Constructing a Speech Translation System using Simultaneous Interpretation Data

Hiroaki Shimizu, Graham Neubig, Sakriani Sakti,

Tomoki Toda, Satoshi Nakamura

Nara Institute of Science and Technology (NAIST), Japan

December 6th 2013

NAIST AHCLAB

Background

Speech translation

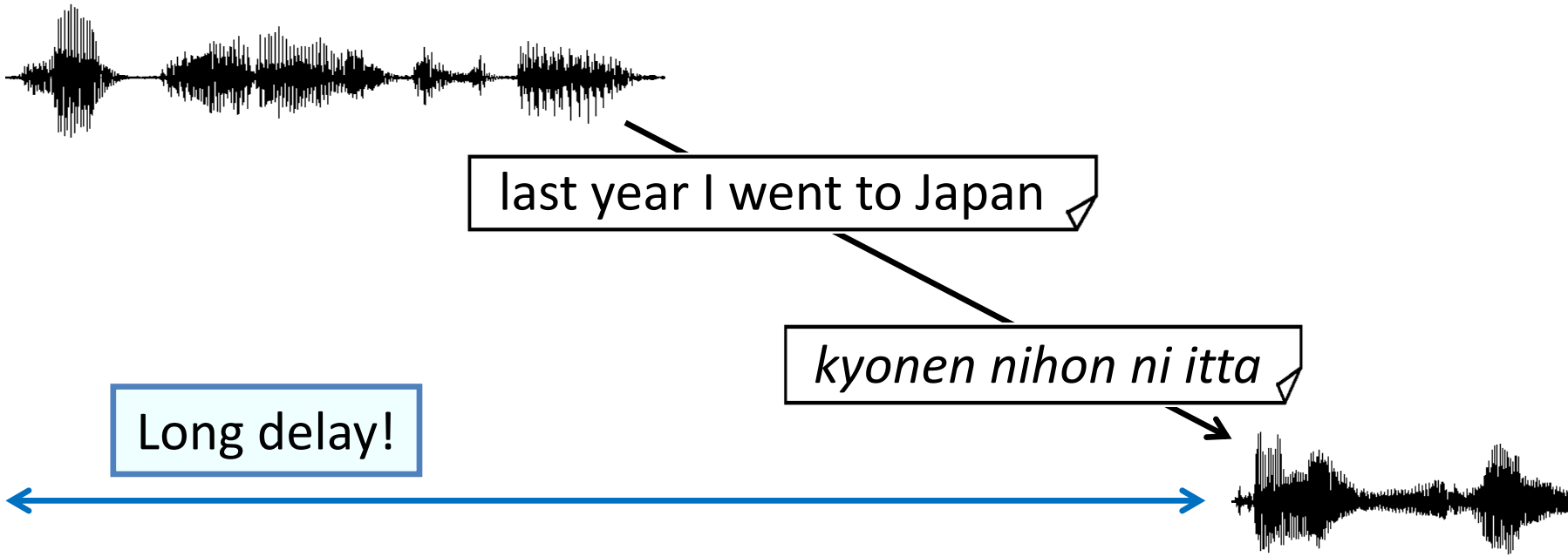
Human interpreters

- What is the source of this difference?
  - accuracy
  - delay

We focus on the problem of delay.

What is the problem of delay?

- Speech translation

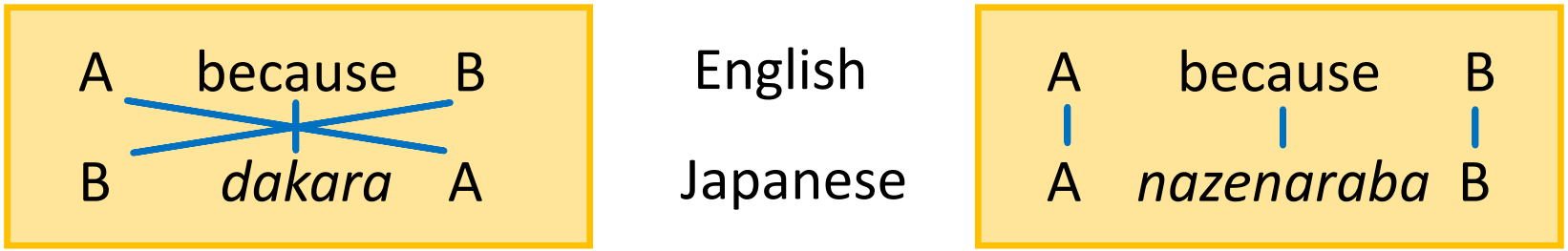


When simultaneous interpreters interpret lectures in real time, they perform **a variety of techniques** to shorten the delay.

Techniques of human simultaneous interpreters

- 1) Salami technique [Jones 02]
  - Divide longer sentences up into a number of shorter ones

- 2) Adjusting lexical choice
  - Reduce word reordering



Translator

Simultaneous interpreter

Techniques of human simultaneous interpreters

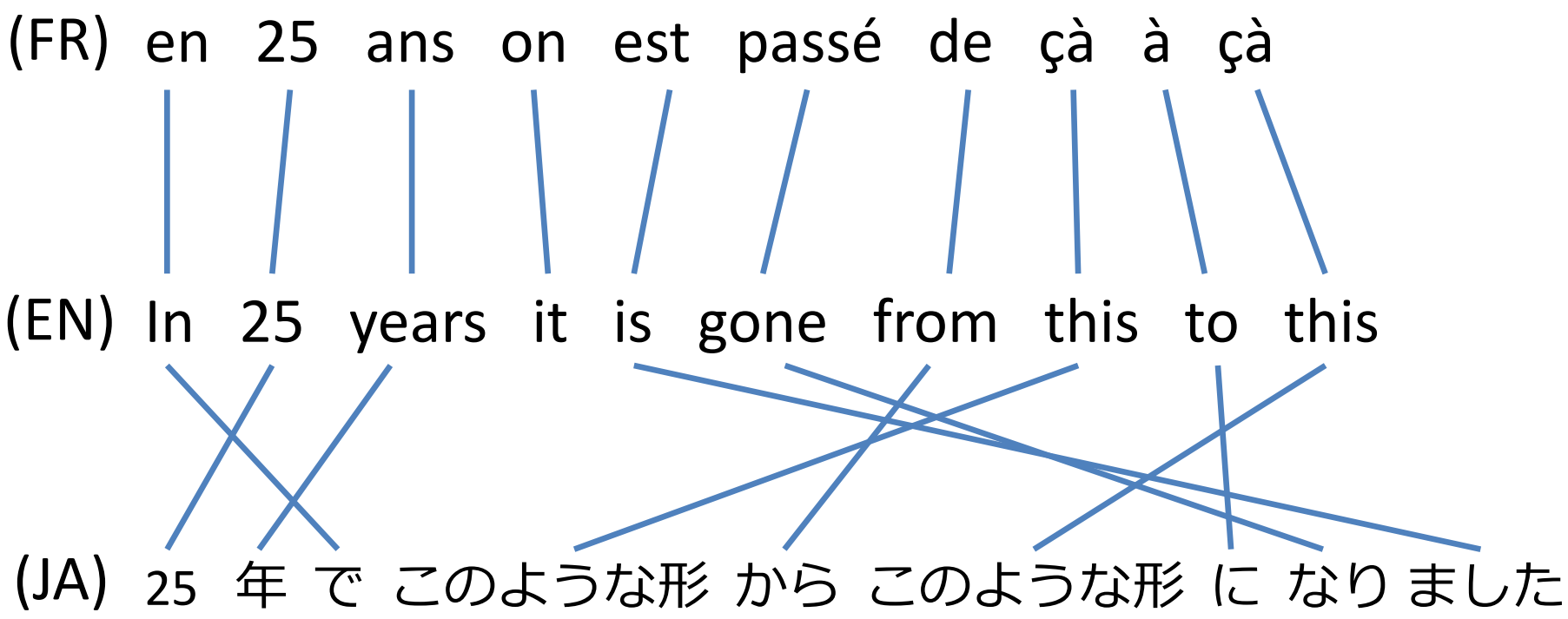
- 1) Salami technique [Jones 02]
  - Divide longer sentences up into a number of shorter ones

Similar techniques have been proposed for automatic speech translation [Fügen+ 07] [Bangalore+ 12] [Fujita+ 13]

- 2) Adjusting lexical choice
  - Reduce word reordering

No similar techniques have been proposed for automatic speech translation

Difficulty of English-Japanese



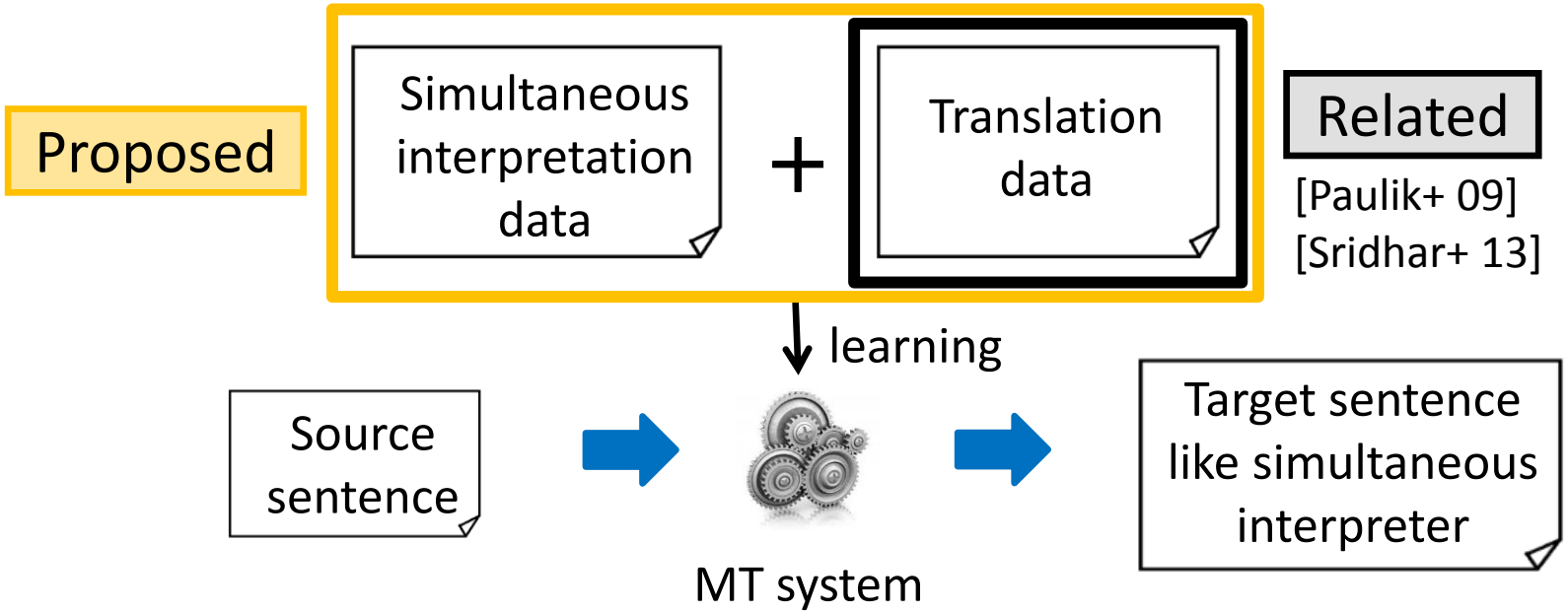
The word order is quite different between English and Japanese.

Purpose

Research purpose

Figure out what speech translation can learn from simultaneous interpreters

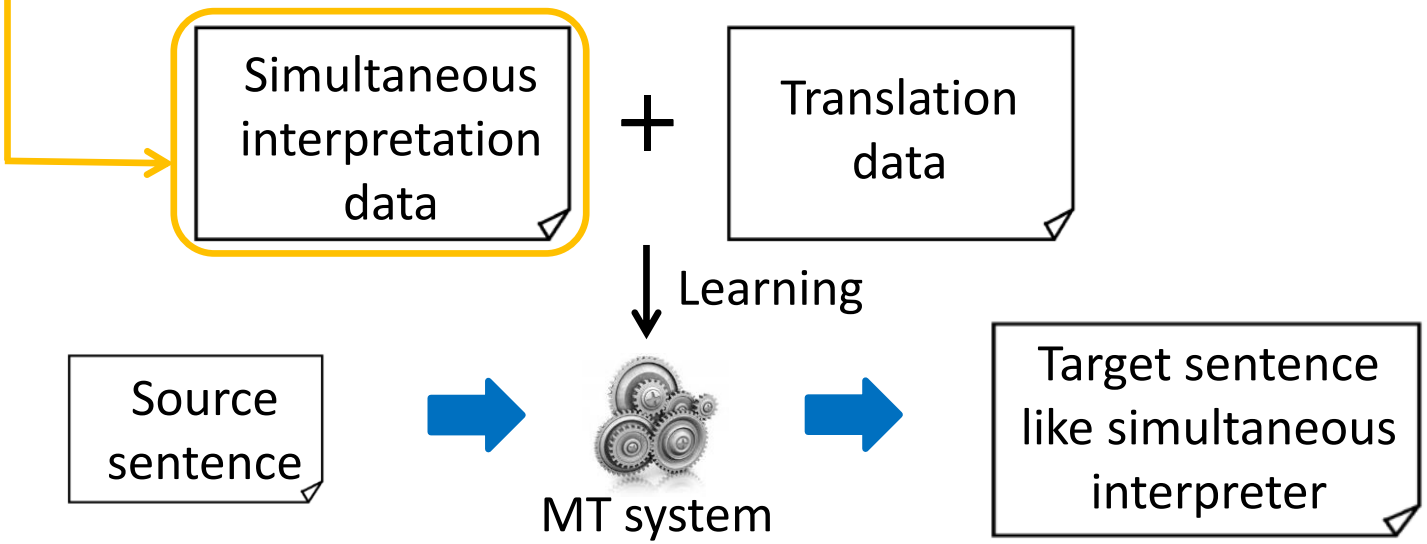
- ST system overall view



Overview

1) Collecting simultaneous interpretation data

2) Examining difference between simultaneous interpretation and translation data



3) Using the simultaneous interpretation data

4) Experiment and Result

Simultaneous interpretation data

- Materials
  - TED (English → Japanese)

Possible to compare translated subtitles with simultaneous interpretation data

- Interpreters
  - Three human simultaneous interpreters
  - Different experience levels

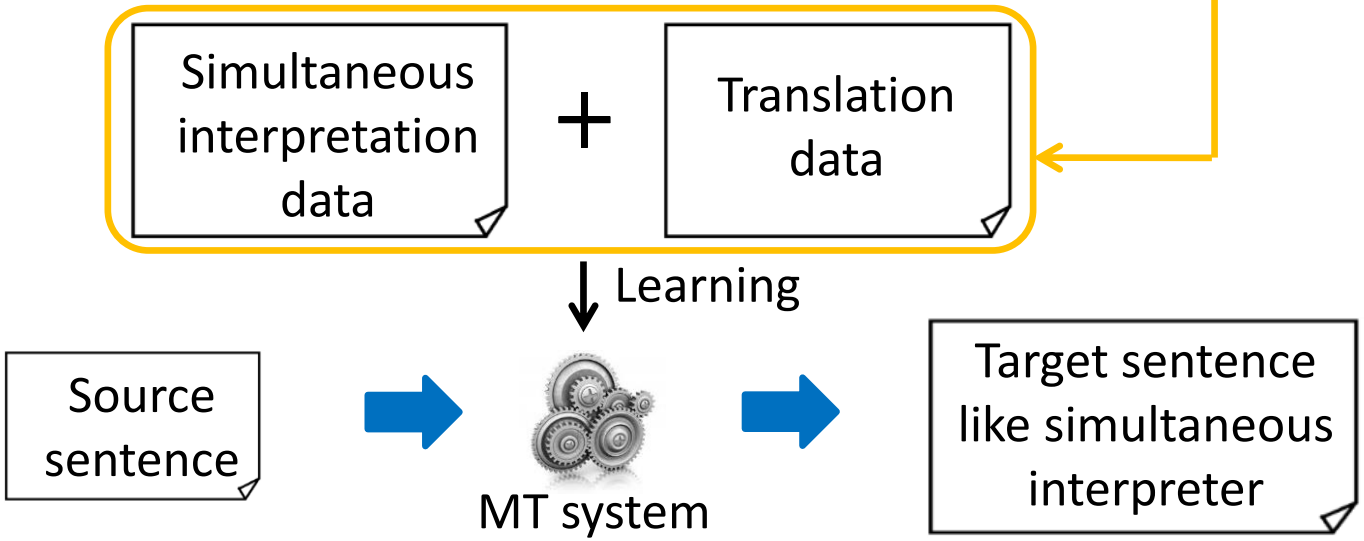
Experience	Rank
15 years	S rank
4 years	A rank
1 year	B rank

Allow us to compare characteristics of human simultaneous interpreters of different levels

Overview

1) Collecting simultaneous interpretation data

2) Examining difference between simultaneous interpretation and translation data



3) Using the simultaneous interpretation data

4) Experiment and Result

Difference between simultaneous interpretation data and translation data

- Motivation

Translation

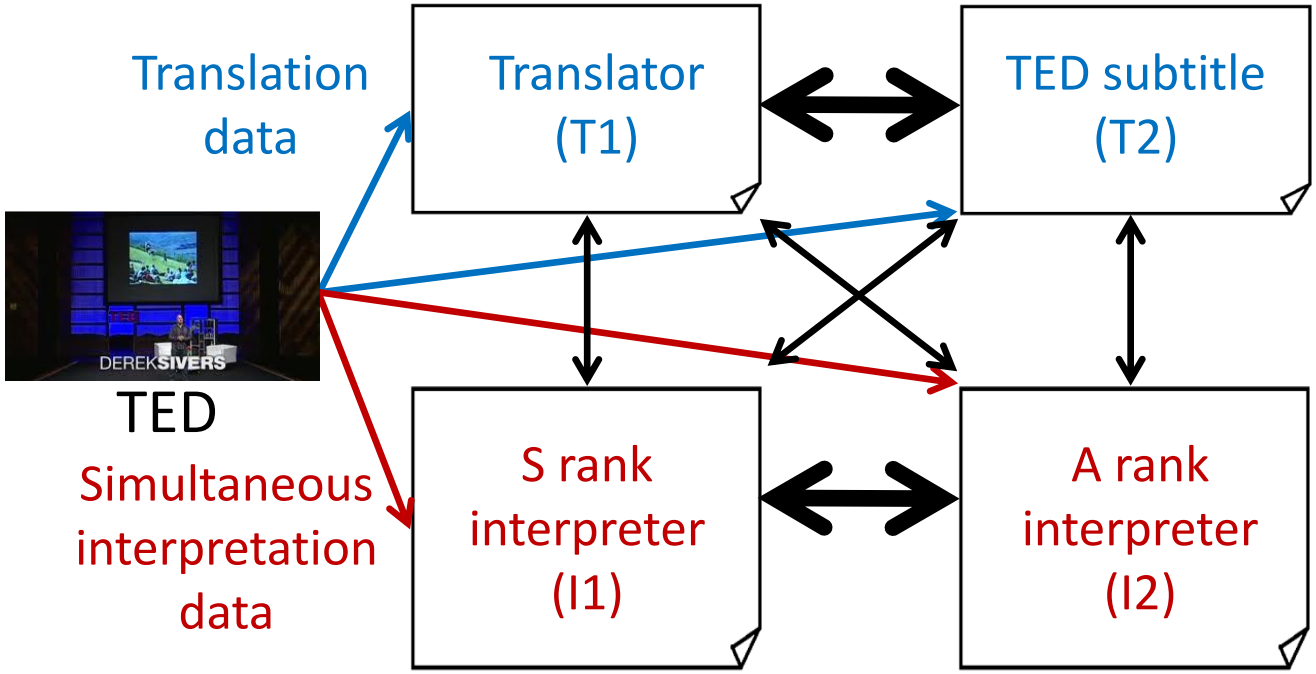
Time-unconstrained

Simultaneous interpretation

Time-constrained
Including tricks

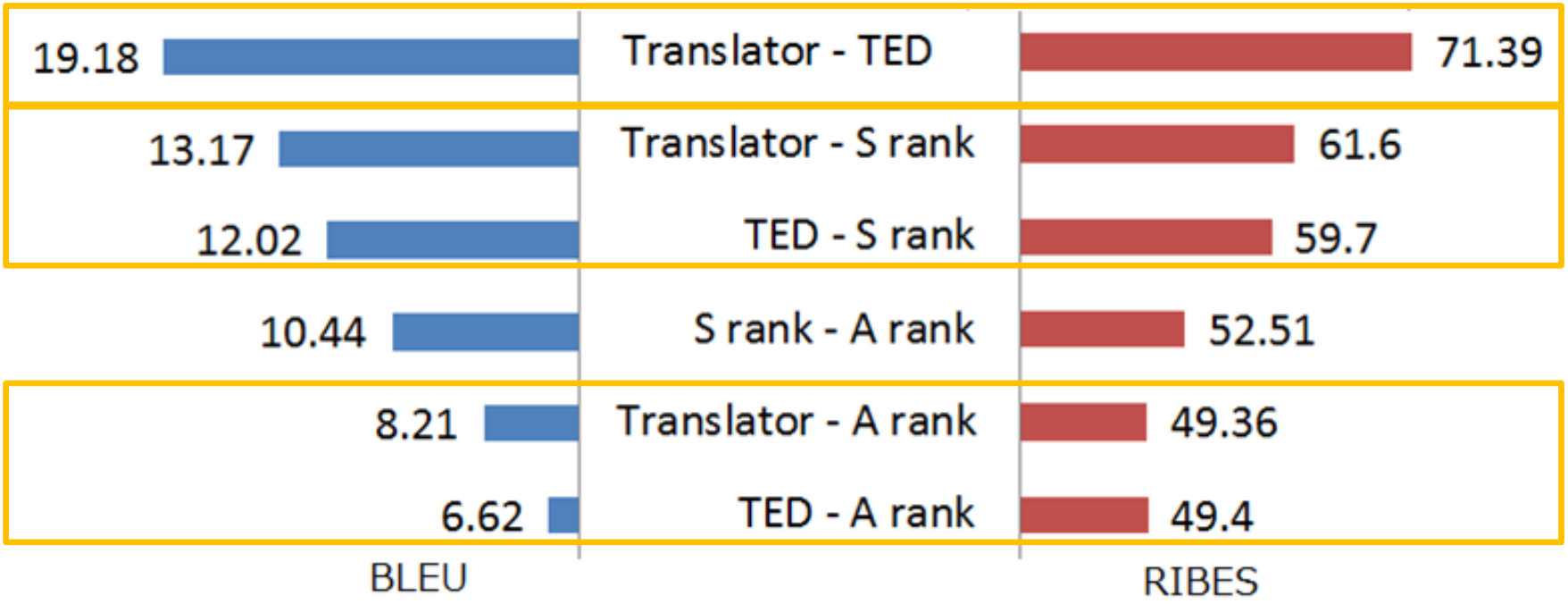
We compare translation data with the simultaneous interpretation data to find the difference.

Preliminary experiment design



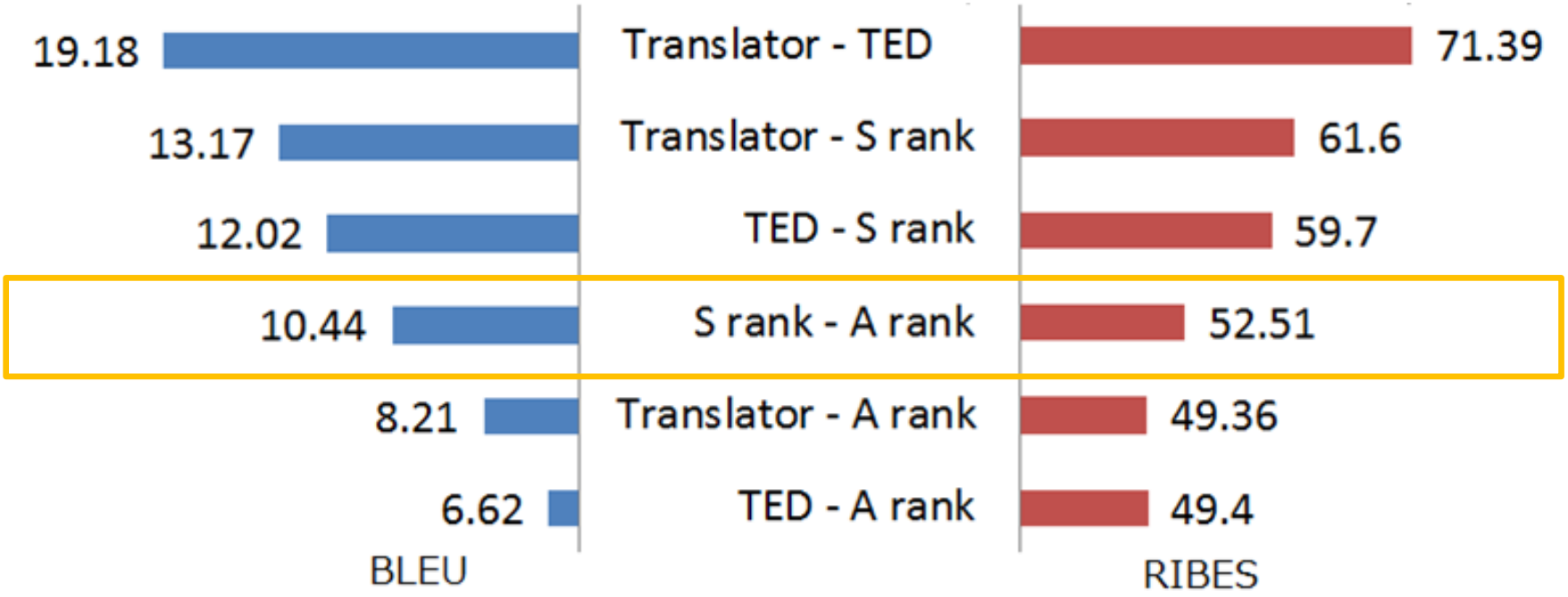
We hypothesize the similarities of T1-T2 and I1-I2 are higher than any other combinations.

Result: difference simultaneous interpretation data and translation data



- Translation data pair is highest of all combinations.
- Translation and simultaneous interpretation data pairs are lower than translation data pair.

Result: difference simultaneous interpretation data and translation data (Cont'd)



- Simultaneous interpretation data pair is unexpectedly lower.

Discussion

- The reason that simultaneous interpretation data pair is unexpectedly low

Data		Words (Ja)
Translation	Translator	4.58k
	TED subtitle	4.64k
Simultaneous interpretation	S rank	4.44k
	A rank	3.67k

S rank can interpret, but A rank cannot.

  - A rank is more similar to S rank than any others

Translation data and simultaneous interpretation data are different from the view of the similarity measures

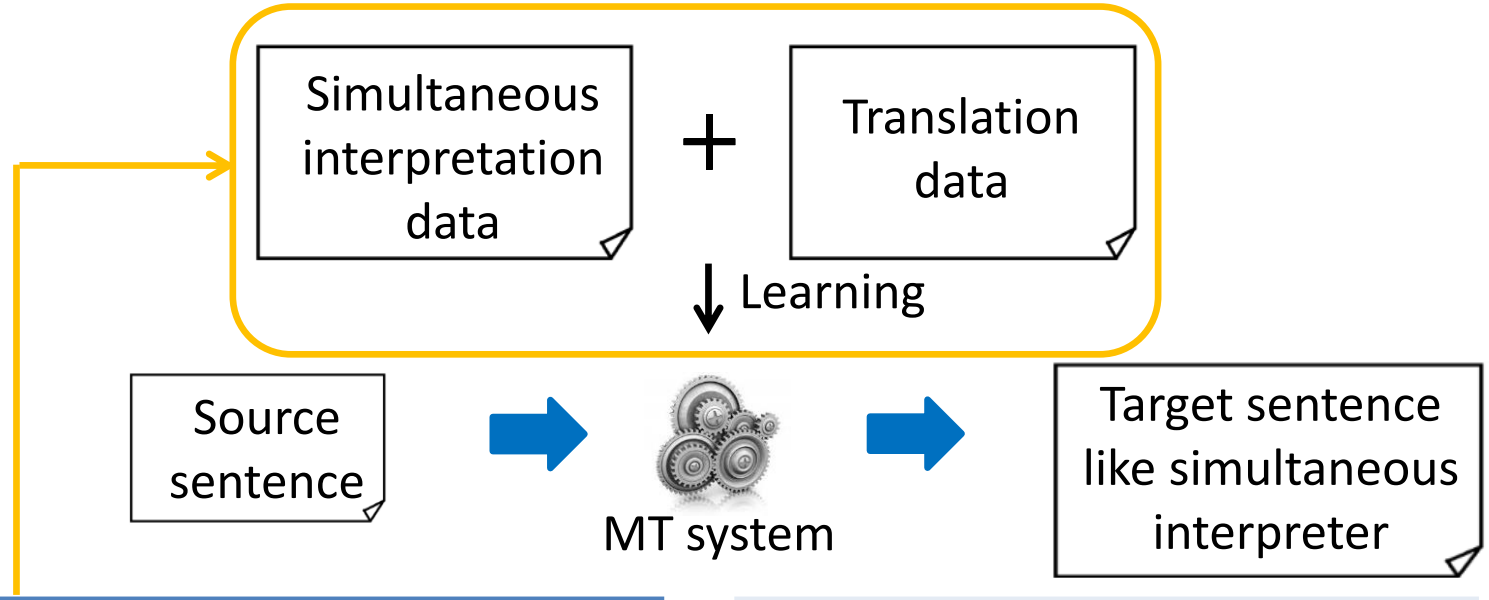
An example of simultaneous interpretation data and translation data

Data	Example
Source	and the disasters around the world have been increasing at an absolutely extraordinary and unprecedented rate
TED subtitle	そして/世界中で/大災害が/これまでに例を見ない率で/増えているのです *and / around the world / disasters / at an absolutely extraordinary and unprecedented rate / have been increasing*
S rank	世界中の/自然災害は/急速に/最近/増加しておりまして *around the world / disasters / rapidly / recently / have been increasing*
A rank	異常に/これまでにない例で *unprecedentedly / at unexpected example*

Overview

1) Collecting simultaneous interpretation data

2) Examining difference between simultaneous interpretation and translation data



3) Using the simultaneous interpretation data

4) Experiment and Result

Learning of the MT system

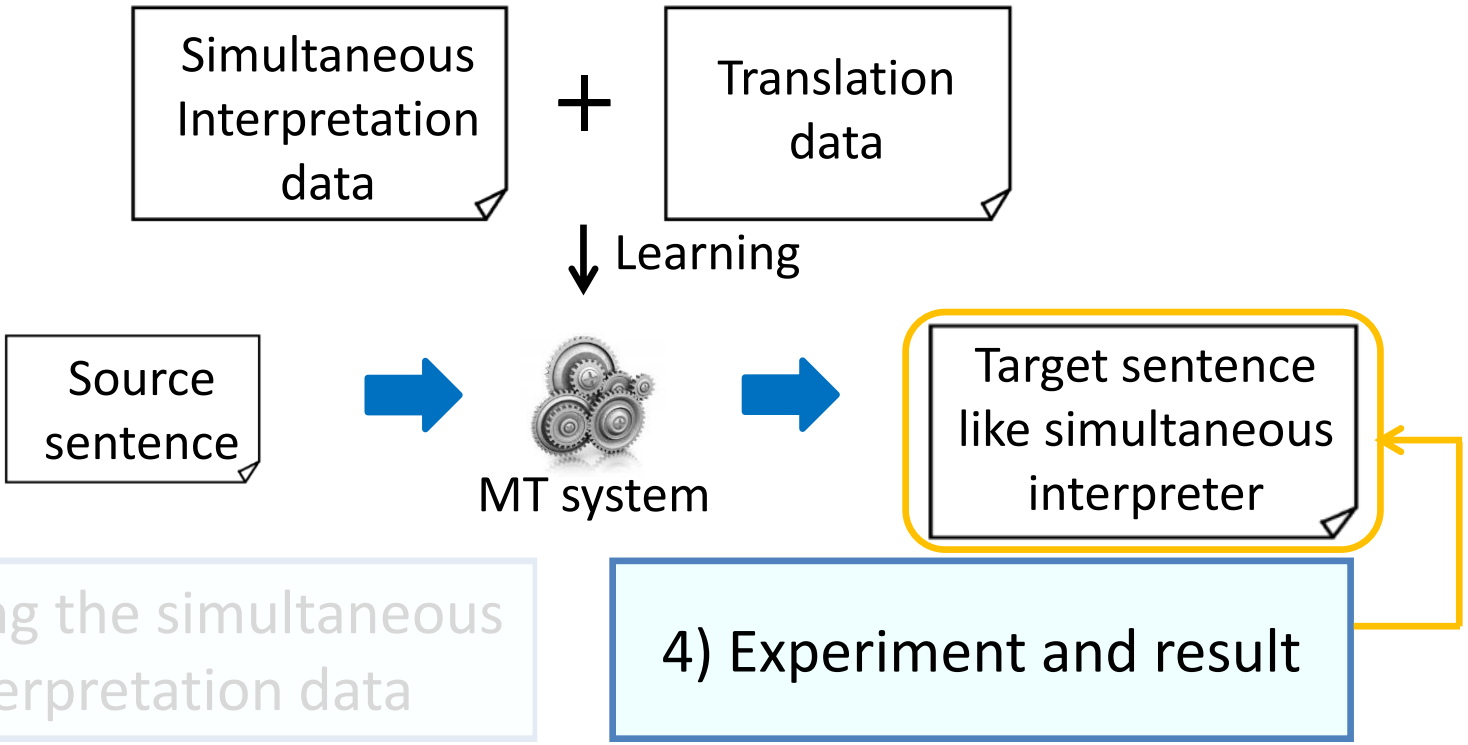
We use simultaneous interpretation data for three steps

- Tuning (Tu)
  - The parameters such as the reordering probabilities and word penalty to learn the style of simultaneous interpreters.
- Language model (LM): linear interpolation
  - The word order and lexical choice of translation is similar to simultaneous interpretation.
- Translation model (TM): fill-up [Bisazza+ 11]
  - Like LM, lexical choice is similar to simultaneous interpretation.

Overview

1) Collecting simultaneous interpretation data

2) Examining difference between simultaneous interpretation and translation data



3) Using the simultaneous interpretation data

4) Experiment and result

Data

- Task
  - TED talks (English→Japanese)

The number of words	Translation data	Simultaneous interpretation data
TM, LM (en/ja)	1.57M / 2.24M	29.7k / 33.9k
Tune (en/ja)	12.9k / 19.1k	12.9k / 16.1k
Test (en/ja)		11.5k / 14.9k

1) Using only the data from the S rank interpreter
2) Using the simultaneous interpretation data (NOT translation data) in evaluation

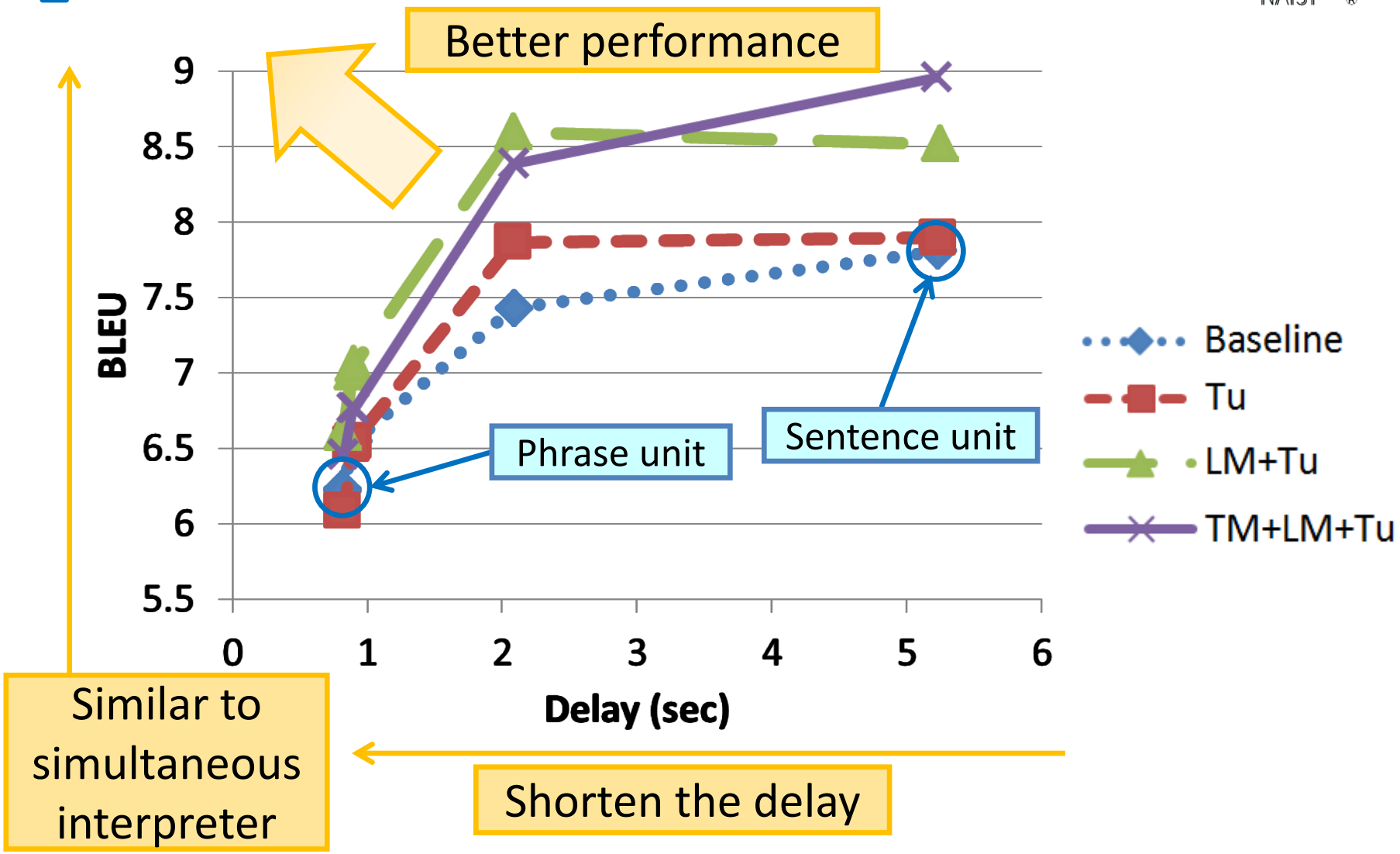
Experimental setup

- Dividing method using right probability [Fujita+ 13]
  - Method for deciding translation timing
  - Allow us to adjust the speed and accuracy of translation
- Evaluation method

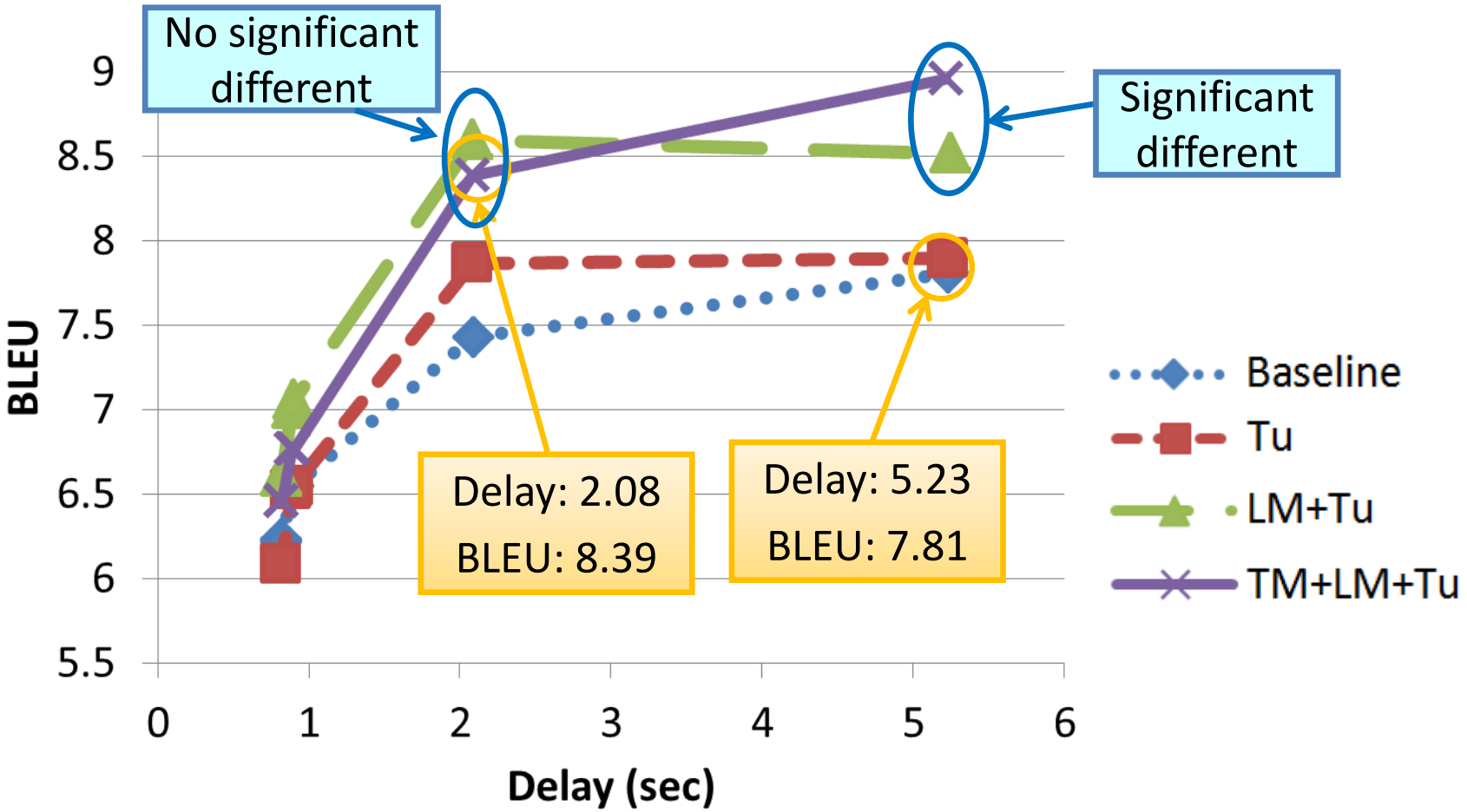
1) Translation accuracy
   - BLEU, RIBES
2) Delay
   - Time from start of input to completion of translation

(100% accurate ASR and do not consider speech synthesis)

Result: learning of the MT system (BLEU)

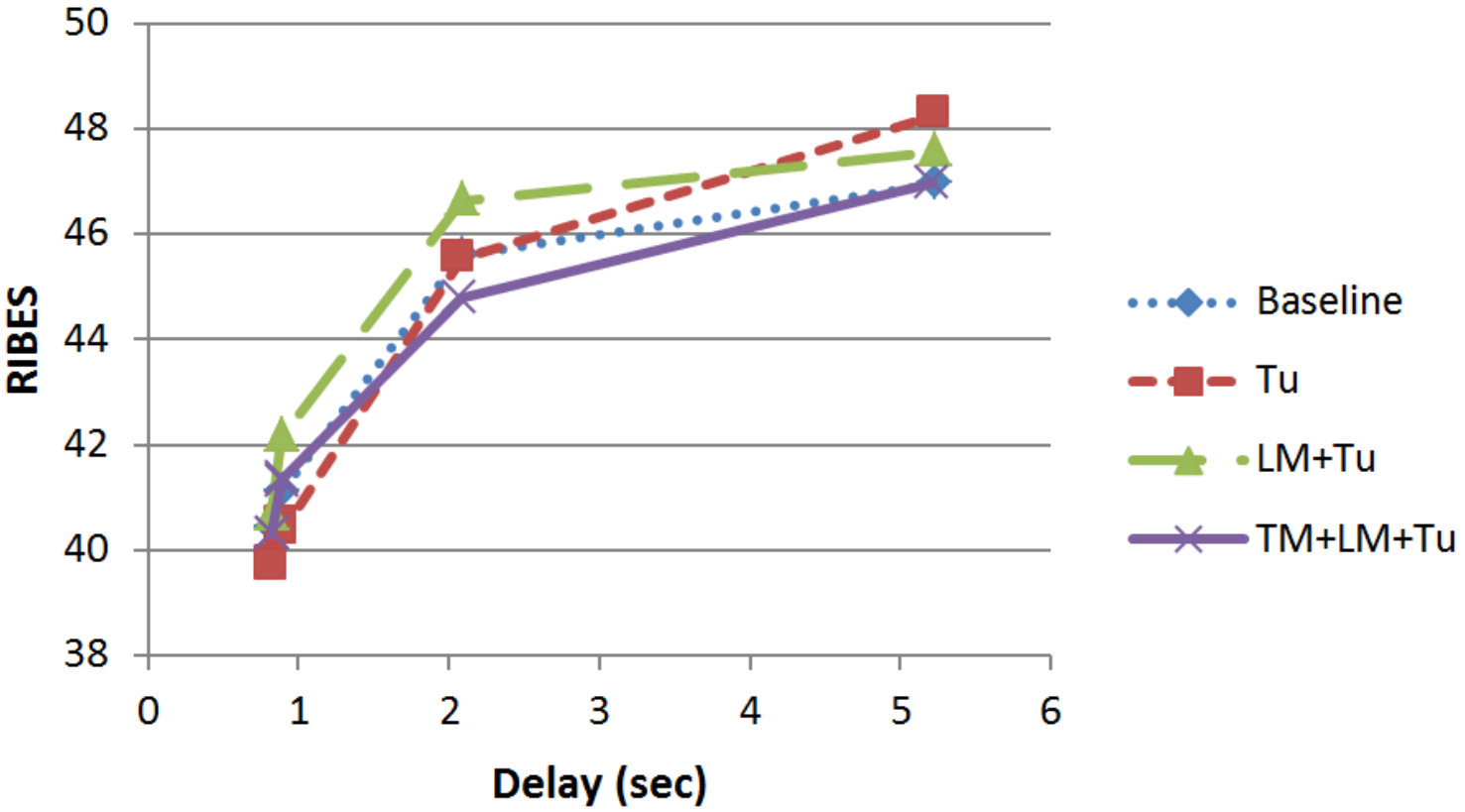


Result: learning of the MT system (BLEU)



More similar to human simultaneous interpreters

Result: learning of the MT system (RIBES)



Proposed system does not show improvement from the view of RIBES, because tuning is optimized for BLEU.

Example of translation results

	Sentence
Src	If you look at in the context of the history you can see what this is doing
Ref	過去から/流れを見てみますと/災害は/このように/増えています *from the past / look at the context and / disasters are / like this increasing*
Base-line	見てみると/歴史の中で/見ることができます/これがやっていること *looking at / in the history / you can see / what this is doing*
Pro-posed	では/歴史の中で/見ることができます/これがやっていること *ok / in the history / you can see / what this is doing*

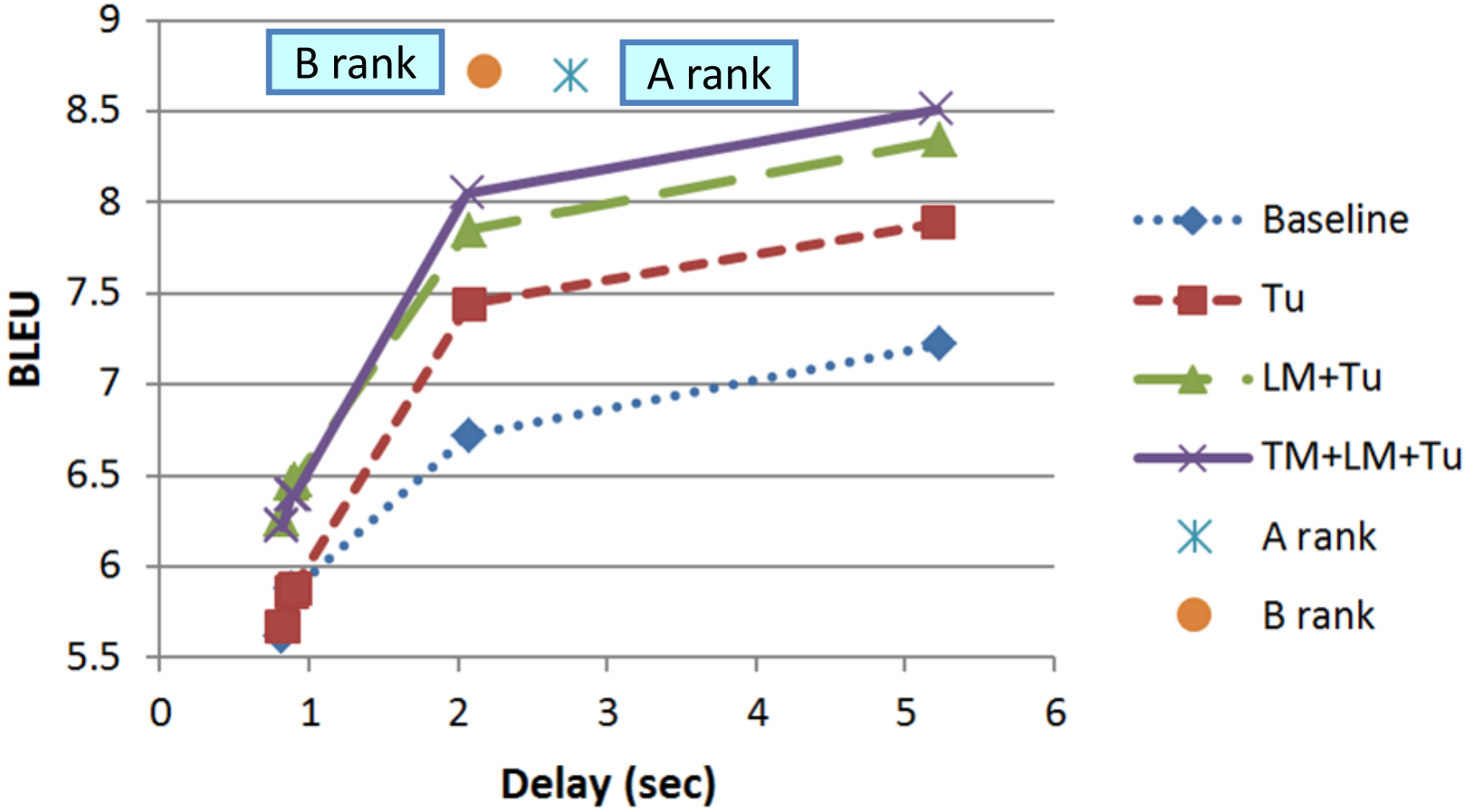
- Choose shorter phrase to reduce the number of the words
  - Tuning has adjusted the parameters to prefer shorter
- Start a sentence with the word “で (*and*)” (over 25% sentence)
  - Avoid long empty pauses while the interpreter is waiting

Setup: comparing the system with human simultaneous interpreters

We compare our proposed system with the human simultaneous interpreters

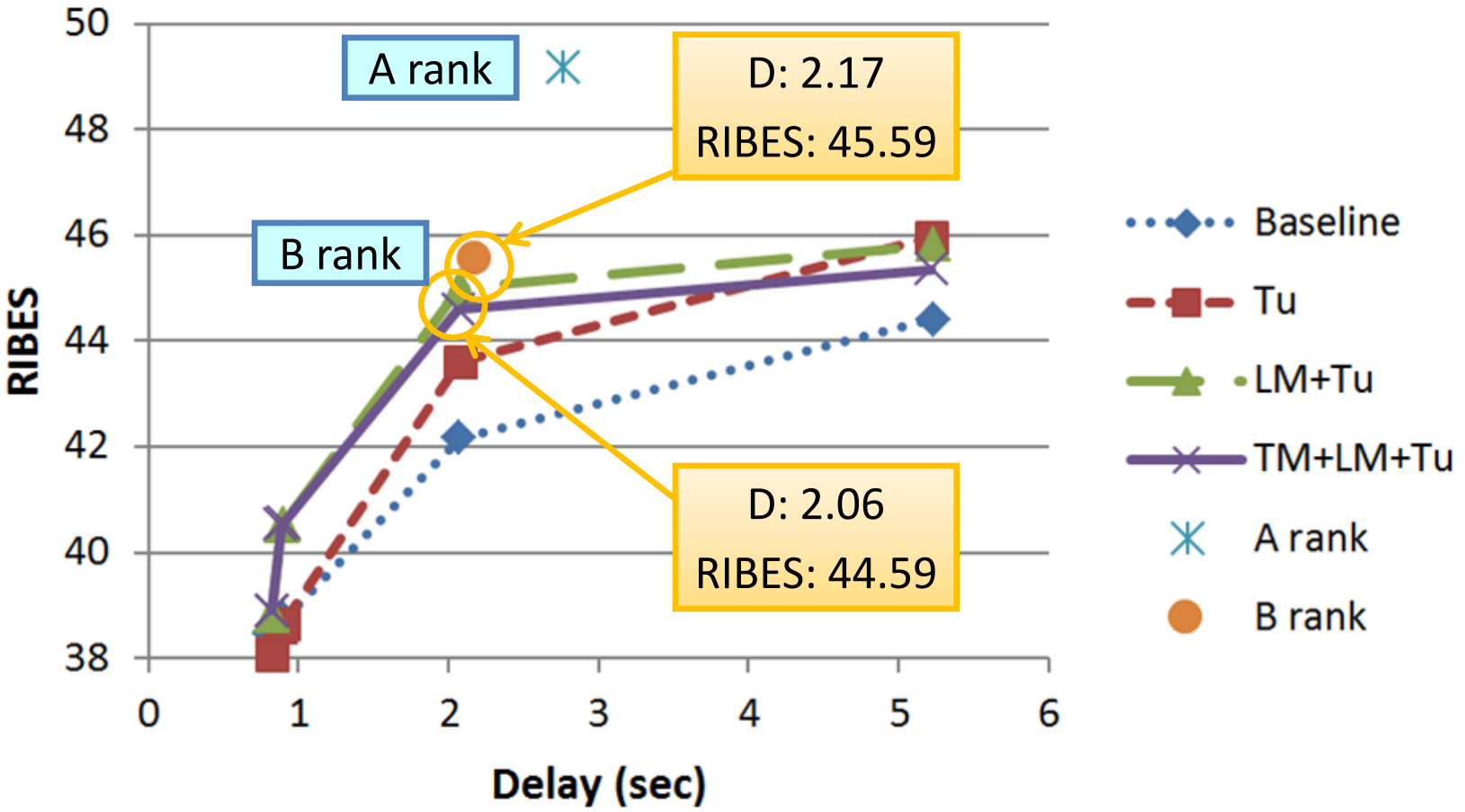
- Compare with the human simultaneous interpreters
  - A rank (4 years)
  - B rank (1 year)
- We use ASR results as input to the translation system
  - WER is 19.36%

Result: comparing the system with human simultaneous interpreters (BLEU)



The system achieves result slightly lower than human simultaneous interpreters from the view of BLEU.

Result: comparing the system with human simultaneous interpreters (RIBES)



From the view of RIBES, the system and B rank (1 year) interpreter achieve similar result.

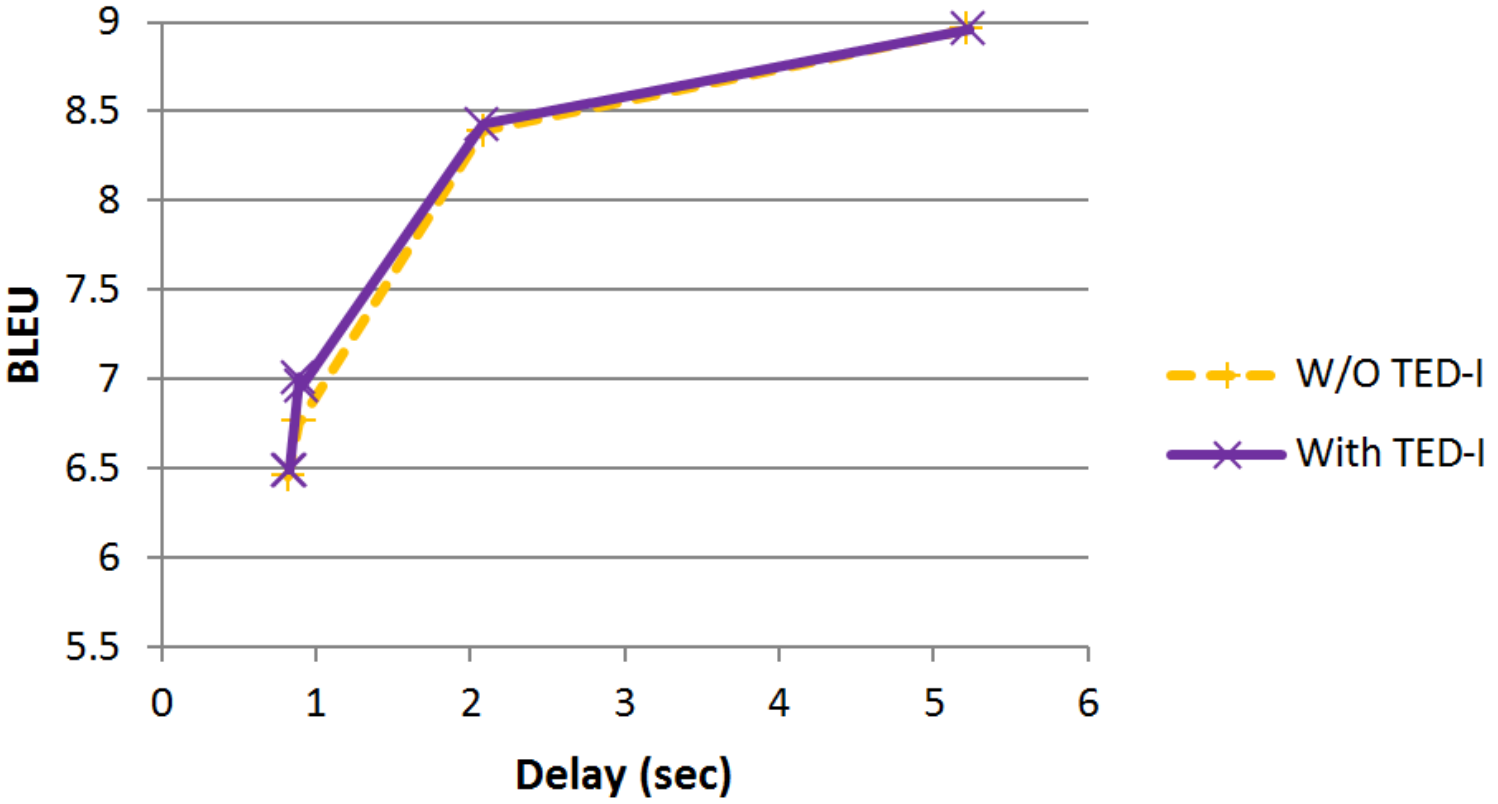
Conclusion

- Purpose
  - Examine the potential of simultaneous interpretation data
- Proposed
  - Use simultaneous interpretation data for learning of MT
- Result
  - Output is more similar to simultaneous interpreter
- Future works
  - Subjective evaluation

Thank you!
Questions?

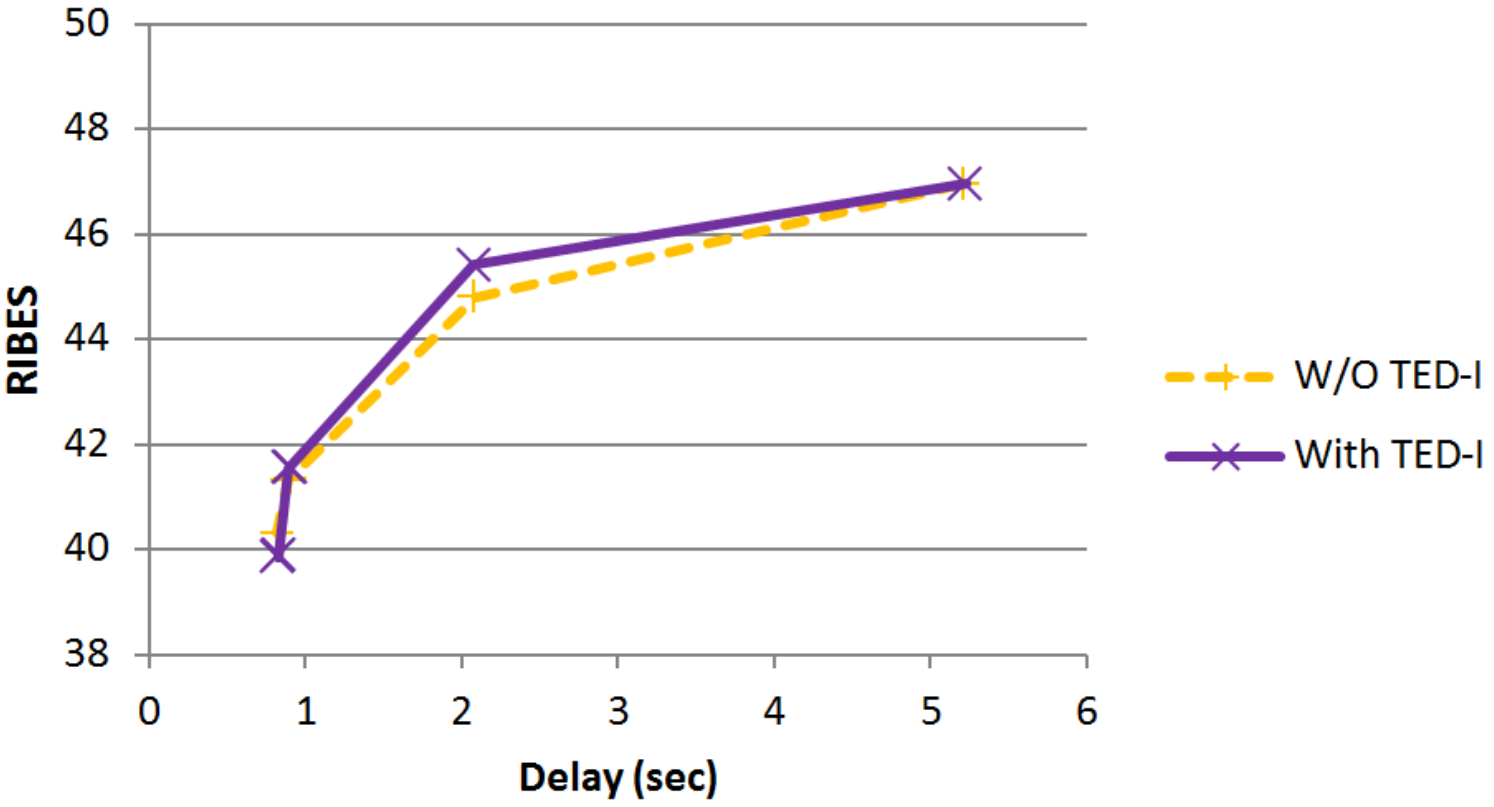
Appendix

Result: learning of translation timing (BLEU)



There is no difference to use the simultaneous interpretation data for learning right probability.

Result: learning of translation timing (RIBES)



There is no difference to use the simultaneous interpretation data for learning right probability.

Evaluation method

- Delay

$$D = U + T$$

U: Waiting time before we can start translating

T: Time required for MT decoding

Significance Testing

- Using the bootstrap resampling
  - Using 10000 samples
  - α =0.05

Sample data of human simultaneous interpreters (Japanese voice)

The S rank interpreter
(15 years)

The A rank interpreter
(4 years)

The B rank interpreter
(1year)