

# **EBMT, SMT, Hybrid, and More: ATR Spoken Language Translation Systems**

**Eiichiro SUMITA, Yasuhiro AKIBA, Takao DOI,  
Andrew FINCH, Kenji IMAMURA, Hideo OKUMA,  
Michael PAUL, Mitsuo SHIMOHATA, and Taro WATANABE**

# Outline

1. **Hybrid MT System**
2. **Phrase-based HMM SMT System**
3. **Paraphrasing for MT**

<b>Unrestricted</b>	<b>J-to-E</b>	
<b>Supplied</b>	<b>J-to-E</b>	<b>C-to-E</b>

→ The technologies are combined in the near future.

# Hybrid MT System (Unrestricted J-to-E Track)

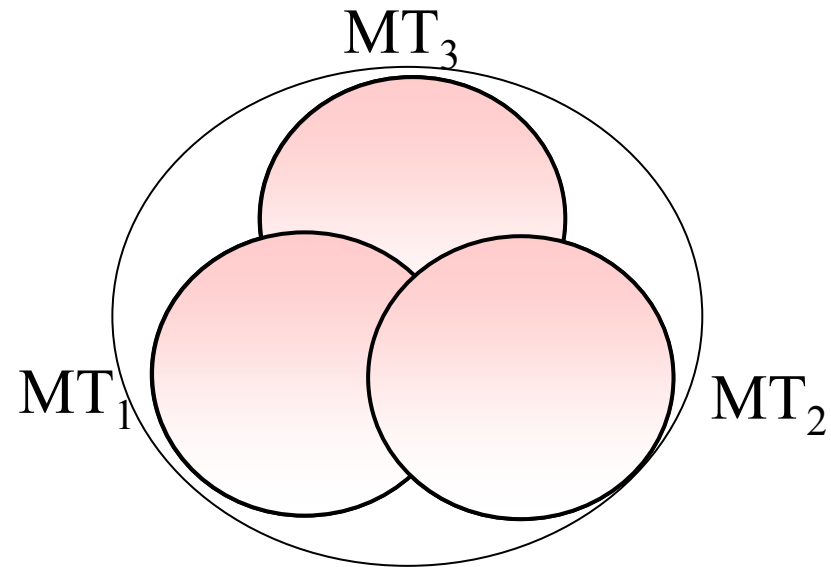
(J) o-shiharai wa genkin desu ka kurejitto kaado desu ka

(MT<sub>1</sub>) Is the payment cash? Or is it by credit card? [Good]

(MT<sub>2</sub>) Would you like to pay by cash or credit card? [Perfect]

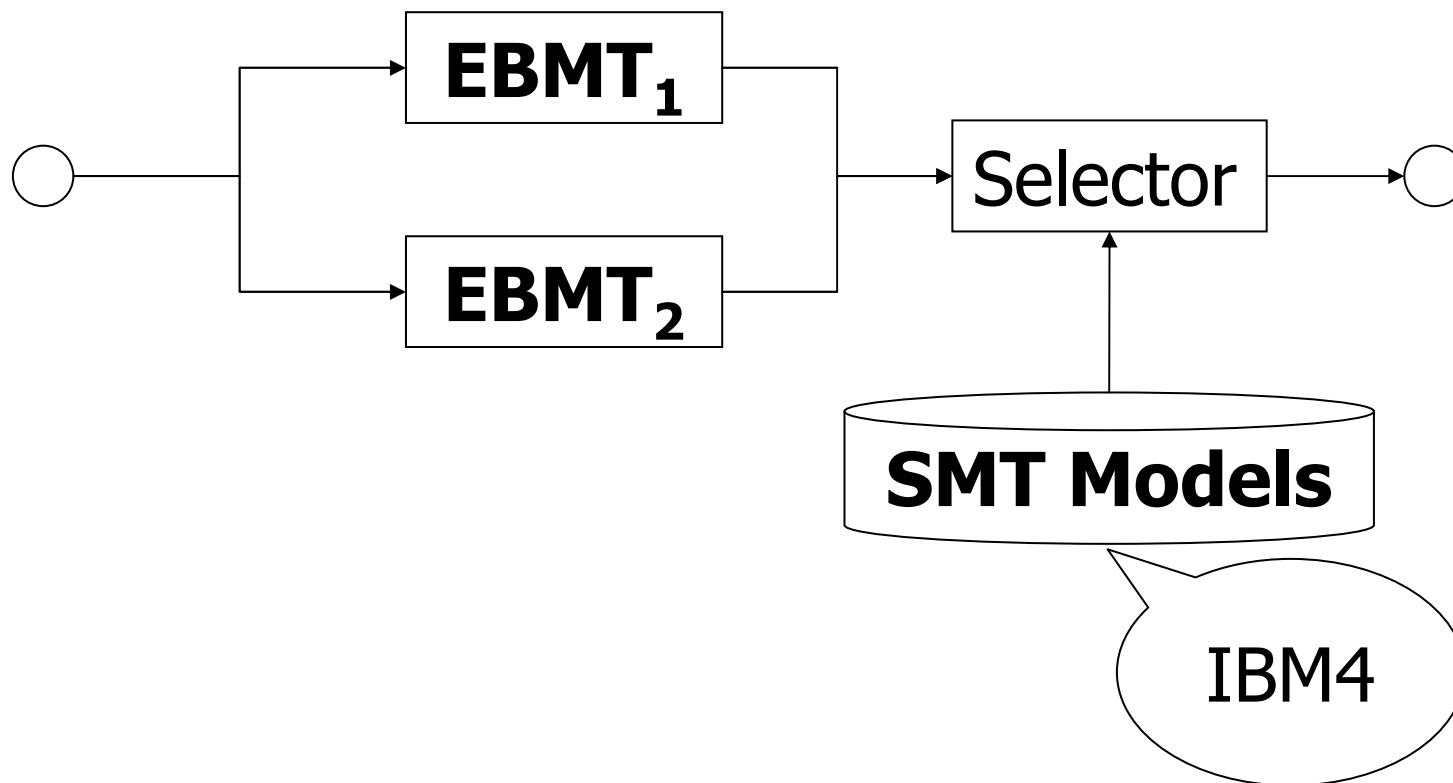
(MT<sub>3</sub>) Could you cash or credit card? [Fair]

# So many men, so many minds.



# Today's Hybridization

“Multiple **EBMTs** Followed By  
A Selector Based On **SMT Models**”



# EBMT<sub>1</sub>

## D<sup>3</sup>(Dp-match Driven transDucer)

**Input** (J) iro/ga/ki/ni/iri/masen

RETRIEVE

Thesaurus

**Example** { (J) dezain/ga/ki/ni/iri/masen  
(E) I do not like the design.

ADAPT

Corpus

**Output** (E) I do not like the color.

Bilingual Dic.

**EBMT<sub>1</sub>**  
**D<sup>3</sup> (2)**

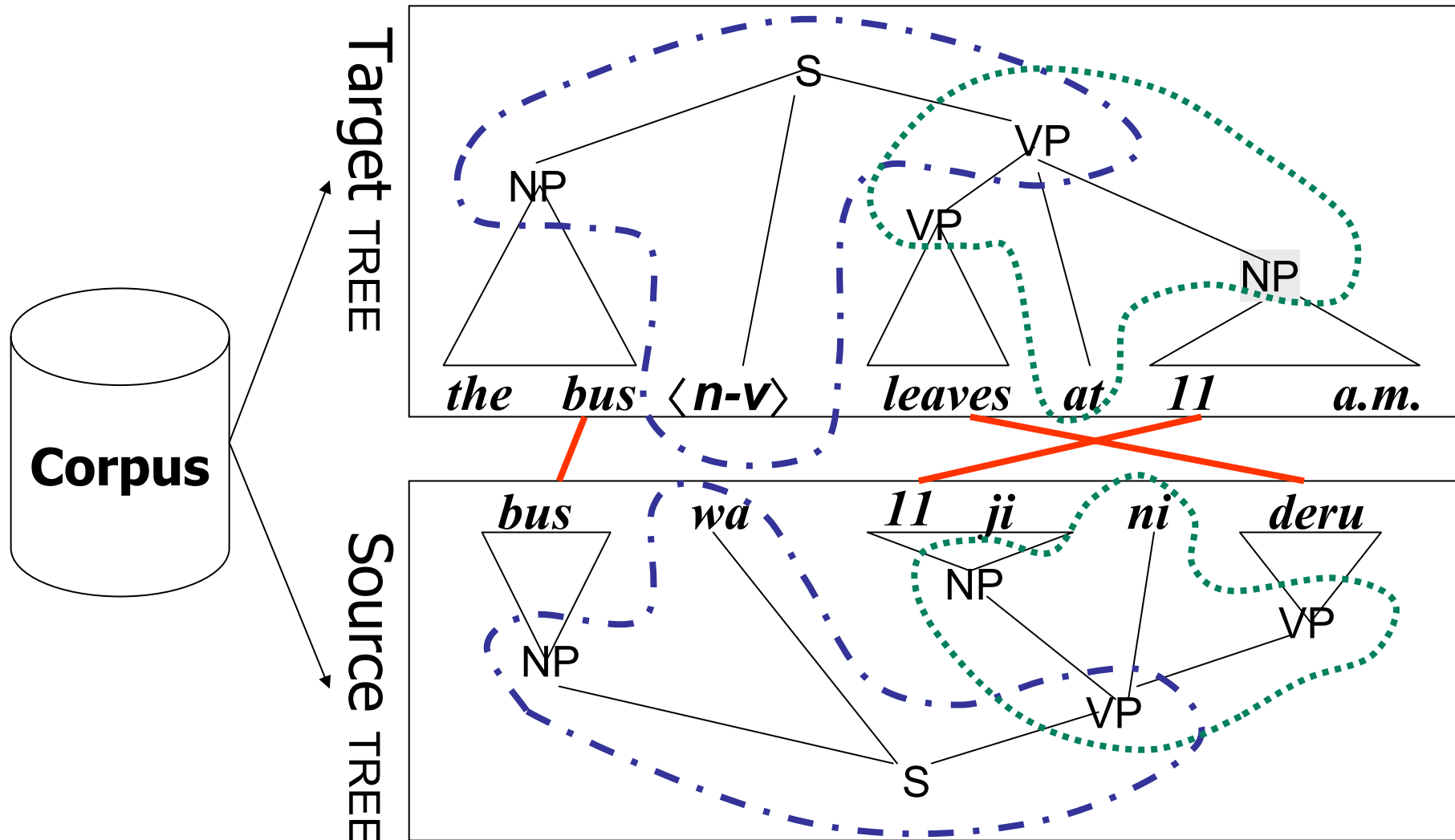
$$\textit{Distance} = \frac{I + D + 2 \sum \textit{SEMDIST}}{L_{input} + L_{example}}$$

Height of Most Specific Common  
Abstraction in the thesaurus

$$\textit{SEMDIST} = \frac{K}{N}$$

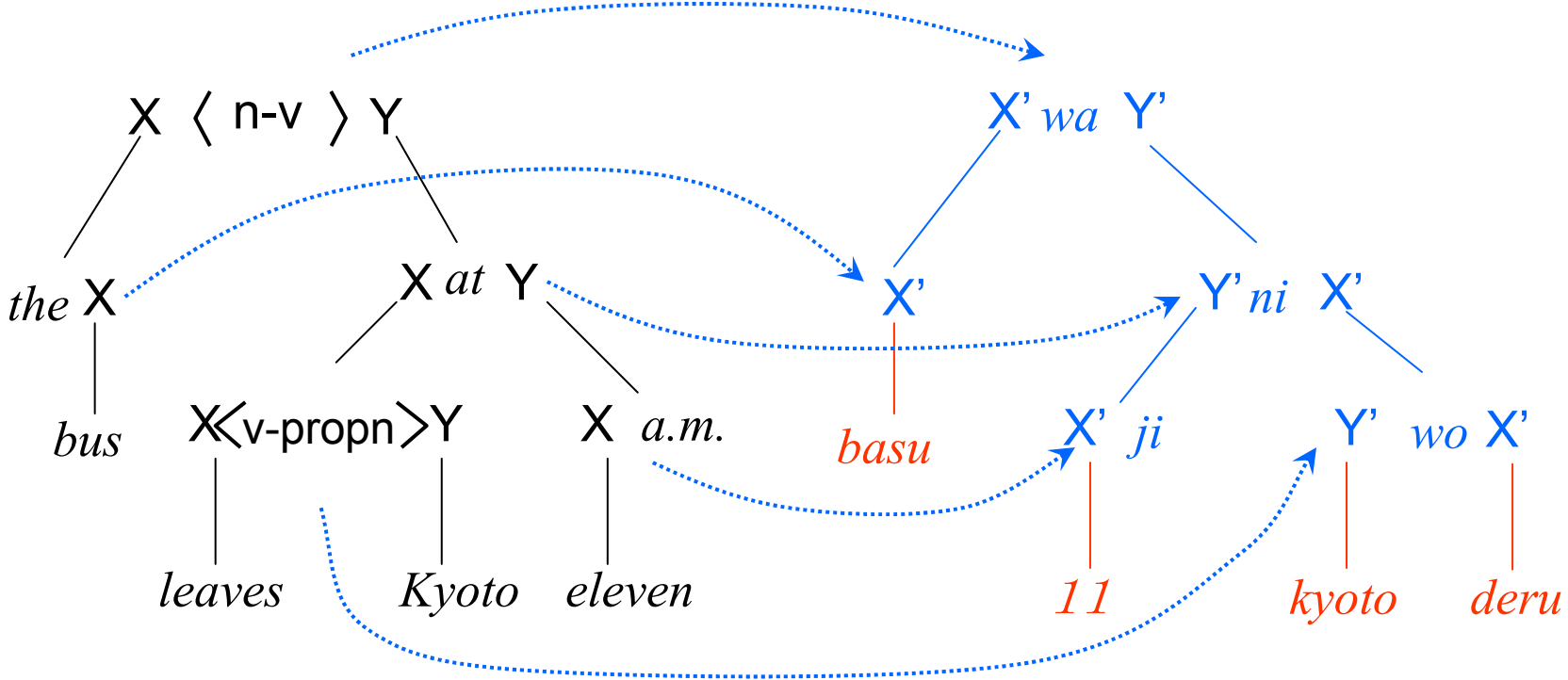
Height of the Thesaurus

# EBMT<sub>2</sub> Hierarchical Phrase Alignment based Translation (HPAT)





# EBMT<sub>2</sub> HPAT(2)



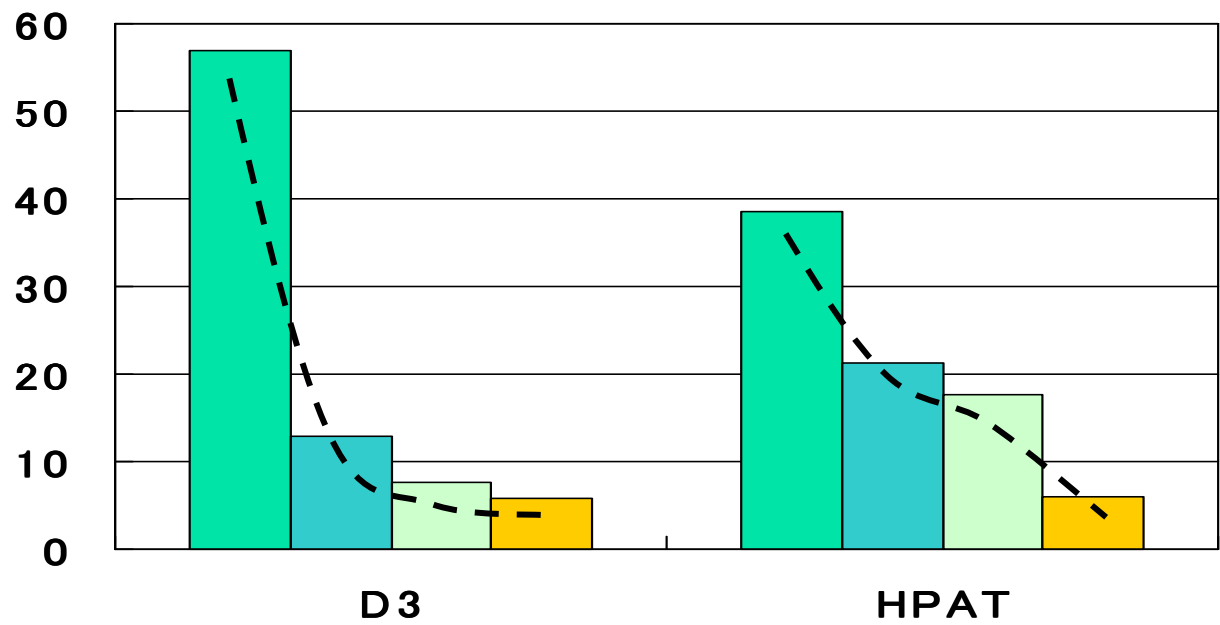
- (1) Parse source language using source patterns.
- (2) Map source patterns to target patterns.
- (3) Translate leaves by referring to a dictionary.

# Comparison of Two EBMTs

Table 2: *Features of the two EBMTs*

	<b>D3</b>	<b>HPAT</b>
<i>Unit</i>	<i>Sentence</i>	<i>Grammatical Unit</i>
Coverage	Narrow	Wide
Quality	Good	Modest

# Comparison of Two EBMTs (2)



## Quality Ranks

- (S) Perfect;
- (A) Very Good;
- (B) Good;
- (C) Fair;
- (D) Bad.

# SMT-based Selector

■ Conventional selector:

**1. Language Model**

■ Our selector:

**1. Language Model**

**2. Translation Model**

**3. Multiple  
comparison test**

■ **Our selector outperforms component MTs and conventional selectors based on LM.**

# Results (Unrestricted J-to-E Track): Selecting Effect

Table 4: *Objective evaluation*

	<b>D3</b>	<b>HPAT</b>	<b>SELECT</b>	<b>DIFF.</b>
BLEU	60.36	49.33	63.36	+2.7
NIST	10.35	9.78	10.72	+0.37
GTM	77.70	76.88	79.67	+1.97
mWER	28.86	37.18	26.31	-2.55
mPER	26.07	31.06	23.33	-2.74

# Results (Unrestricted J-to-E Track): Selecting Effect

*Table 5: ATR's Overall Subjective Evaluation - cumulative percentages of S, A, B, C, and D ranks.*

	D3	HPAT	SELECT	DIFF.
S	57.00	38.60	59.80	+2.80
S,A	70.00	59.80	73.00	+3.00
S,A,B	77.60	77.40	82.40	+4.80
S,A,B,C	83.40	83.40	87.80	+4.40
D	16.60	16.60	12.20	-4.40

Quality Ranks  
 (S) Perfect;  
 (A) Very Good;  
 (B) Good;  
 (C) Fair;  
 (D) Bad.

# Corpus Size

Table 7+: *ATR's Overall Subjective Evaluation vs. Corpus Size*

	Supplied BTEC (20K)	Full BTEC (200 K)	DIFF.
S	34.00	59.80	25.80
S, A	50.60	73.00	22.40
S, A, B	72.20	82.40	10.20
S, A, B, C	81.80	87.80	6.00
D	18.20	12.20	-6.00

Quality Ranks  
 (S) Perfect;  
 (A) Very Good;  
 (B) Good;  
 (C) Fair;  
 (D) Bad.

# Phrase-based HMM SMT System (Supplied J-to-E and C-to-E Tracks)

$$\hat{e} = \arg \max_e P(e | f) \quad (1)$$

$$\hat{e} = \arg \max_e P(f | e)P(e) \quad (2)$$

$$P(f | e) = \sum_{\bar{f}, \bar{e}} P(f, \bar{f}, \bar{e} | e) \quad (4)$$

$$P(f, \bar{f}, \bar{e} | e) = P(f | \bar{f}, \bar{e}, e)P(\bar{f} | \bar{e}, e)P(\bar{e} | e) \quad (5)$$

Phrase Segmentation

Phrase Translation

Phrase Ngram



# Phrase Ngram Model

$$P(\bar{e} | e) \approx \prod_i P(\bar{e}_i | \bar{e}_{i-1}) \quad (6)$$

- **Forward-backward algorithm** to estimate probabilities

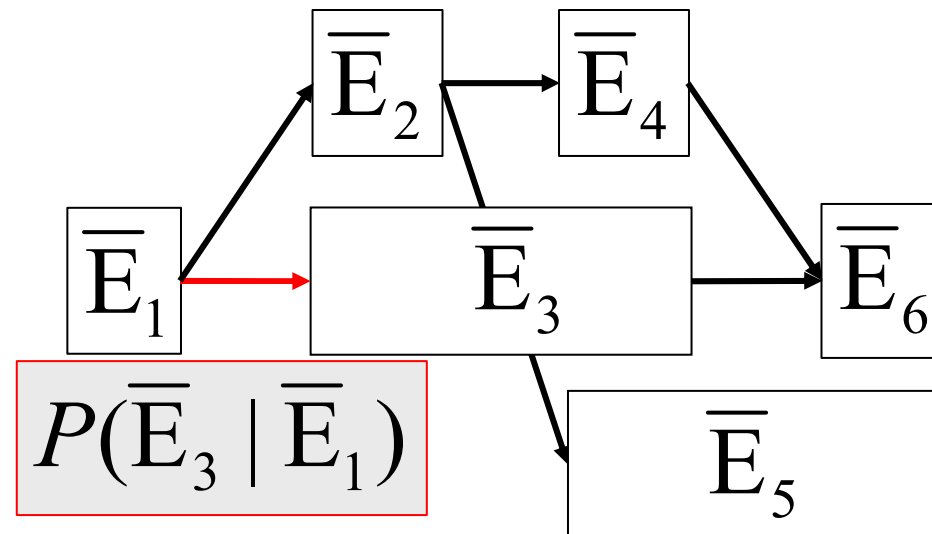


Figure 1: *Phrase Ngram Model*

# Phrase Segmentation Model

- Likelihood of a particular phrase segment  $\bar{f}_j$  observed in  $\mathbf{f}$

$$P(\mathbf{f} | \bar{\mathbf{f}}, \bar{\mathbf{e}}, \mathbf{e}) \propto P(\bar{\mathbf{f}} | \mathbf{f}) \approx \prod_j P(\bar{f}_j | \mathbf{f})$$

- **Forward-backward algorithm** to estimate probabilities

## Phrase Translation Model

$$P(\bar{f} | \bar{e}, e) \approx \prod_j P(\bar{f}_j | \bar{e}_{a_j}) \quad (11)$$

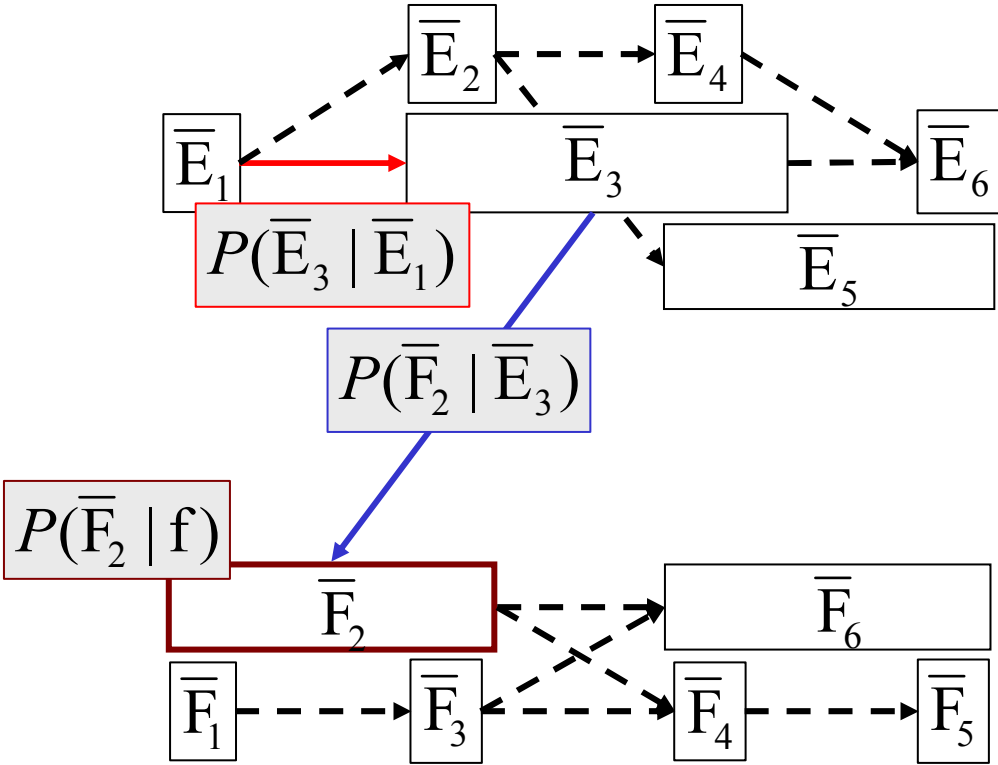
# Combined

$$P(\mathbf{f} | \mathbf{e}) \approx \sum_{\bar{\mathbf{e}}, \bar{\mathbf{f}}} \prod_{j,i} P(\bar{f}_j | \mathbf{f}) P(\bar{f}_j | \bar{e}_i) P(\bar{e}_i | \bar{e}_{i-1}) \quad (12)$$

Phrase Segmentation

Phrase Translation

Phrase Ngram



# Parameter Estimation

- Please consult Section 3.5 in the paper.

# Phrase Segment Induction

Extract phrase pair using the following criterion:

$$P(\bar{e}|\bar{f})P(\bar{f}|\bar{e}) = \frac{\text{count}(\bar{e}, \bar{f})^2}{\sum_{\bar{f}} \text{count}(\bar{e}, \bar{f}) \sum_{\bar{g}} \text{count}(\bar{e}, \bar{f})}$$

# Decoding

$$\hat{e} = \operatorname{argmax}_{\mathbf{e}} \frac{1}{Z(\mathbf{f})} \sum_j \lambda_j \log Pr_j(\mathbf{e}, \mathbf{f})$$

- ◆ Discriminative training to determine  $\lambda_j$   
(Och and Ney 2002; Och 2003)
- ◆ Word graph based search  
(Ueffing et al. 2002)

# Results: Supplied task JE and CE

Table 8+: *Evaluation (Supplied Task)*

J-to-E	System	mWER	Fluency	Adequacy
	Top	41.8	34.8	34.1
	<b>Our</b>	<b>61.4</b>	<b>34.8</b>	<b>19.4</b>
	Bottom	<b>61.4</b>	31.0	<b>19.4</b>

C-to-E	System	mWER	Fluency	Adequacy
	Top	45.5	<b>38.2</b>	33.3
	<b>Our</b>	<b>46.9</b>	<b>38.2</b>	<b>29.5</b>
	Bottom	61.6	25.0	29.0



# Paraphrasing for **sentential variants**

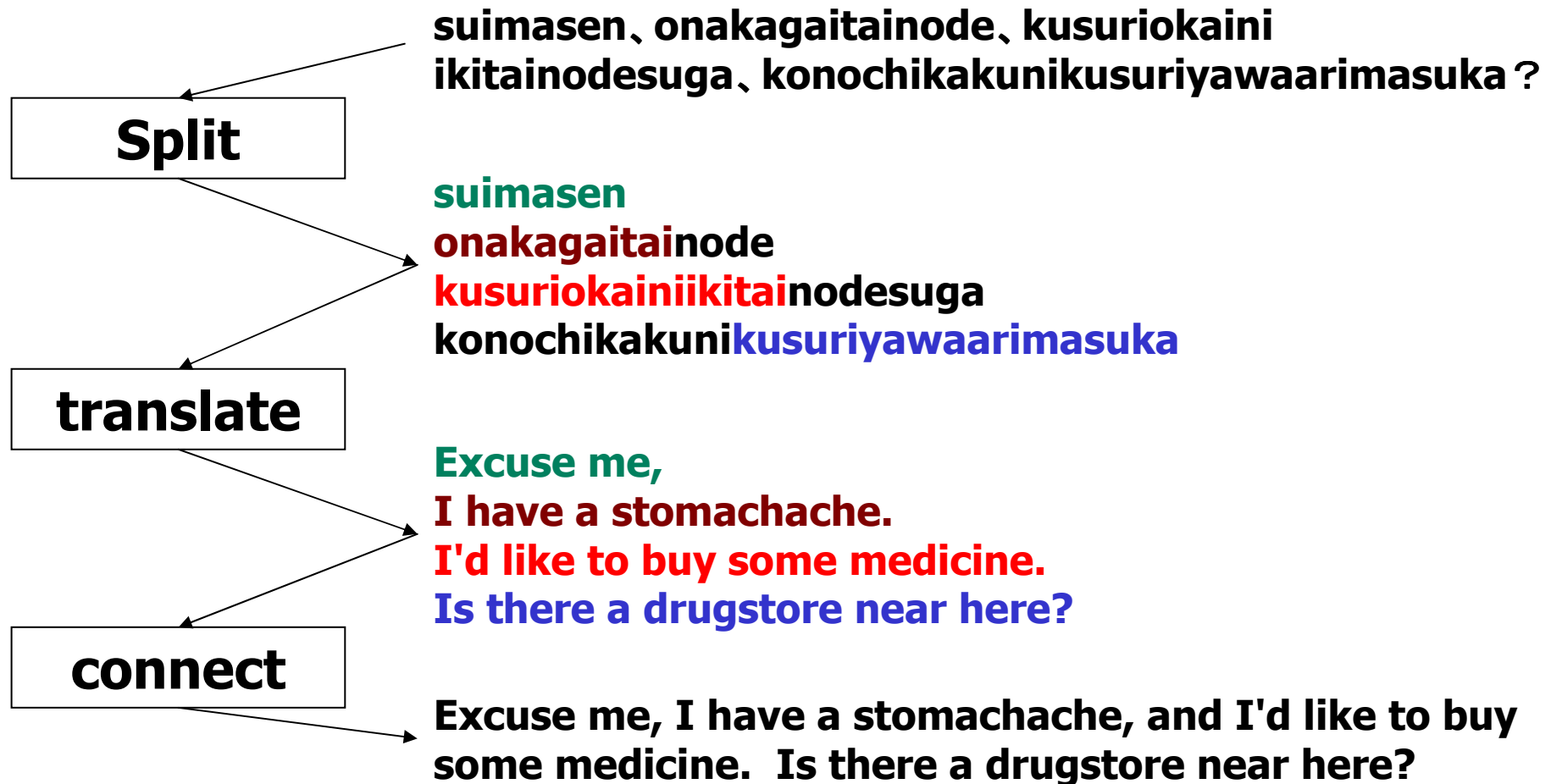
Three paraphrasers

- 1. based on *DP-matching*.**  
[Shimohata et al. 2002]
- 2. based on *SMT*.**  
[Finch et al. 2002]
- 3. based on data-oriented parsing.**  
[Finch et al. 2004]

Two Effects

- **Increased coverage and reduced word error rate** for MT.
- **Effective expansion of reference sentences** for translation evaluation.

# Paraphrasing for long sentence translation



Doi, T.: "Splitting input sentences for machine translation using language model with sentence similarity," Coling, 2004.

# Concluding Remarks

- (1) hybridization of multiple EBMTs followed by a statistical selector,
- (2) new SMT, phrase-based HMM SMT, and
- (3) paraphrasing methods.