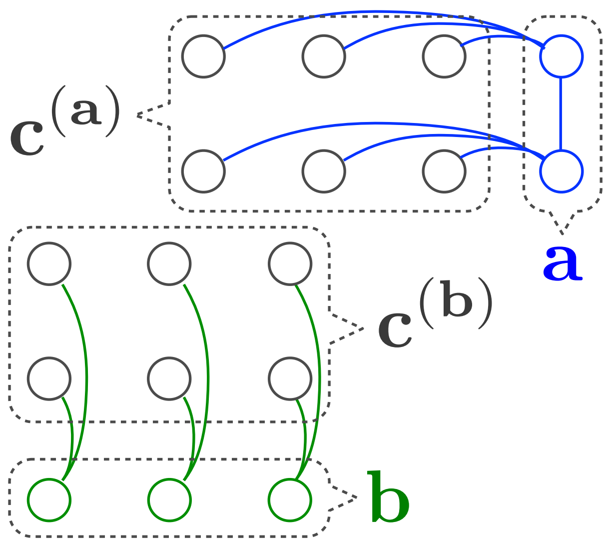
$\log \mathrm{P}(\mathbf{f}, \mathbf{a} | \mathbf{e})$ $\quad$ $\log \mathrm{P}(\mathbf{e}, \mathbf{b} | \mathbf{f})$

$$\max_{\mathbf{a}, \mathbf{b}, \mathbf{c}^{(\mathbf{a})}, \mathbf{c}^{(\mathbf{b})}} f(\mathbf{a}, \mathbf{c}^{(\mathbf{a})}) + g(\mathbf{b}, \mathbf{c}^{(\mathbf{b})})$$

$$\textit{such that: } c_{ij}^{(\mathbf{a})} = c_{ij}^{(\mathbf{b})} \quad \forall\ (i, j) \in \mathcal{I}$$

Dual Problem Derivation

Primal problem:



$\log \mathrm{P}(\mathbf{f}, \mathbf{a} | \mathbf{e})$ $\quad$ $\log \mathrm{P}(\mathbf{e}, \mathbf{b} | \mathbf{f})$

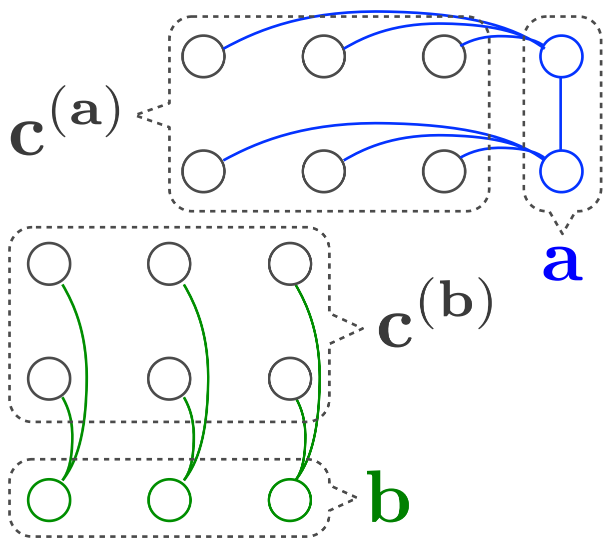
$$\max_{\mathbf{a},\mathbf{b},\mathbf{c}^{(\mathbf{a})},\mathbf{c}^{(\mathbf{b})}} f(\mathbf{a}, \mathbf{c}^{(\mathbf{a})}) + g(\mathbf{b}, \mathbf{c}^{(\mathbf{b})})$$

$$\textit{such that: } c_{ij}^{(\mathbf{a})} = c_{ij}^{(\mathbf{b})} \quad \forall\ (i, j) \in \mathcal{I}$$

Lagrange relaxation:

Dual Problem Derivation

Primal problem:



$\log \mathrm{P}(\mathbf{f}, \mathbf{a}|\mathbf{e})$ $\log \mathrm{P}(\mathbf{e}, \mathbf{b}|\mathbf{f})$

$$\max_{\mathbf{a},\mathbf{b},\mathbf{c}^{(\mathbf{a})},\mathbf{c}^{(\mathbf{b})}} f(\mathbf{a}, \mathbf{c}^{(\mathbf{a})}) + g(\mathbf{b}, \mathbf{c}^{(\mathbf{b})})$$

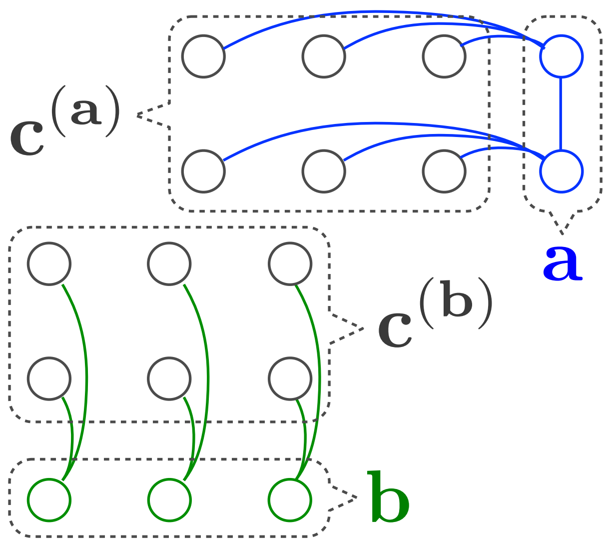
$$\textit{such that: } c_{ij}^{(\mathbf{a})} = c_{ij}^{(\mathbf{b})} \quad \forall \ (i, j) \in \mathcal{I}$$

Lagrange relaxation:

$$f(\mathbf{a}, \mathbf{c}^{(\mathbf{a})}) + g(\mathbf{b}, \mathbf{c}^{(\mathbf{b})}) + \sum_{(i,j)\in\mathcal{I}} u(i, j)(\mathbf{c}_{i,j}^{(\mathbf{a})} - \mathbf{c}_{i,j}^{(\mathbf{b})})$$

Dual Problem Derivation

Primal problem:



$\log \mathrm{P}(\mathbf{f}, \mathbf{a}|\mathbf{e})$

$\log \mathrm{P}(\mathbf{e}, \mathbf{b}|\mathbf{f})$

$$\max_{\mathbf{a},\mathbf{b},\mathbf{c}^{(\mathbf{a})},\mathbf{c}^{(\mathbf{b})}} f(\mathbf{a}, \mathbf{c}^{(\mathbf{a})}) + g(\mathbf{b}, \mathbf{c}^{(\mathbf{b})})$$

$$\textit{such that: } c_{ij}^{(\mathbf{a})} = c_{ij}^{(\mathbf{b})} \quad \forall\ (i,j) \in \mathcal{I}$$

Lagrange relaxation:

Disagreement penalty

$$f(\mathbf{a}, \mathbf{c}^{(\mathbf{a})}) + g(\mathbf{b}, \mathbf{c}^{(\mathbf{b})}) + \sum_{(i,j)\in\mathcal{I}} u(i,j)(\mathbf{c}_{i,j}^{(\mathbf{a})} - \mathbf{c}_{i,j}^{(\mathbf{b})})$$

Dual Problem Derivation

$$f(\mathbf{a}, \mathbf{c}^{(\mathbf{a})}) + g(\mathbf{b}, \mathbf{c}^{(\mathbf{b})}) + \sum_{(i,j) \in \mathcal{I}} u(i,j)(\mathbf{c}^{(\mathbf{a})}_{i,j} - \mathbf{c}^{(\mathbf{b})}_{i,j})$$

Dual Problem Derivation

Primal problem:

$$f(\mathbf{a}, \mathbf{c}^{(\mathbf{a})}) + g(\mathbf{b}, \mathbf{c}^{(\mathbf{b})}) + \sum_{(i,j)\in\mathcal{I}} u(i,j)(\mathbf{c}_{i,j}^{(\mathbf{a})} - \mathbf{c}_{i,j}^{(\mathbf{b})})$$

Dual Problem Derivation

Primal problem: $\max_{\mathbf{a},\mathbf{b},\mathbf{c}^{(\mathbf{a})},\mathbf{c}^{(\mathbf{b})}} \min_{\mathbf{u}}$

$$f(\mathbf{a}, \mathbf{c}^{(\mathbf{a})}) + g(\mathbf{b}, \mathbf{c}^{(\mathbf{b})}) + \sum_{(i,j)\in\mathcal{I}} u(i,j)(\mathbf{c}_{i,j}^{(\mathbf{a})} - \mathbf{c}_{i,j}^{(\mathbf{b})})$$

Dual Problem Derivation

Primal problem: $$\max_{\mathbf{a},\mathbf{b},\mathbf{c}^{(\mathbf{a})},\mathbf{c}^{(\mathbf{b})}} \min_{\mathbf{u}}$$

$$f(\mathbf{a},\mathbf{c}^{(\mathbf{a})}) + g(\mathbf{b},\mathbf{c}^{(\mathbf{b})}) + \sum_{(i,j)\in\mathcal{I}} u(i,j)(\mathbf{c}^{(\mathbf{a})}_{i,j} - \mathbf{c}^{(\mathbf{b})}_{i,j})$$

Dual Problem:

Dual Problem Derivation

Primal problem: $\max_{\mathbf{a},\mathbf{b},\mathbf{c}^{(\mathbf{a})},\mathbf{c}^{(\mathbf{b})}} \min_{\mathbf{u}}$

$$f(\mathbf{a},\mathbf{c}^{(\mathbf{a})}) + g(\mathbf{b},\mathbf{c}^{(\mathbf{b})}) + \sum_{(i,j)\in\mathcal{I}} u(i,j)(\mathbf{c}^{(\mathbf{a})}_{i,j} - \mathbf{c}^{(\mathbf{b})}_{i,j})$$

Dual Problem:

$$\min_{\mathbf{u}} \left(\max_{\mathbf{a},\mathbf{c}^{(\mathbf{a})}} \left[f(\mathbf{a},\mathbf{c}^{(\mathbf{a})}) + \sum_{i,j} u(i,j) c^{(\mathbf{a})}_{ij} \right] + \max_{\mathbf{b},\mathbf{c}^{(\mathbf{b})}} \left[g(\mathbf{b},\mathbf{c}^{(\mathbf{b})}) - \sum_{i,j} u(i,j) c^{(\mathbf{b})}_{ij} \right] \right)$$

Dual Problem Derivation

Google

Primal problem: $$\max_{\mathbf{a},\mathbf{b},\mathbf{c}^{(\mathbf{a})},\mathbf{c}^{(\mathbf{b})}} \min_{\mathbf{u}}$$

$$f(\mathbf{a},\mathbf{c}^{(\mathbf{a})}) + g(\mathbf{b},\mathbf{c}^{(\mathbf{b})}) + \sum_{(i,j)\in\mathcal{I}} u(i,j)(\mathbf{c}^{(\mathbf{a})}_{i,j} - \mathbf{c}^{(\mathbf{b})}_{i,j})$$

Dual Problem:

$$\min_{\mathbf{u}} \left(\max_{\mathbf{a},\mathbf{c}^{(\mathbf{a})}} \left[f(\mathbf{a},\mathbf{c}^{(\mathbf{a})}) + \sum_{i,j} u(i,j) c^{(\mathbf{a})}_{ij} \right] + \max_{\mathbf{b},\mathbf{c}^{(\mathbf{b})}} \left[g(\mathbf{b},\mathbf{c}^{(\mathbf{b})}) - \sum_{i,j} u(i,j) c^{(\mathbf{b})}_{ij} \right] \right)$$

Dual Problem Optimization

Dual Problem:

$$\min_{\mathbf{u}} \left(\max_{\mathbf{a},\mathbf{c}^{(\mathbf{a})}} \left[f(\mathbf{a}, \mathbf{c}^{(\mathbf{a})}) + \sum_{i,j} u(i,j) c_{ij}^{(\mathbf{a})} \right] + \max_{\mathbf{b},\mathbf{c}^{(\mathbf{b})}} \left[g(\mathbf{b}, \mathbf{c}^{(\mathbf{b})}) - \sum_{i,j} u(i,j) c_{ij}^{(\mathbf{b})} \right] \right)$$

Dual Problem Optimization

Dual Objective:

$$h(\mathbf{u}) = \left(\max_{\mathbf{a},\mathbf{c}^{(\mathbf{a})}} \left[f(\mathbf{a}, \mathbf{c}^{(\mathbf{a})}) + \sum_{i,j} u(i,j) c_{ij}^{(\mathbf{a})} \right] + \max_{\mathbf{b},\mathbf{c}^{(\mathbf{b})}} \left[g(\mathbf{b}, \mathbf{c}^{(\mathbf{b})}) - \sum_{i,j} u(i,j) c_{ij}^{(\mathbf{b})} \right] \right)$$

Dual Problem Optimization

Dual Objective:

$$h(\mathbf{u}) = \left(\max_{\mathbf{a},\mathbf{c}^{(\mathbf{a})}} \left[f(\mathbf{a}, \mathbf{c}^{(\mathbf{a})}) + \sum_{i,j} u(i,j) c_{ij}^{(\mathbf{a})} \right] + \max_{\mathbf{b},\mathbf{c}^{(\mathbf{b})}} \left[g(\mathbf{b}, \mathbf{c}^{(\mathbf{b})}) - \sum_{i,j} u(i,j) c_{ij}^{(\mathbf{b})} \right] \right)$$

Gradient:

Dual Problem Optimization

Dual Objective:

$$h(\mathbf{u}) = \left(\max_{\mathbf{a},\mathbf{c}^{(\mathbf{a})}} \left[f(\mathbf{a}, \mathbf{c}^{(\mathbf{a})}) + \sum_{i,j} u(i,j) c_{ij}^{(\mathbf{a})} \right] + \max_{\mathbf{b},\mathbf{c}^{(\mathbf{b})}} \left[g(\mathbf{b}, \mathbf{c}^{(\mathbf{b})}) - \sum_{i,j} u(i,j) c_{ij}^{(\mathbf{b})} \right] \right)$$

Gradient:

$$\frac{\partial h(\mathbf{u})}{\partial u(i,j)} = \widehat{\mathbf{c}_{ij}^{(\mathbf{a})}} - \widehat{\mathbf{c}_{ij}^{(\mathbf{b})}}$$

Dual Problem Optimization

Dual Objective:

$$h(\mathbf{u}) = \left(\max_{\mathbf{a},\mathbf{c}^{(\mathbf{a})}} \left[f(\mathbf{a}, \mathbf{c}^{(\mathbf{a})}) + \sum_{i,j} u(i,j) c_{ij}^{(\mathbf{a})} \right] + \max_{\mathbf{b},\mathbf{c}^{(\mathbf{b})}} \left[g(\mathbf{b}, \mathbf{c}^{(\mathbf{b})}) - \sum_{i,j} u(i,j) c_{ij}^{(\mathbf{b})} \right] \right)$$

Gradient:

$$\frac{\partial h(\mathbf{u})}{\partial u(i,j)} = \widehat{\mathbf{c}_{ij}^{(\mathbf{a})}} - \widehat{\mathbf{c}_{ij}^{(\mathbf{b})}}$$

Results of optimizing each term independently

Implementing Dual Decomposition Inference

- **for** $t = 1$ to max iterations **do**

$$r \leftarrow \frac{1}{t}$$

$$\widehat{\mathbf{c}^{(\mathbf{a})}} \leftarrow \arg\max f(\mathbf{a}, \mathbf{c}^{(\mathbf{a})}) + \textstyle\sum_{i,j} u(i,j) c_{ij}^{(\mathbf{a})}$$

$$\widehat{\mathbf{c}^{(\mathbf{b})}} \leftarrow \arg\max g(\mathbf{b}, \mathbf{c}^{(\mathbf{b})}) - \textstyle\sum_{i,j} u(i,j) c_{ij}^{(\mathbf{b})}$$

if $\widehat{\mathbf{c}^{(\mathbf{a})}} = \widehat{\mathbf{c}^{(\mathbf{b})}}$ **then**

return $\widehat{\mathbf{c}^{(\mathbf{a})}}$

$$\mathbf{u} \leftarrow \mathbf{u} + r \cdot \left(\widehat{\mathbf{c}^{(\mathbf{b})}} - \widehat{\mathbf{c}^{(\mathbf{a})}} \right)$$

return $\text{symm}\left(\widehat{\mathbf{c}^{(\mathbf{a})}}, \widehat{\mathbf{c}^{(\mathbf{b})}} \right)$

Implementing Dual Decomposition Inference

for $t = 1$ to max iterations **do**

$$r \leftarrow \frac{1}{t}$$

$$\widehat{\mathbf{c}^{(\mathbf{a})}} \leftarrow \arg\max f(\mathbf{a}, \mathbf{c}^{(\mathbf{a})}) + \sum_{i,j} u(i,j) c_{ij}^{(\mathbf{a})}$$

$$\widehat{\mathbf{c}^{(\mathbf{b})}} \leftarrow \arg\max g(\mathbf{b}, \mathbf{c}^{(\mathbf{b})}) - \sum_{i,j} u(i,j) c_{ij}^{(\mathbf{b})}$$

if $\widehat{\mathbf{c}^{(\mathbf{a})}} = \widehat{\mathbf{c}^{(\mathbf{b})}}$ **then**

return $\widehat{\mathbf{c}^{(\mathbf{a})}}$

$$\mathbf{u} \leftarrow \mathbf{u} + r \cdot \left(\widehat{\mathbf{c}^{(\mathbf{b})}} - \widehat{\mathbf{c}^{(\mathbf{a})}} \right)$$

return $\text{symm}\left(\widehat{\mathbf{c}^{(\mathbf{a})}}, \widehat{\mathbf{c}^{(\mathbf{b})}} \right)$

Implementing Dual Decomposition Inference

for $t = 1$ to max iterations **do**

$$r \leftarrow \frac{1}{t}$$

$$\widehat{\mathbf{c}^{(\mathbf{a})}} \leftarrow \arg\max f(\mathbf{a}, \mathbf{c}^{(\mathbf{a})}) + \sum_{i,j} u(i,j) c_{ij}^{(\mathbf{a})}$$

$$\widehat{\mathbf{c}^{(\mathbf{b})}} \leftarrow \arg\max g(\mathbf{b}, \mathbf{c}^{(\mathbf{b})}) - \sum_{i,j} u(i,j) c_{ij}^{(\mathbf{b})}$$

if $\widehat{\mathbf{c}^{(\mathbf{a})}} = \widehat{\mathbf{c}^{(\mathbf{b})}}$ **then**

return $\widehat{\mathbf{c}^{(\mathbf{a})}}$

$$\mathbf{u} \leftarrow \mathbf{u} + r \cdot \left(\widehat{\mathbf{c}^{(\mathbf{b})}} - \widehat{\mathbf{c}^{(\mathbf{a})}} \right)$$

return $\text{symm}\left(\widehat{\mathbf{c}^{(\mathbf{a})}}, \widehat{\mathbf{c}^{(\mathbf{b})}} \right)$

Implementing Dual Decomposition Inference

for $t = 1$ to max iterations **do**

$\quad r \leftarrow \frac{1}{t}$

$\quad \widehat{\mathbf{c}^{(\mathbf{a})}} \leftarrow \arg\max f(\mathbf{a}, \mathbf{c}^{(\mathbf{a})}) + \sum_{i,j} u(i,j) c_{ij}^{(\mathbf{a})}$

$\quad \widehat{\mathbf{c}^{(\mathbf{b})}} \leftarrow \arg\max g(\mathbf{b}, \mathbf{c}^{(\mathbf{b})}) - \sum_{i,j} u(i,j) c_{ij}^{(\mathbf{b})}$

$\quad$ **if** $\widehat{\mathbf{c}^{(\mathbf{a})}} = \widehat{\mathbf{c}^{(\mathbf{b})}}$ **then**

$\quad\quad$ **return** $\widehat{\mathbf{c}^{(\mathbf{a})}}$

$\quad \mathbf{u} \leftarrow \mathbf{u} + r \cdot \left(\widehat{\mathbf{c}^{(\mathbf{b})}} - \widehat{\mathbf{c}^{(\mathbf{a})}} \right)$

return $\text{symm}\left(\widehat{\mathbf{c}^{(\mathbf{a})}}, \widehat{\mathbf{c}^{(\mathbf{b})}} \right)$

Implementing Dual Decomposition Inference

$$\textbf{for } t = 1 \text{ to max iterations } \textbf{do}$$

$$\quad r \leftarrow \frac{1}{t}$$

$$\quad \widehat{\mathbf{c}^{(\mathbf{a})}} \leftarrow \arg\max f(\mathbf{a}, \mathbf{c}^{(\mathbf{a})}) + \textstyle\sum_{i,j} u(i,j) c_{ij}^{(\mathbf{a})}$$

$$\quad \widehat{\mathbf{c}^{(\mathbf{b})}} \leftarrow \arg\max g(\mathbf{b}, \mathbf{c}^{(\mathbf{b})}) - \textstyle\sum_{i,j} u(i,j) c_{ij}^{(\mathbf{b})}$$

$$\quad \textbf{if } \widehat{\mathbf{c}^{(\mathbf{a})}} = \widehat{\mathbf{c}^{(\mathbf{b})}} \textbf{ then}$$

$$\qquad \textbf{return } \widehat{\mathbf{c}^{(\mathbf{a})}}$$

$$\quad \mathbf{u} \leftarrow \mathbf{u} + r \cdot \left(\widehat{\mathbf{c}^{(\mathbf{b})}} - \widehat{\mathbf{c}^{(\mathbf{a})}} \right)$$

$$\textbf{return } \operatorname{symm}\left(\widehat{\mathbf{c}^{(\mathbf{a})}}, \widehat{\mathbf{c}^{(\mathbf{b})}} \right)$$

Implementing Dual Decomposition Inference

for $t = 1$ to max iterations **do**

$\quad r \leftarrow \frac{1}{t}$

$\quad \widehat{\mathbf{c}^{(\mathbf{a})}} \leftarrow \arg\max f(\mathbf{a}, \mathbf{c}^{(\mathbf{a})}) + \sum_{i,j} u(i,j) c_{ij}^{(\mathbf{a})}$

$\quad \widehat{\mathbf{c}^{(\mathbf{b})}} \leftarrow \arg\max g(\mathbf{b}, \mathbf{c}^{(\mathbf{b})}) - \sum_{i,j} u(i,j) c_{ij}^{(\mathbf{b})}$

$\quad$ **if** $\widehat{\mathbf{c}^{(\mathbf{a})}} = \widehat{\mathbf{c}^{(\mathbf{b})}}$ **then**

$\quad\quad$ **return** $\widehat{\mathbf{c}^{(\mathbf{a})}}$

$\quad \mathbf{u} \leftarrow \mathbf{u} + r \cdot \left(\widehat{\mathbf{c}^{(\mathbf{b})}} - \widehat{\mathbf{c}^{(\mathbf{a})}} \right)$

return symm$\left(\widehat{\mathbf{c}^{(\mathbf{a})}}, \widehat{\mathbf{c}^{(\mathbf{b})}} \right)$

Optimality Upon Convergence

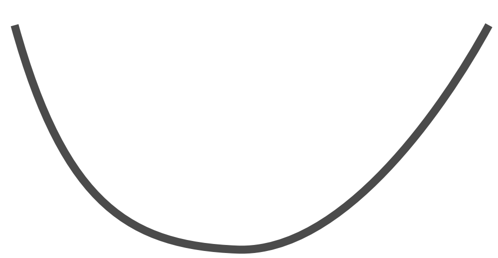
Optimality Upon Convergence

- Dual objective is **convex**
  - Dual optimum reached if **gradient descent** converges

Optimality Upon Convergence

- Dual objective is **convex**
  - Dual optimum reached if **gradient descent** converges

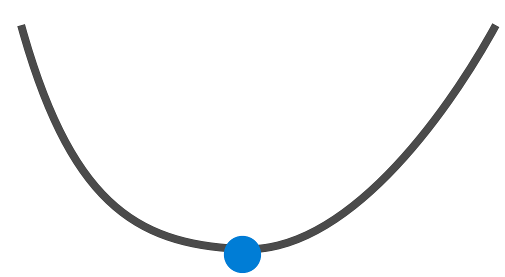
Dramatization:



Optimality Upon Convergence

- Dual objective is **convex**
  - Dual optimum reached if **gradient descent** converges
- Converged dual optimum **satisfies all constraints** of the primal
  - Converged dual optimum is a **feasible** primal solution

Dramatization:

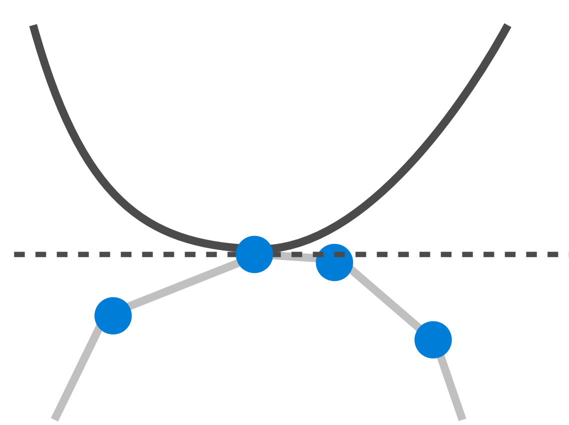


Optimality Upon Convergence

- Dual objective is **convex**
  - Dual optimum reached if **gradient descent** converges
- Converged dual optimum **satisfies all constraints** of the primal
  - Converged dual optimum is a **feasible** primal solution
- The dual optimum is an **upper bound** on the primal optimum

∎

Dramatization:



Experimental Design

Experimental Design

- Trained on 6.2 million words of Chinese-English FBIS data

Experimental Design

- Trained on 6.2 million words of Chinese-English FBIS data
- Evaluated on 150 hand-aligned sentences of NIST 2002 data

Experimental Design

- Trained on 6.2 million words of Chinese-English FBIS data
- Evaluated on 150 hand-aligned sentences of NIST 2002 data
- Training regimen:

Experimental Design

- Trained on 6.2 million words of Chinese-English FBIS data
- Evaluated on 150 hand-aligned sentences of NIST 2002 data
- Training regimen:
  - 5 iterations of Model 1, with agreement (Liang et al, 2006)

Experimental Design

- Trained on 6.2 million words of Chinese-English FBIS data
- Evaluated on 150 hand-aligned sentences of NIST 2002 data
- Training regimen:
  - 5 iterations of Model 1, with agreement (Liang et al, 2006)
  - 5 iterations of HMM, no agreement training (*it's better*)

Experimental Design

- Trained on 6.2 million words of Chinese-English FBIS data
- Evaluated on 150 hand-aligned sentences of NIST 2002 data
- Training regimen:
  - 5 iterations of Model 1, with agreement (Liang et al, 2006)
  - 5 iterations of HMM, no agreement training (*it's better*)
- Aligners implemented in the Go language, but ≈ Berkeley Aligner

Experimental Design

- Trained on 6.2 million words of Chinese-English FBIS data
- Evaluated on 150 hand-aligned sentences of NIST 2002 data
- Training regimen:
  - 5 iterations of Model 1, with agreement (Liang et al, 2006)
  - 5 iterations of HMM, no agreement training (*it's better*)
- Aligners implemented in the Go language, but ≈ Berkeley Aligner
- HMM parameters are fixed for all experiments

Convergence and Agreement Rates

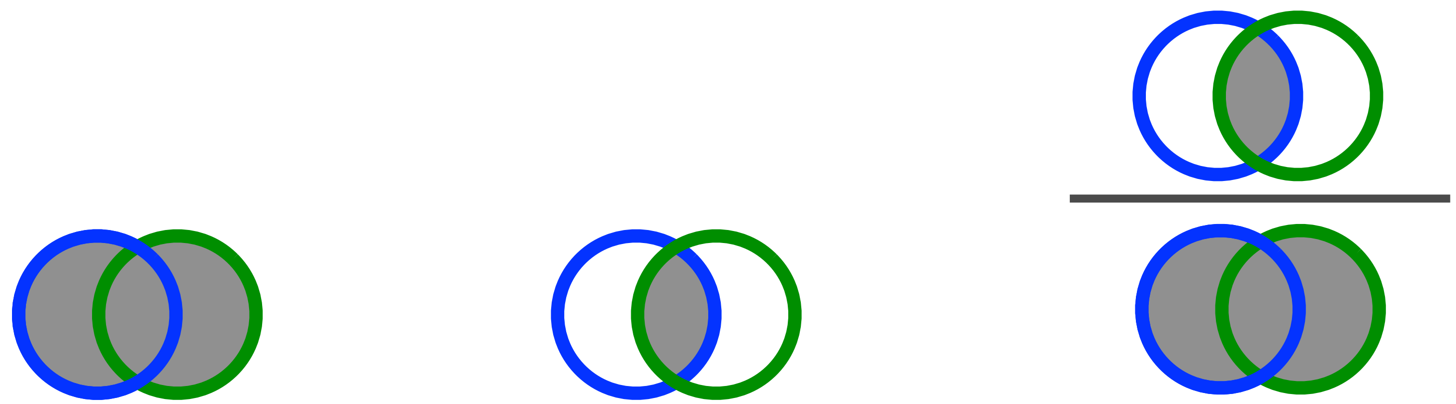
After 250 iterations, inference converges 6.2% of the time

Dual solution oscillates, implying a duality gap

Convergence and Agreement Rates

After 250 iterations, inference converges 6.2% of the time

Dual solution oscillates, implying a duality gap

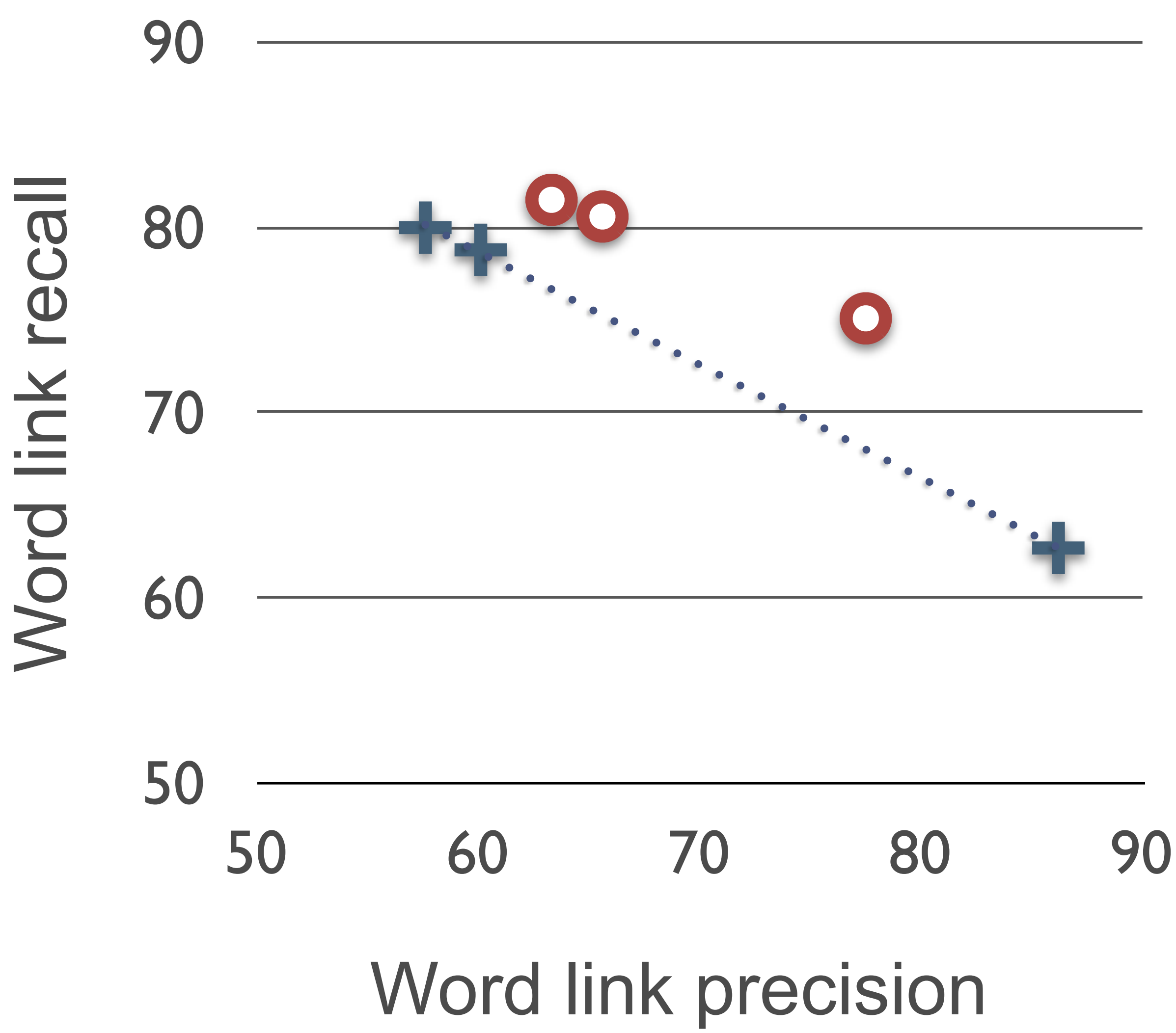


	Union Size	Intersection Size	Ratio
Independent	10,998	5,554	50.5%
Model-Based	10,262	7,620	74.3%

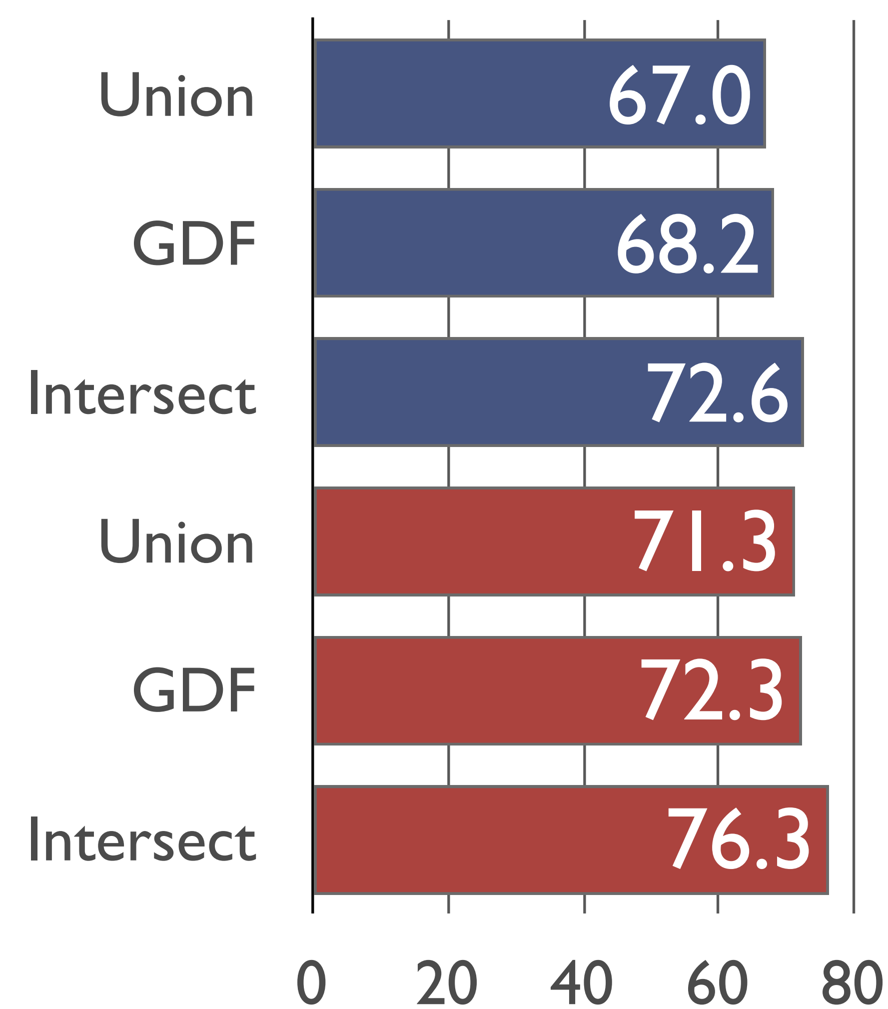
Alignment Error

Combination methods: ✚ Heuristic ○ Model-Based

Word-Level Accuracy



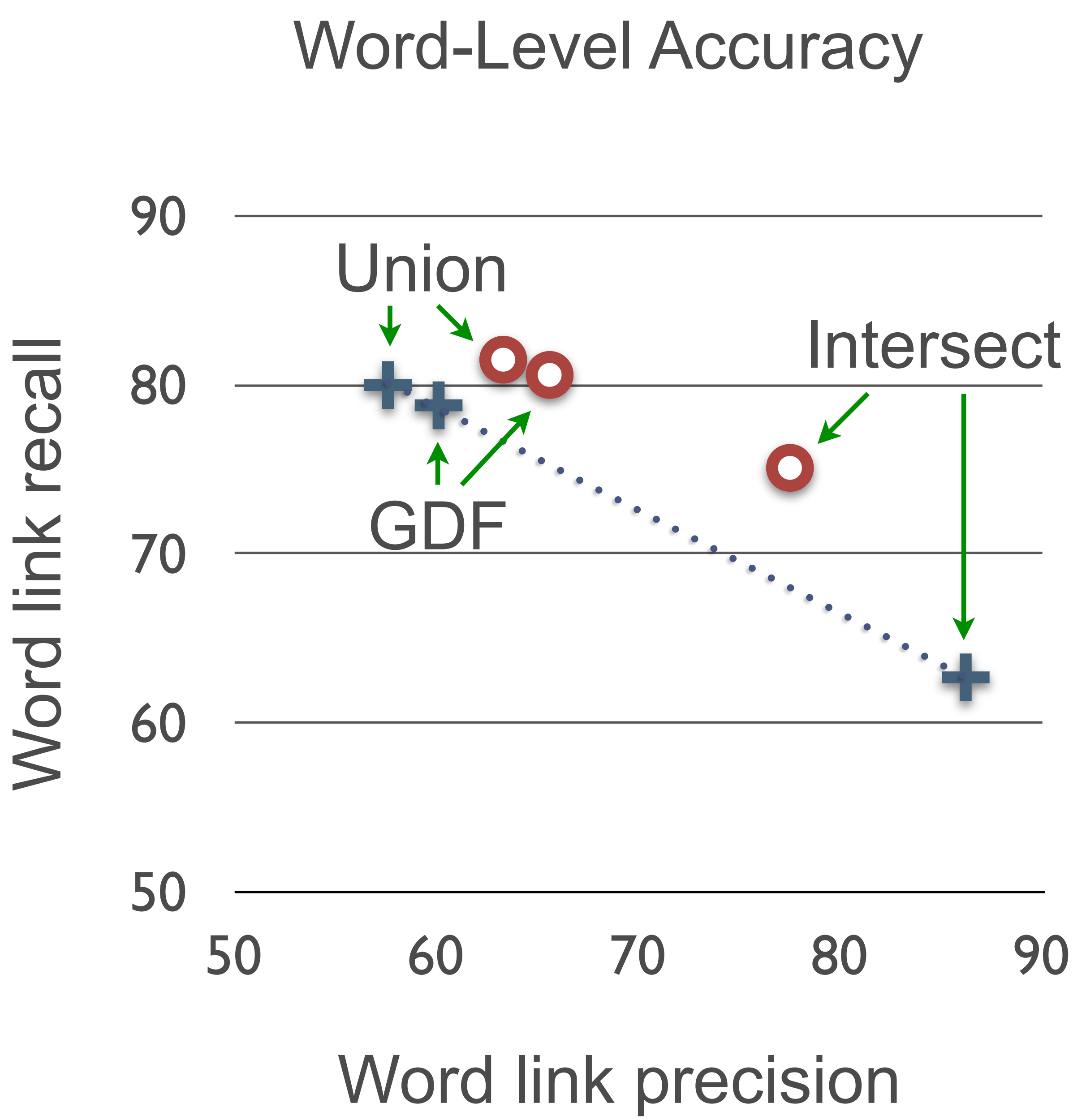
Link F1

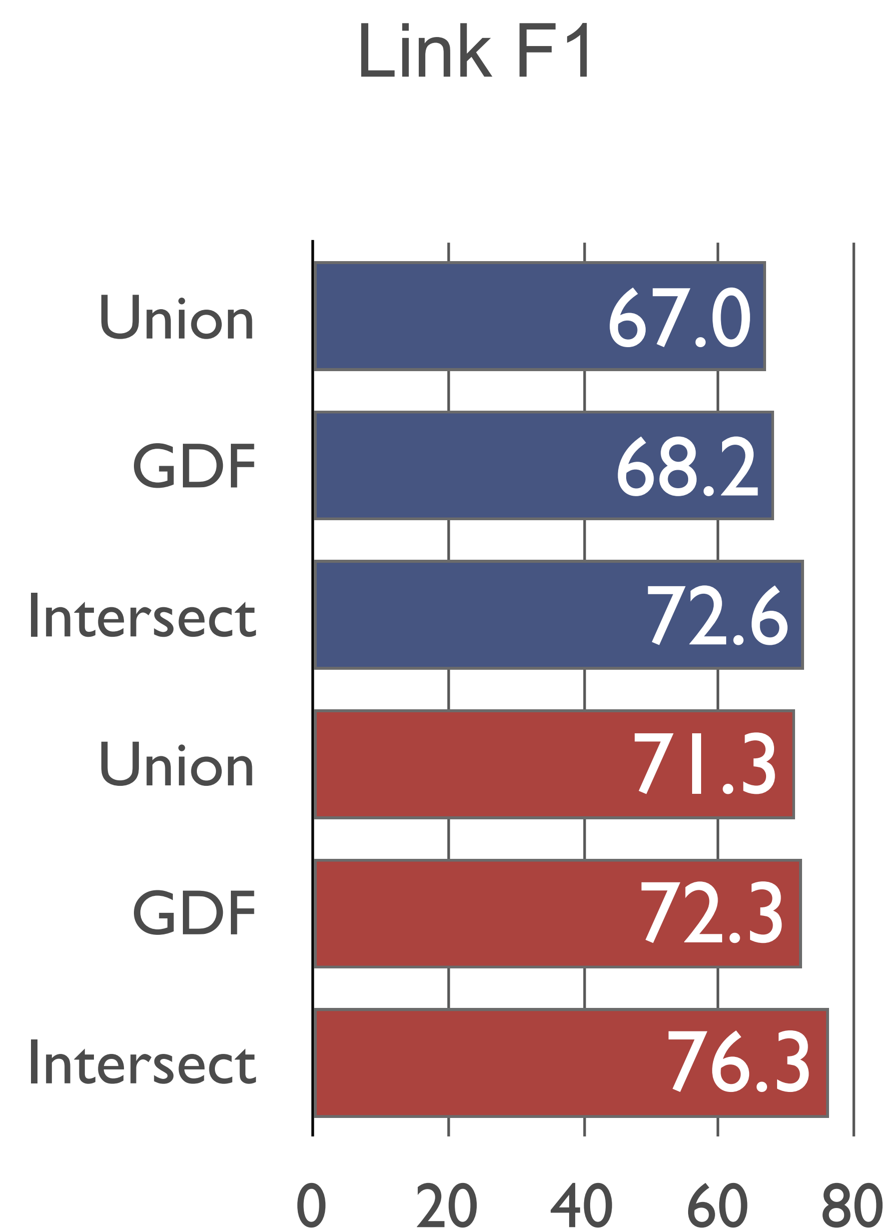


Saturday, June 18, 2011

Alignment Error

Combination methods: ✚ Heuristic ⭘ Model-Based

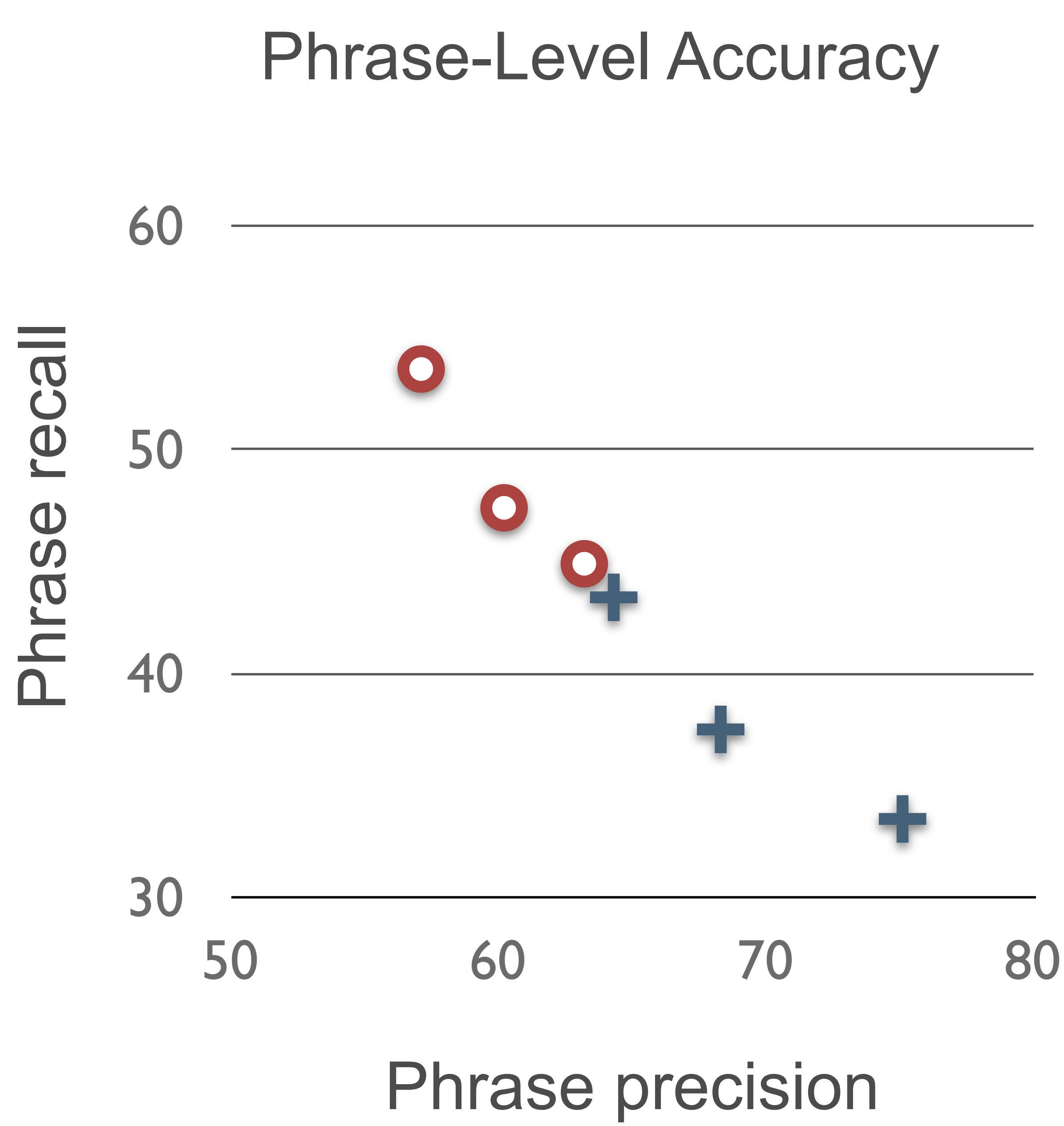


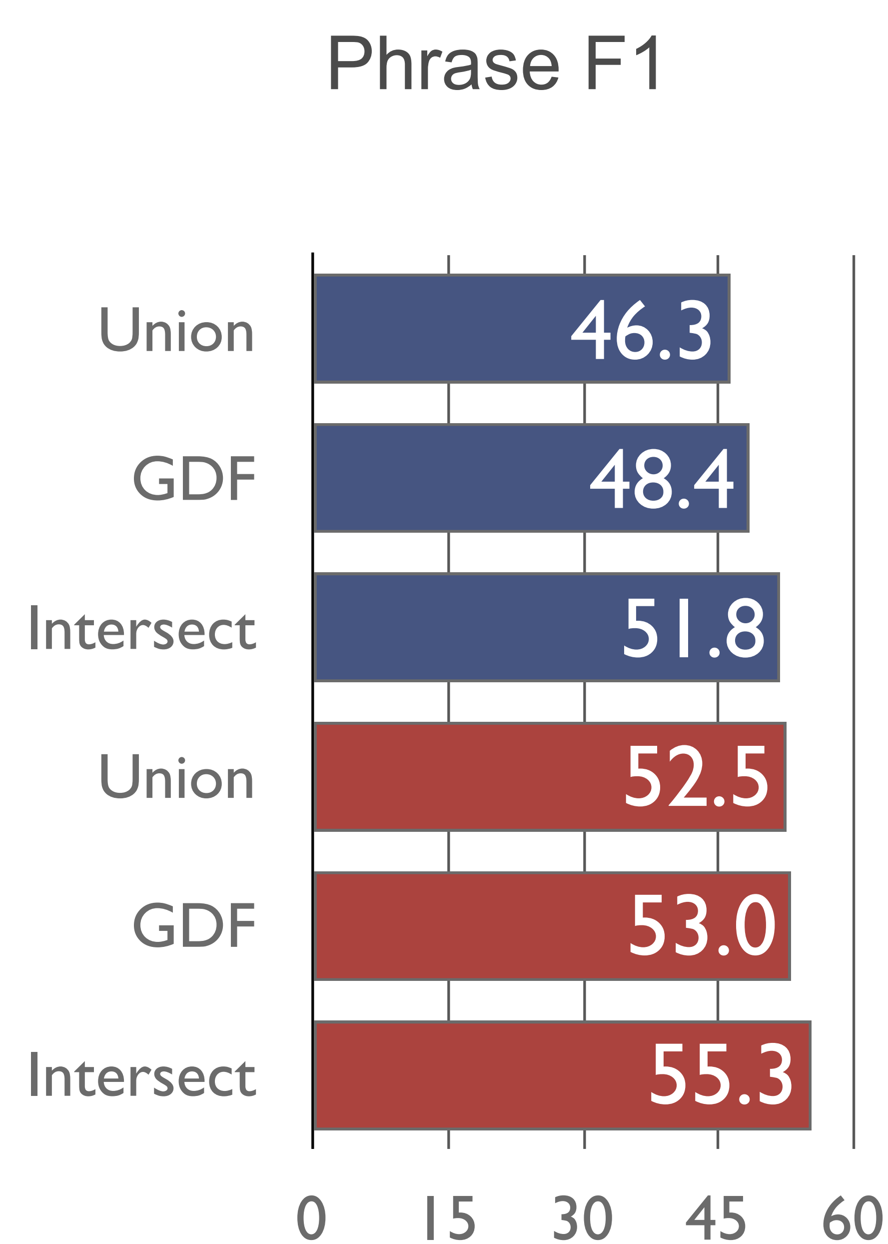


Saturday, June 18, 2011

Phrase Extraction Results

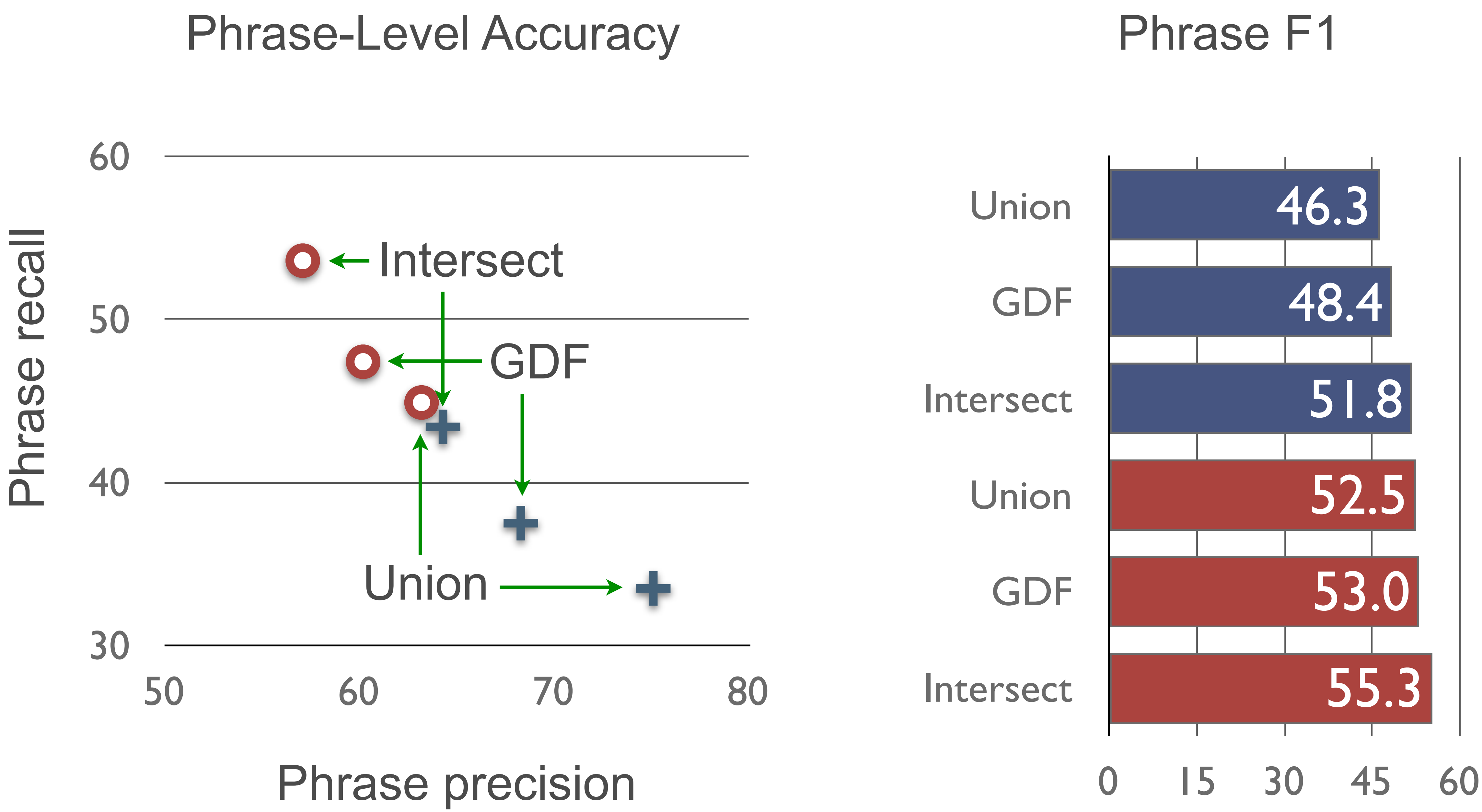
Combination methods: ✚ Heuristic ○ Model-Based





Phrase Extraction Results

Combination methods: ✚ Heuristic ○ Model-Based



End-to-End Translation

- Google research Chinese-to-English alignment template system
- Union outperformed other symmetrization heuristics
- Model 1 & HMM each trained for 3 iterations
- Training and test examples collected from the web
- Single-reference test set commissioned from professional translators

	BLEU
Heuristic	29.59
Model-Based	29.82

Conclusion

Conclusion

- Extensible graphical model framework for aligner combination

Conclusion

- Extensible graphical model framework for aligner combination

- Easy-to-implement dual decomposition inference

Conclusion

- Extensible graphical model framework for aligner combination
- Easy-to-implement dual decomposition inference
- Phrase alignments without pruning, greedy search, ILPs, or ITGs

Conclusion

- Extensible graphical model framework for aligner combination
- Easy-to-implement dual decomposition inference
- Phrase alignments without pruning, greedy search, ILPs, or ITGs
- Convergence is elusive, but early stopping gives useful results

Conclusion

- Extensible graphical model framework for aligner combination
- Easy-to-implement dual decomposition inference
- Phrase alignments without pruning, greedy search, ILPs, or ITGs
- Convergence is elusive, but early stopping gives useful results

T h a n k s
T h a n k s

Q u e s t i o n s ?
Q u e s t i o n s ?