

# Machine translation system for patent documents combining rule-based translation and statistical post-editing applied to the NTCIR-10 PatentMT Task



Terumasa EHARA  
Yamanashi Eiwa College

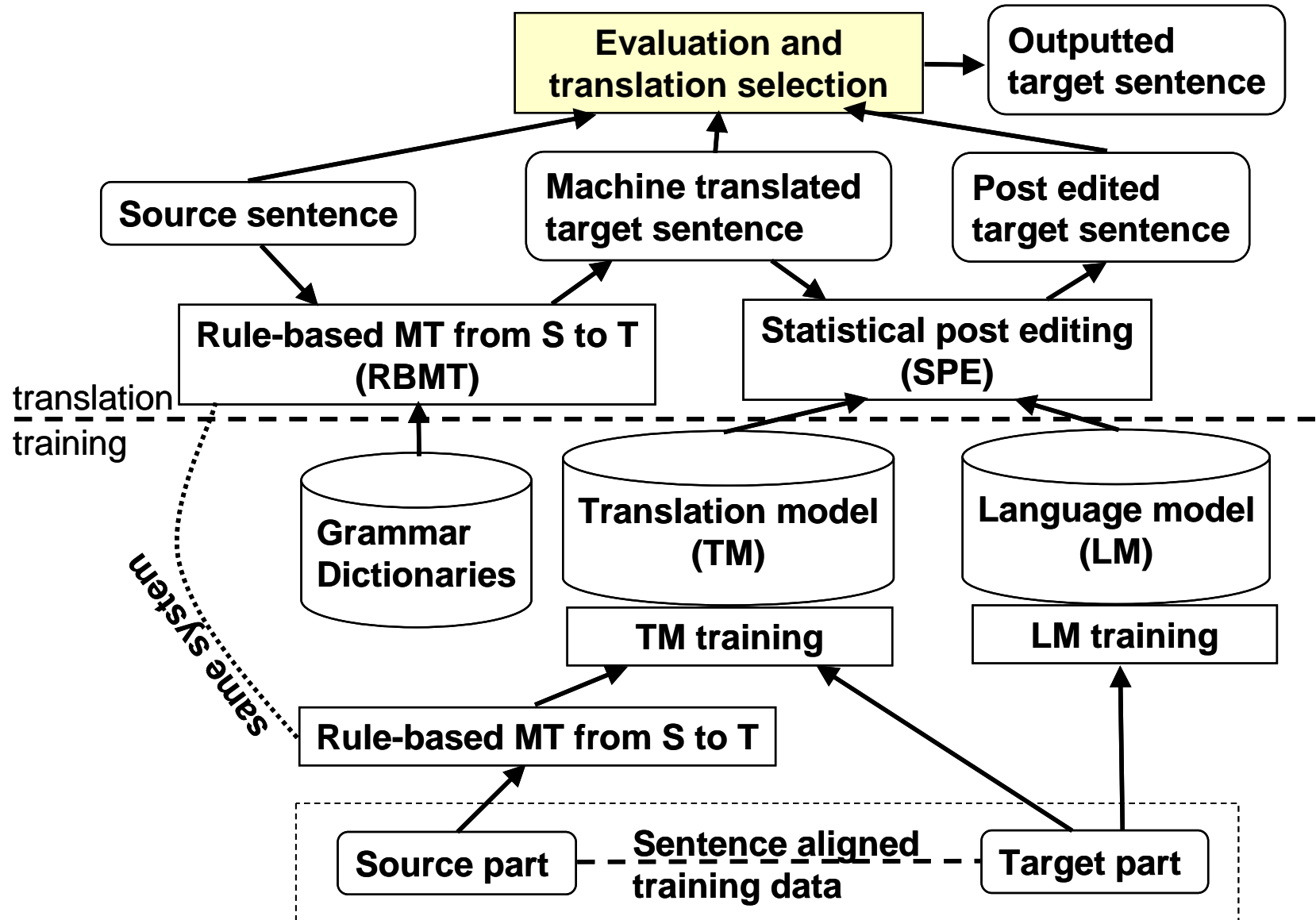
# Table of contents

- Motivation
- System architecture
- Evaluation and translation selection
- Translation model training
- Experiments
- Related works
- Conclusion

# Motivation

- Hybrid system combining rule-based MT (RBMT) and statistical post editing (SPE) may make MT more accurate.
- NTCIR-9's result shows that simple RBMT system outperforms RBMT+SPE system.
- Automatic evaluating of two outputs and selecting the best output may make MT more accurate.

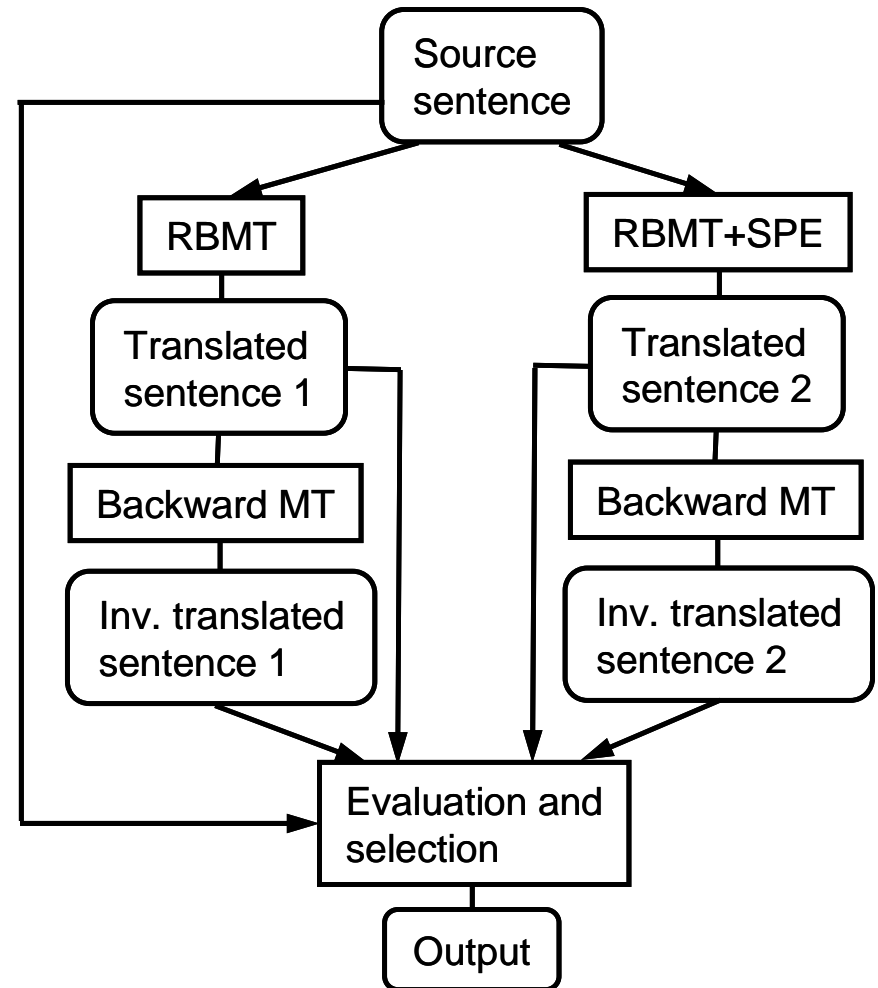
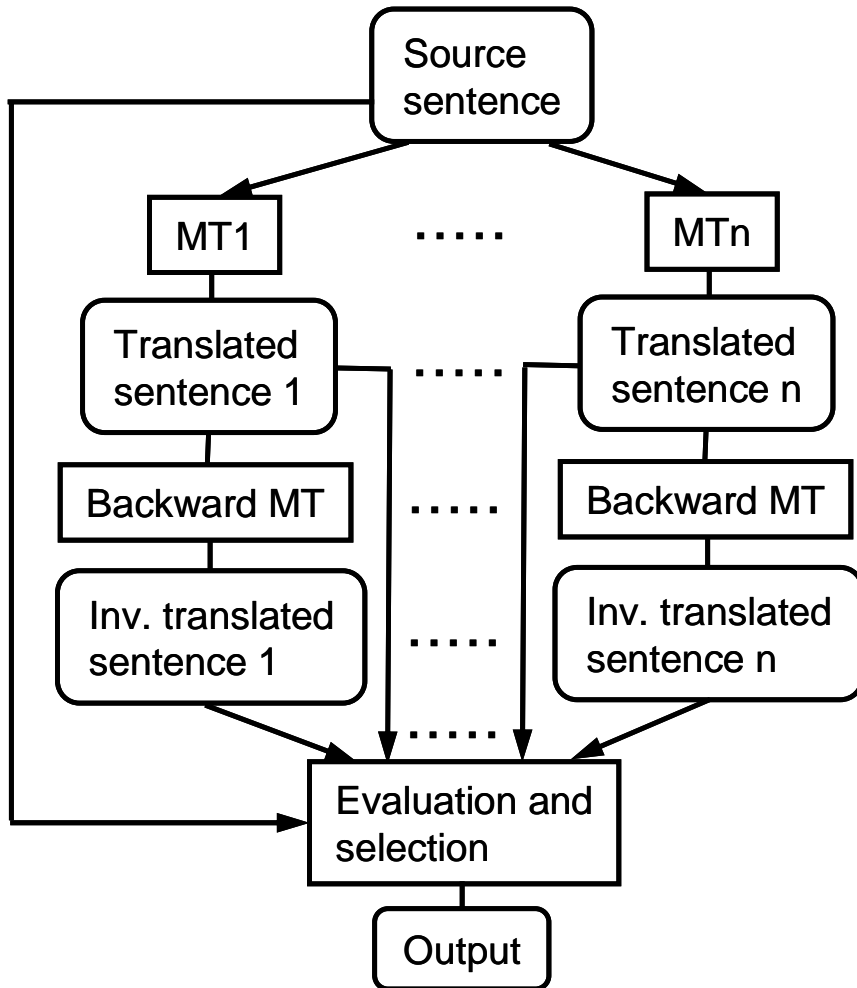
# System architecture (1)



# System architecture (2)

- Tools that are used in our experiments
  - ▶ RBMT part: commercial based MT systems
  - ▶ SPE processor: phrase-based Moses (Rev. 4343, distortion limit = 0)
  - ▶ LM training tool: Sriml (ver.1.5.5)
  - ▶ TM training tool: Giza++ (v1.0.3)

# Evaluation and translation selection (1)



# Evaluation and translation selection (2)

- Evaluation criterion: IMPACT
- Bonus score is added to the IMPACT score of SPE outputs:

Subtask	Bonus
JE	0.1
EJ	$\infty$
CE	0.2

# Evaluation and translation selection (3)

- Preliminary test results using NTCIR-9' JE subtask data:

Won system	Counts
RBMT1	97
Tie	134
EIWA	69

Won system	Counts
RBMT1	34
Tie	228
Our method	38

Adequacy

Won system	Counts
RBMT1	86
Tie	147
EIWA	67

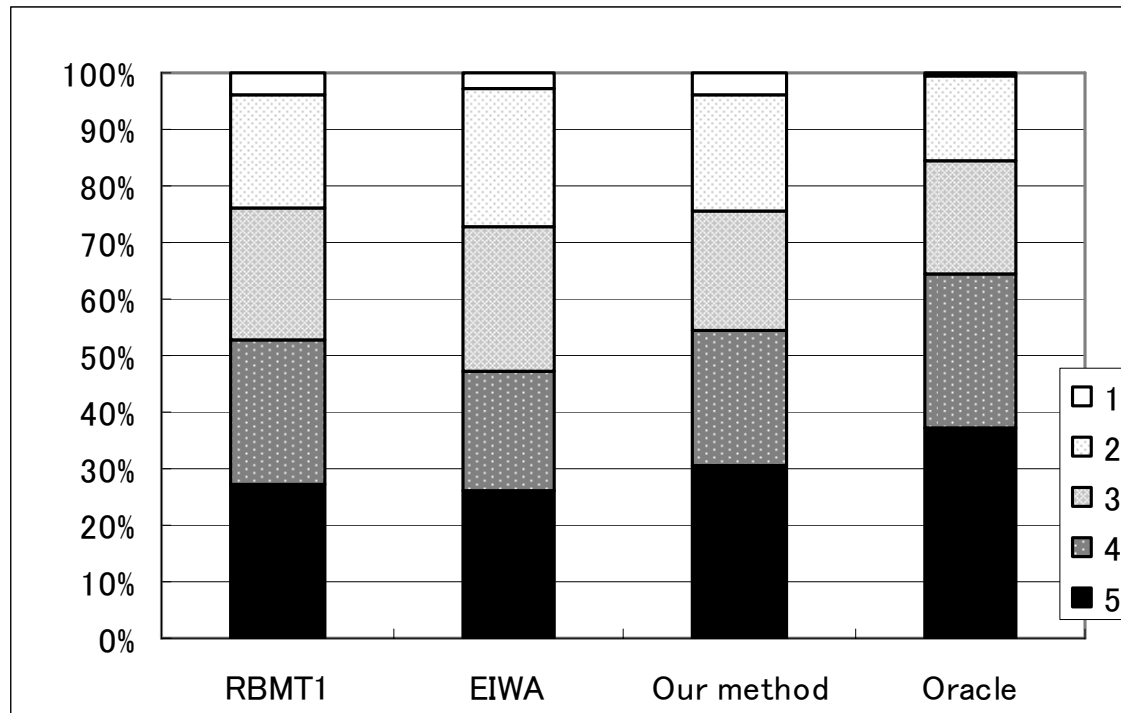
Won system	Counts
RBMT1	26
Tie	235
Our method	39

Acceptability



# Evaluation and translation selection (4)

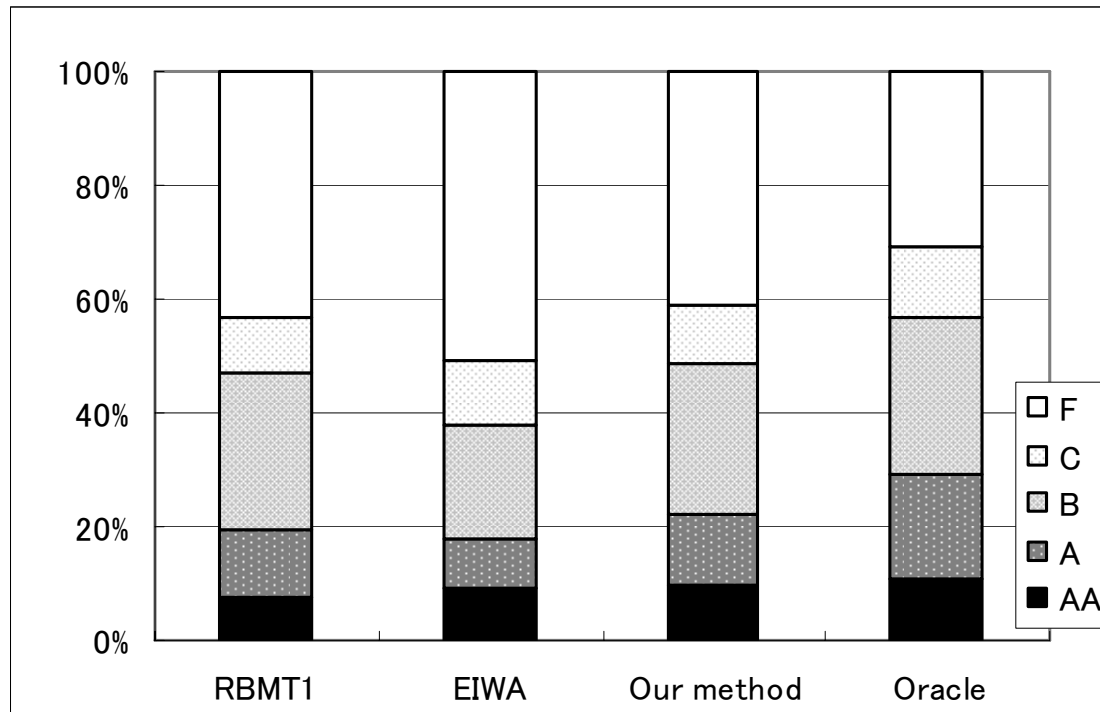
- Preliminary test results using NTCIR-9' JE subtask data:



Adequacy

# Evaluation and translation selection (5)

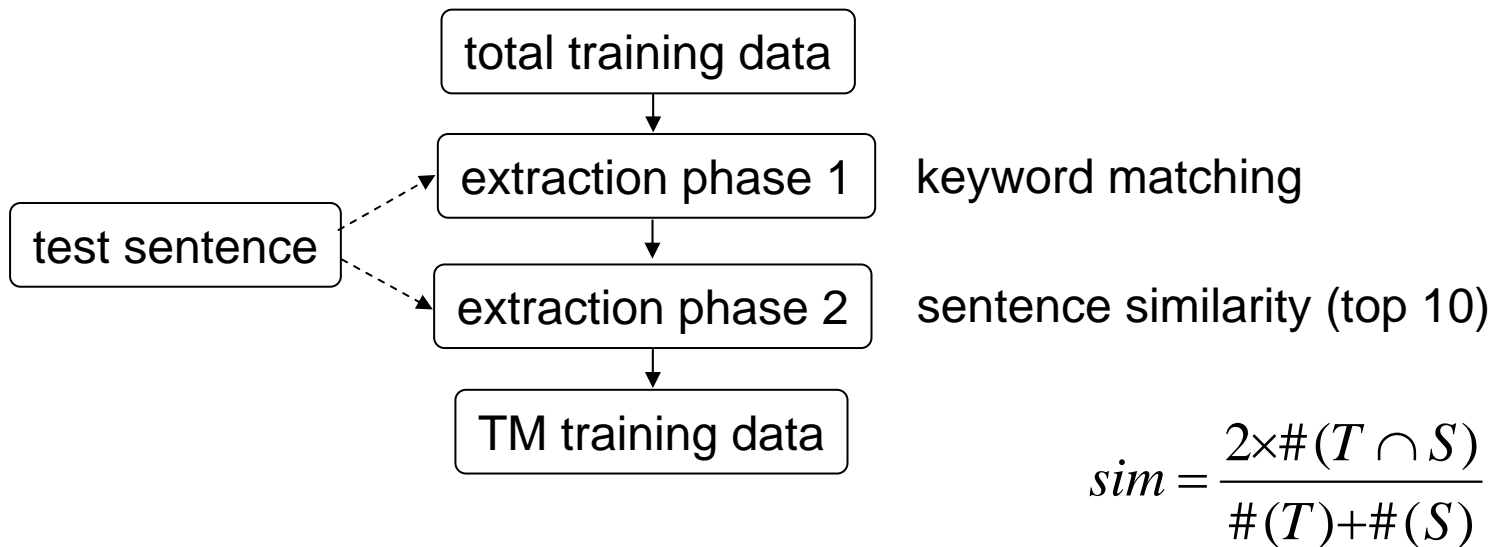
- Preliminary test results using NTCIR-9' JE subtask data:



Acceptability

# Translation model training (1)

- We extract matched data to the test sentences from total training data.
- Matching algorithm is :



# Translation model training (2)

Subtask	Phase and Eval.	Test/dev sentences	Selected training sentences
JE	Development	2,000	253,333
	Test (IE PEE)	2,543	357,443
EJ	Development	2,000	181,000
	Test (IE)	2,300	205,460
	Test ChE	2,000	183,663
CE	Development	2,000	115,528
	Test (IE)	2,300	99,732
	Test (ME PEE)	2,282	126,321

# Experimental results (1)

- Preliminary experiment using NTCIR-9's JE subtask data

	RBMT1	EIWA	Our method	Oracle
Adequacy (average)	3.530	3.430	3.563	3.853
Acceptability (rate of C or higher)	0.57	0.49	0.59	0.69

# Experimental results (2)

- Human judgment (intrinsic evaluation)

	EIWA
Adequacy (average)	2.80
Acceptability (rate of C or higher)	---

CE subtask

	RBMT1	EIWA
Adequacy (average)	3.57	3.53
Acceptability (rate of C or higher)	---	0.44

JE subtask

	RBMT4	EIWA
Adequacy (average)	---	3.42
Acceptability (rate of C or higher)	---	0.59

EJ subtask

# Experimental results (3)

- Automatic evaluation (intrinsic evaluation)

	EIWA
RIBES	0.7403
BLEU	0.2690
NIST	7.5480

CE subtask

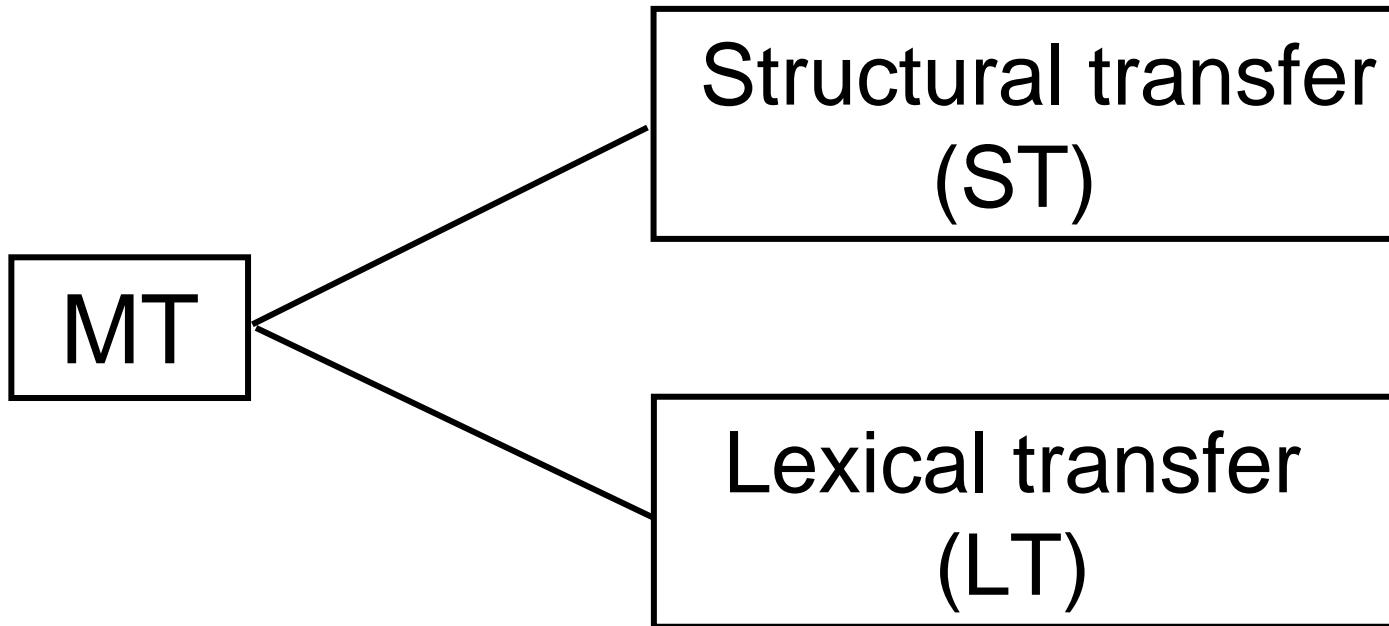
	RBMT1	EIWA
RIBES	0.7106	0.7402
BLEU	0.2035	0.3250
NIST	6.7520	8.2700

JE subtask

	RBMT4	EIWA
RIBES	0.7111	0.7692
BLEU	0.2244	0.3693
NIST	6.2950	8.5010

EJ subtask

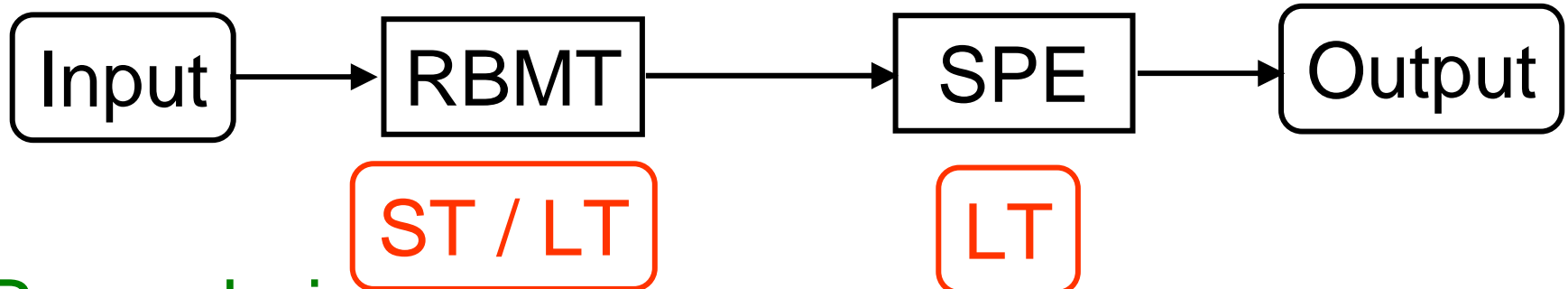
# Related works (1)



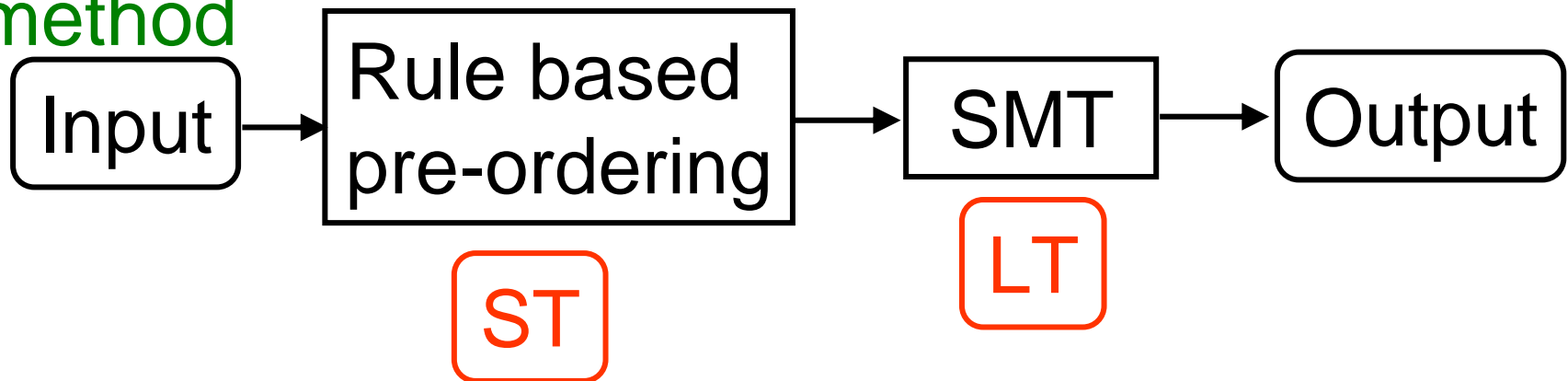


# Related works (2)

## Our method



## Pre-ordering method



# Conclusion

- Combining rule-based machine translation and statistical post-editing, we can improve automatic evaluation score.
- Human judgment score of simple rule-based system is higher than our method, but the difference is not statistically significant.

# Future work

- To improve the parsing accuracy in the RBMT part.
- Syntactically collapsed outputs from the RBMT part can't be recovered by the SPE part.