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**

C3

Outline

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

Unrestricted	J-to-E	
Supplied	J-to-E	C-to-E

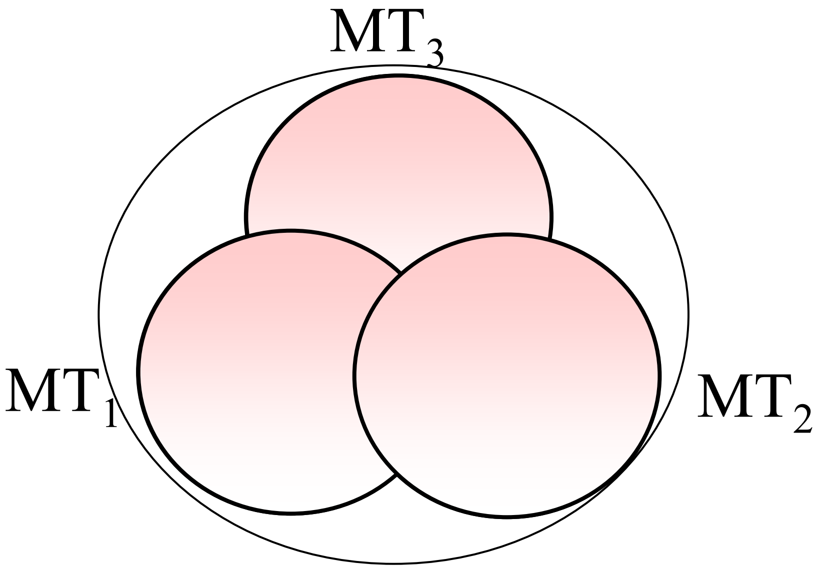
The technologies are combined in the near future.

C^3

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

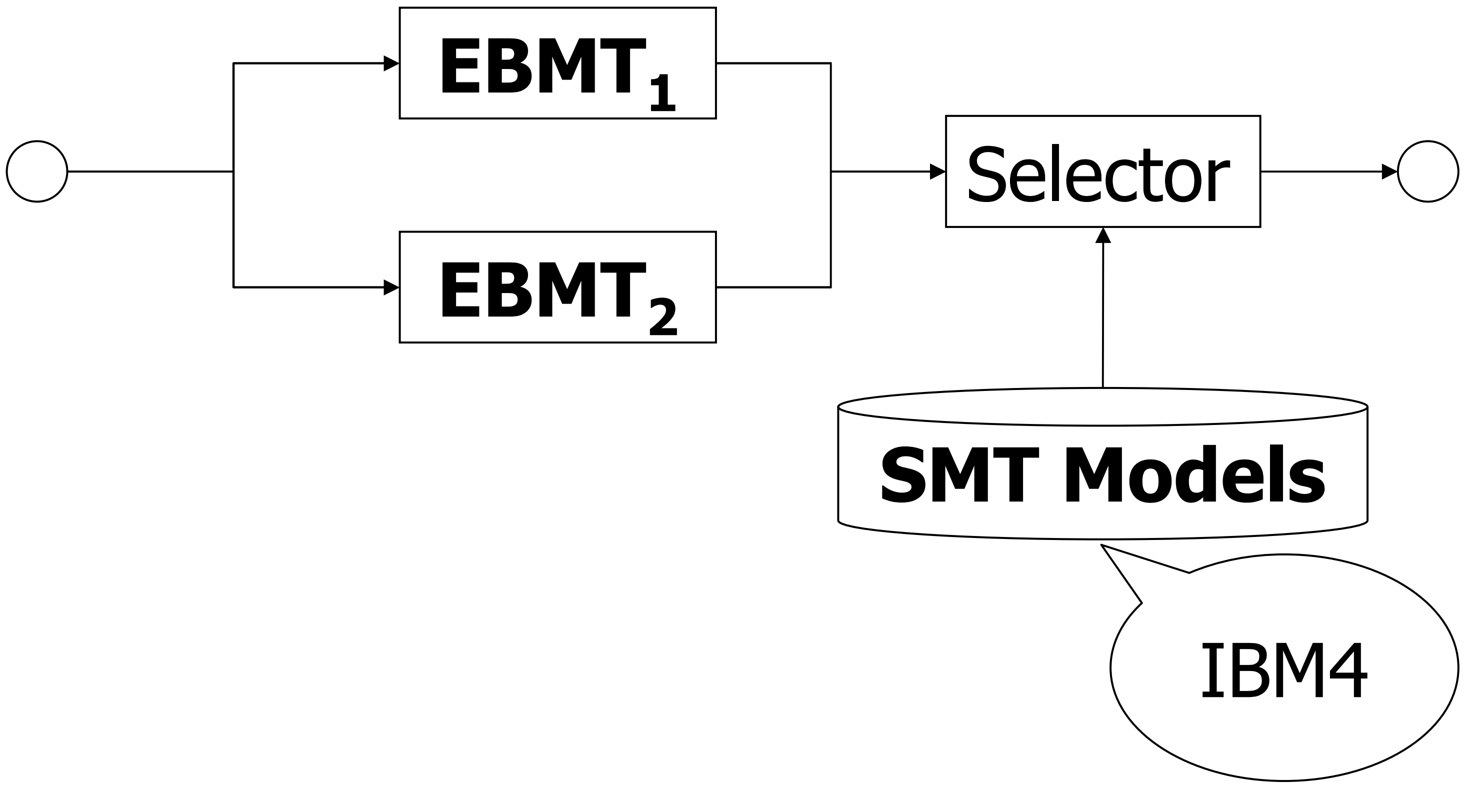
(J)	o-shiharai wa genkin desu ka kurejitto kaado desu ka	
(MT_1)	Is the payment cash? Or is it by credit card?	[**Good**]
(MT_2)	Would you like to pay by cash or credit card?	[**Perfect**]
(MT_3)	Could you cash or credit card?	[**Fair**]

So many men, so many minds.



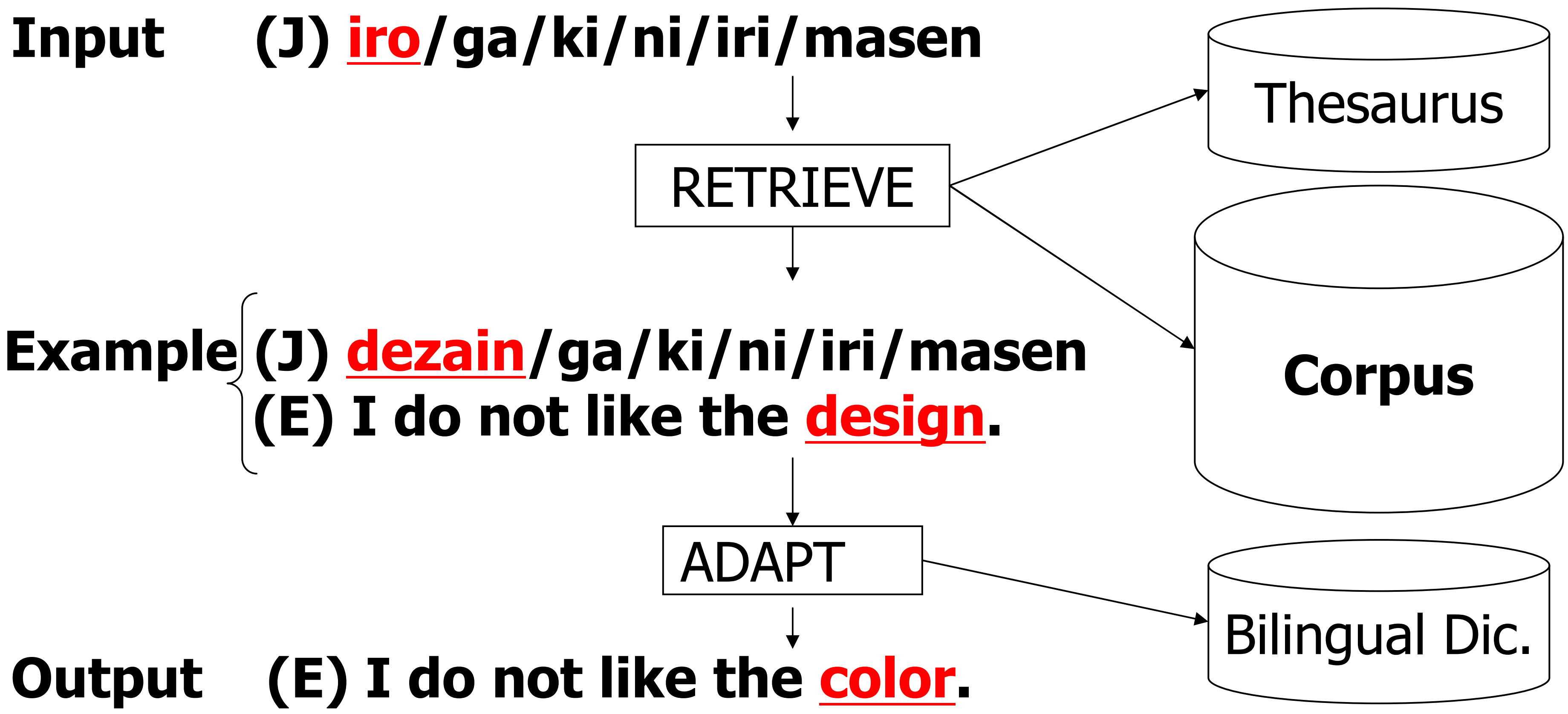
Today's Hybridization

"Multiple **EBMTs Followed By**
A Selector Based On **SMT Models**"



C3

EBMT$_1$
D^3(Dp-match Driven transDucer)



C3

EBMT_1
D^3 (2)

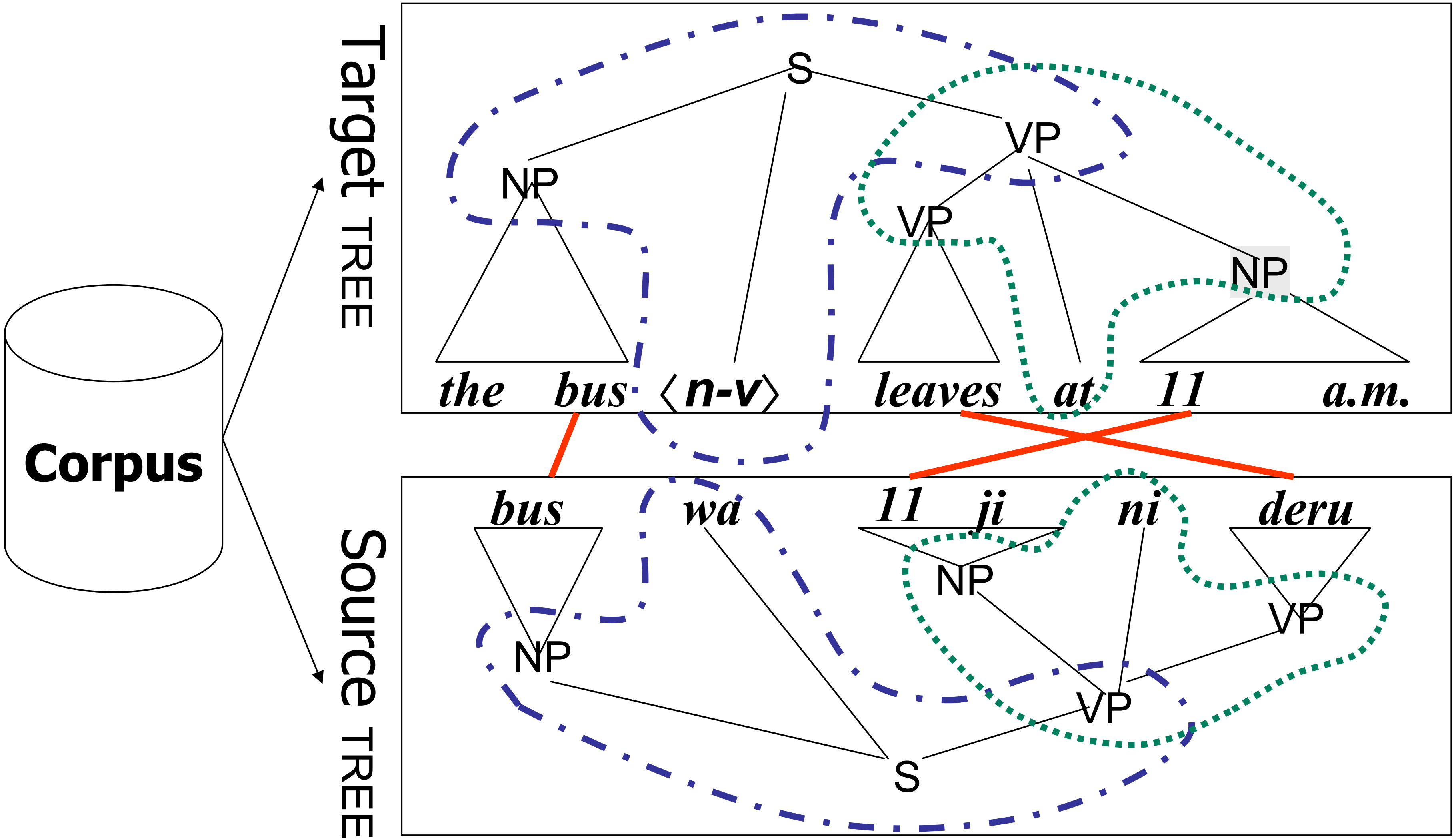
$$Distance = \frac{I + D + 2\sum SEMDIST}{L_{input} + L_{example}}$$

Height of Most Specific Common Abstraction in the thesaurus

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

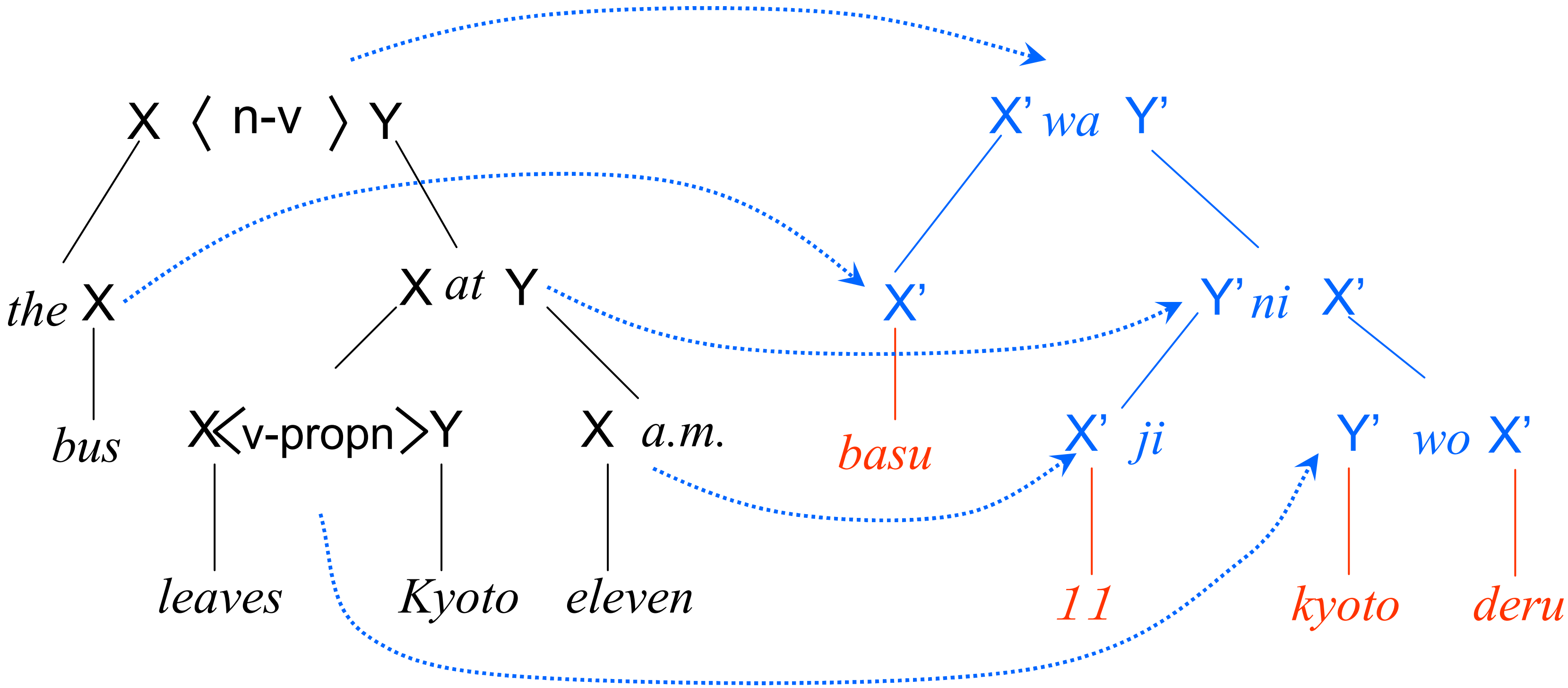
Height of the Thesaurus

EBMT$_2$ Hierarchical Phrase Alignment based Translation (HPAT)



C3

EBMT$_2$
HPAT(2)



(1) Parse source language using source patterns.

(2) Map source patterns to target patterns.

(3) Translate leaves by referring to a dictionary.

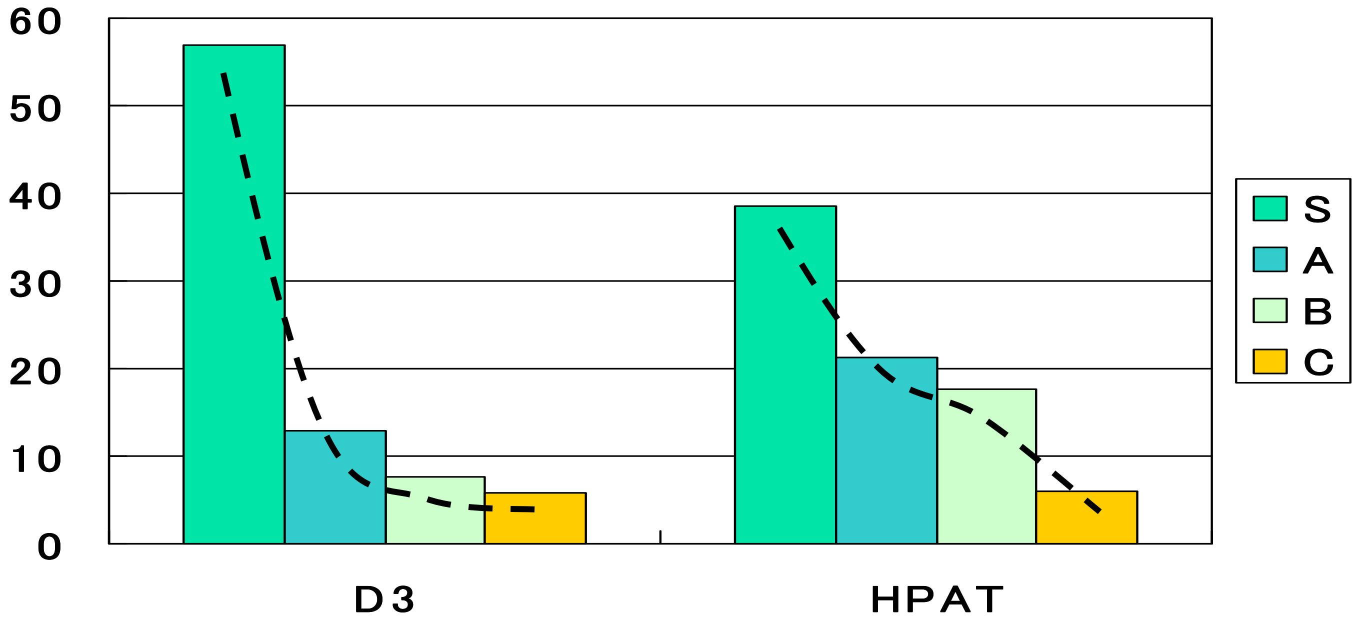
Comparison of Two EBMTs

Table 2: *Features of the two EBMTs*

	D3	**HPAT**
Unit	*Sentence*	*Grammatical Unit*
Coverage	Narrow	Wide
Quality	Good	Modest

C3

Comparison of Two EBMTs (2)



Quality Ranks

(S) Perfect;
(A) Very Good;
(B) Good;
(C) Fair;

(D) Bad.

SMT-based Selector

- Conventional selector:
  1. **L**anguage **M**odel

- Our selector:
  1. **L**anguage **M**odel
  2. **T**ranslation **M**odel
  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*

	D3	HPAT	SELECT	DIFF.
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

C3

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)

$$\hat{\mathrm{e}} = \arg\max_{\mathrm{e}} P(\mathrm{e} \mid \mathrm{f}) \quad (1)$$

$$\hat{\mathrm{e}} = \arg\max_{\mathrm{e}} P(\mathrm{f} \mid \mathrm{e}) P(\mathrm{e}) \quad (2)$$

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

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

Phrase Segmentation　　Phrase Translation　　Phrase Ngram

Phrase Ngram Model

$$P(\overline{\mathrm{e}} \mid \mathrm{e}) \approx \prod_i P(\overline{\mathrm{e}}_i \mid \overline{\mathrm{e}}_{i\text{-}1}) \qquad (6)$$

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

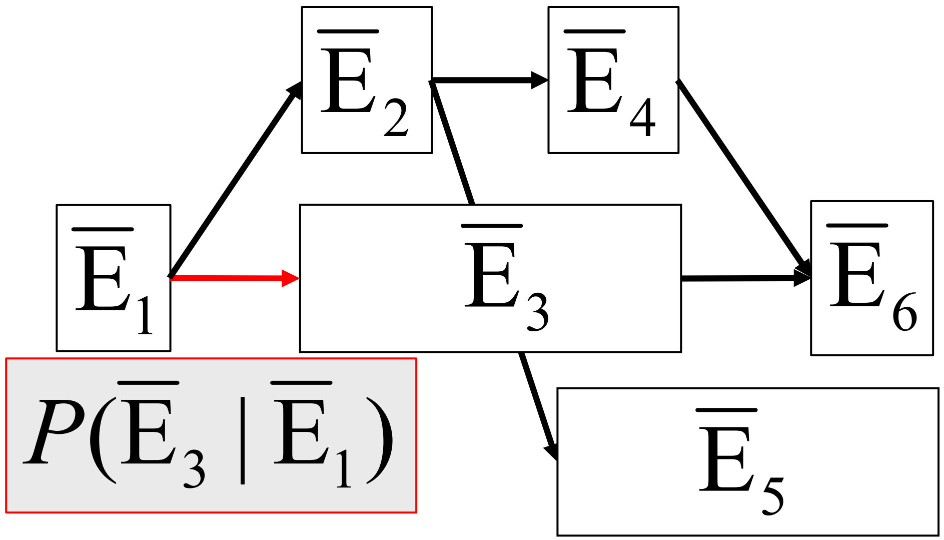


Figure 1: *Phrase Ngram Model*

C^3

Phrase Segmentation Model

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

$$P(\mathrm{f} \mid \bar{\mathrm{f}}, \bar{\mathrm{e}}, \mathrm{e}) \propto P(\bar{\mathrm{f}} \mid \mathrm{f}) \approx \prod_j P(\bar{\mathrm{f}}_j \mid \mathrm{f})$$

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

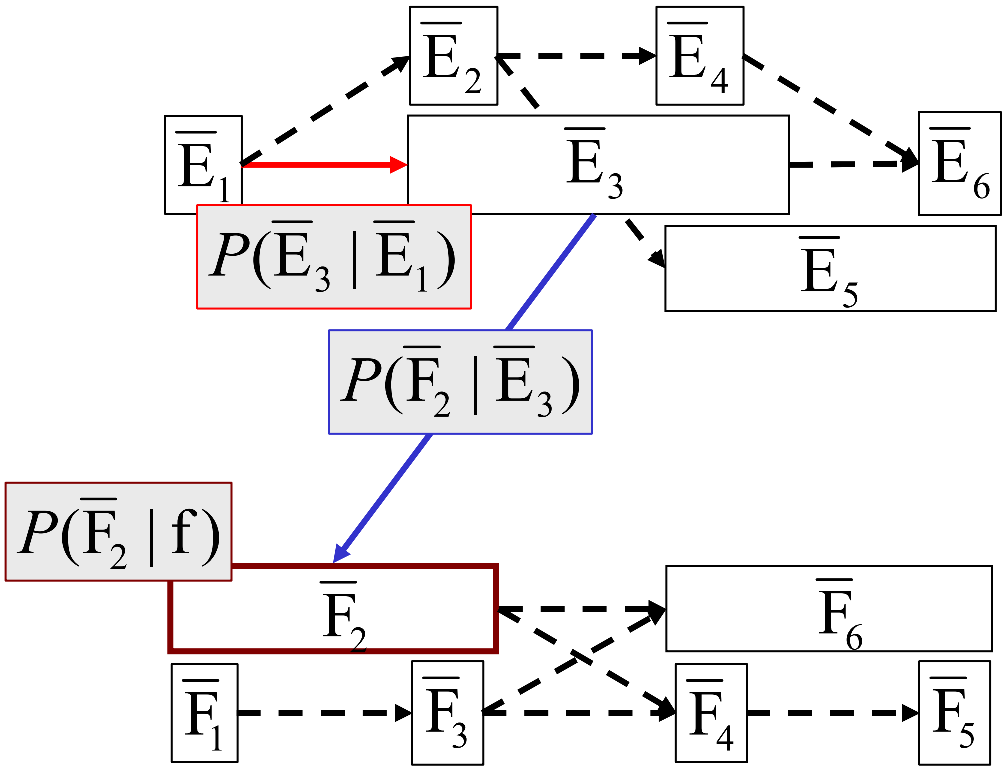
Phrase Translation Model

$$P(\bar{\mathrm{f}} \mid \bar{\mathrm{e}}, \mathrm{e}) \approx \prod_{j} P(\bar{\mathrm{f}}_j \mid \bar{\mathrm{e}}_{\mathrm{a}_j}) \quad (11)$$

Combined

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

Phrase Segmentation Phrase Translation Phrase Ngram



C3

Parameter Estimation

- Please consult Section 3.5 in the paper.

C3

Phrase Segment Induction

Extract phrase pair using the following criterion:

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

C^3

Decoding

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

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

C3

Results: **Supplied task** JE and CE

Table 8+: *Evaluation (Supplied Task)*

	System	mWER	Fluency	Adequacy
J-to-E	Top	41.8	34.8	34.1
	Our	**61.4**	**34.8**	**19.4**
	Bottom	**61.4**	31.0	**19.4**

	System	mWER	Fluency	Adequacy
C-to-E	Top	45.5	**38.2**	33.3
	Our	**46.9**	**38.2**	**29.5**
	Bottom	61.6	25.0	29.0

C3

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

suimasen、onakagaitainode、kusuriokaini ikitainodesuga、konochikakunikusuriyawaarimasuka？

Split

suimasen
onakagaitainode
kusuriokainiikitainodesuga
konochikakunikusuriyawaarimasuka

translate

Excuse me,
I have a stomachache.
I'd like to buy some medicine.
Is there a drugstore near here?

connect

Excuse me, I have a stomachache, and I'd like to buy some medicine. Is there a drugstore near here?

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.