

A machine translation system
combining
rule-based machine translation
and
statistical post-editing

Terumasa EHARA

Yamanashi Eiwa College

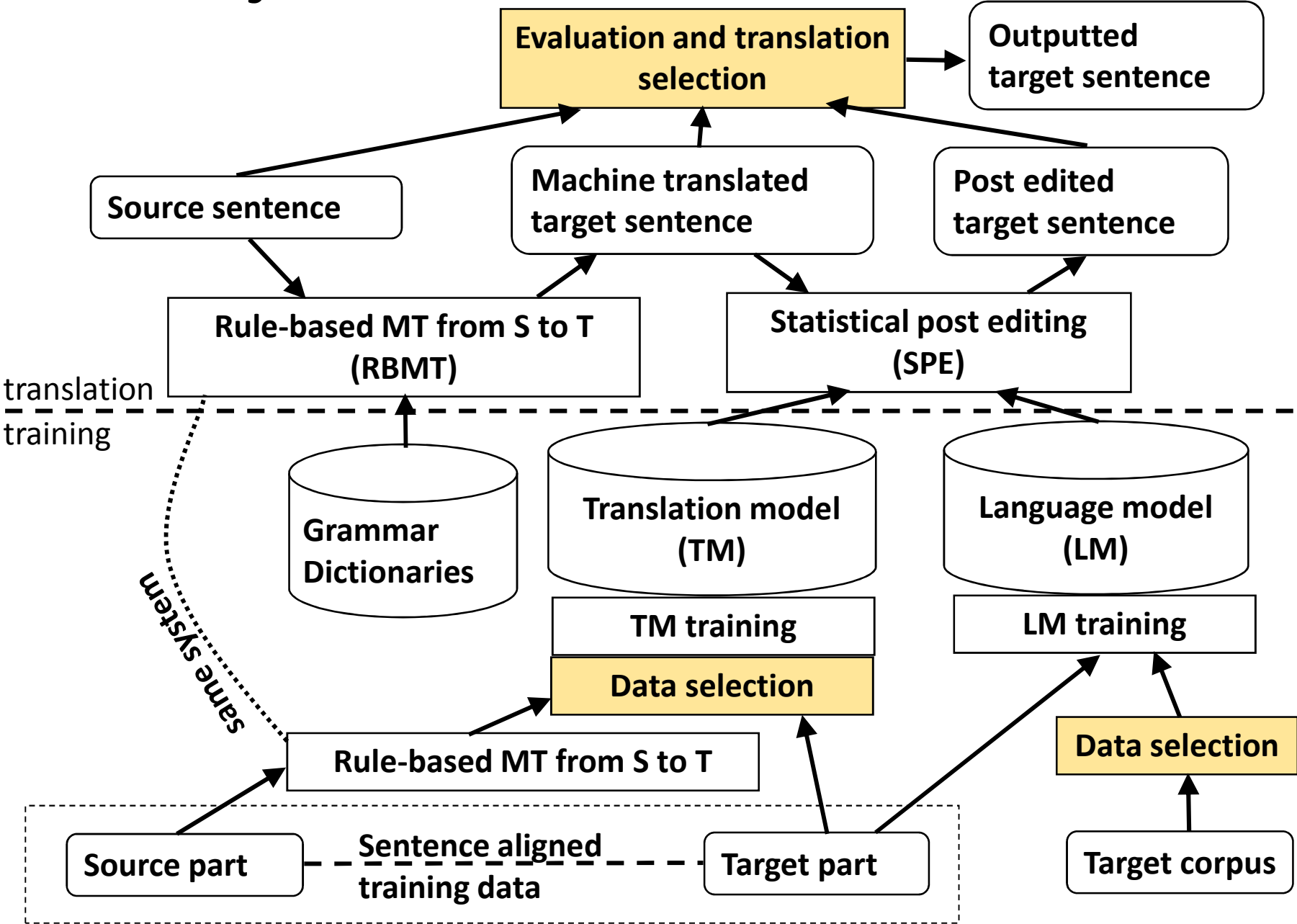


2014. 10. 4

Introduction

- Hybrid system combining rule-based MT (RBTM) and statistical post editing (SPE) may make MT more accurate.
- Several improvements are implemented.

Basic system architecture

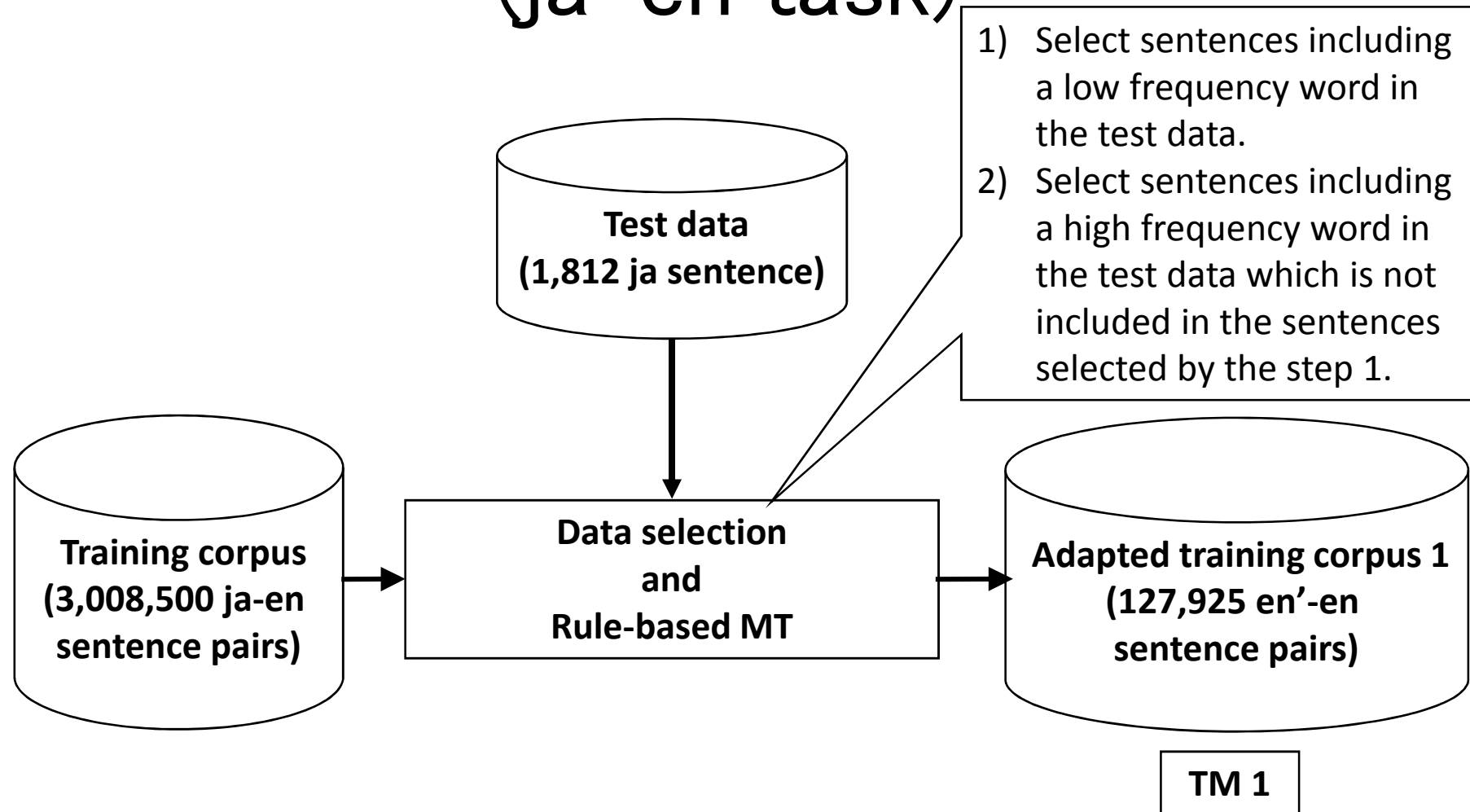


Tools

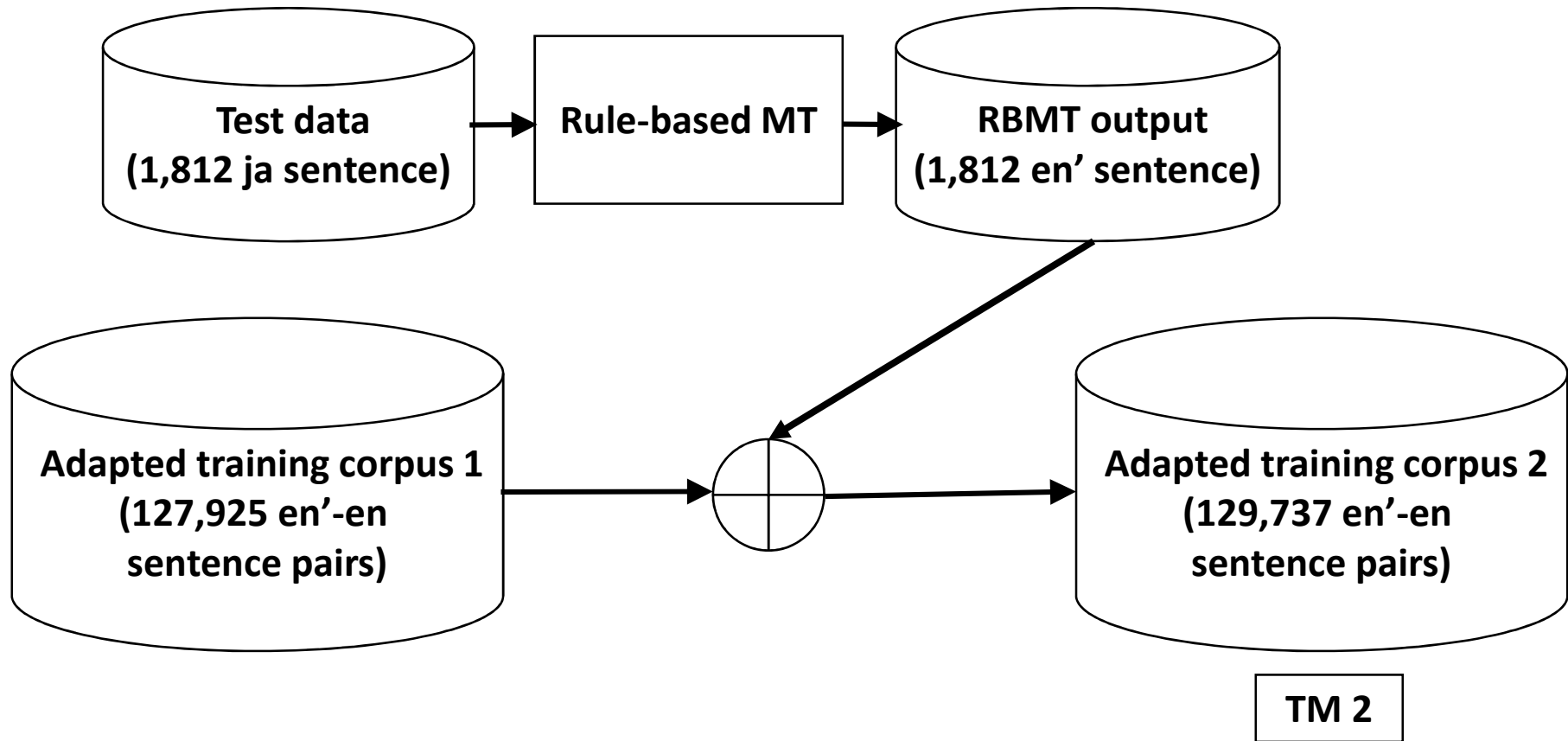
- ▶ RBMT part: commercial based systems
- ▶ SPE part: phrase-based Moses
(Rev. 4343, distortion limit = 1)
- ▶ LM training tool: SRILM (ver.1.5.5)
- ▶ TM training tool: GIZA++ (v1.0.3)

Translation model adaptation (1)

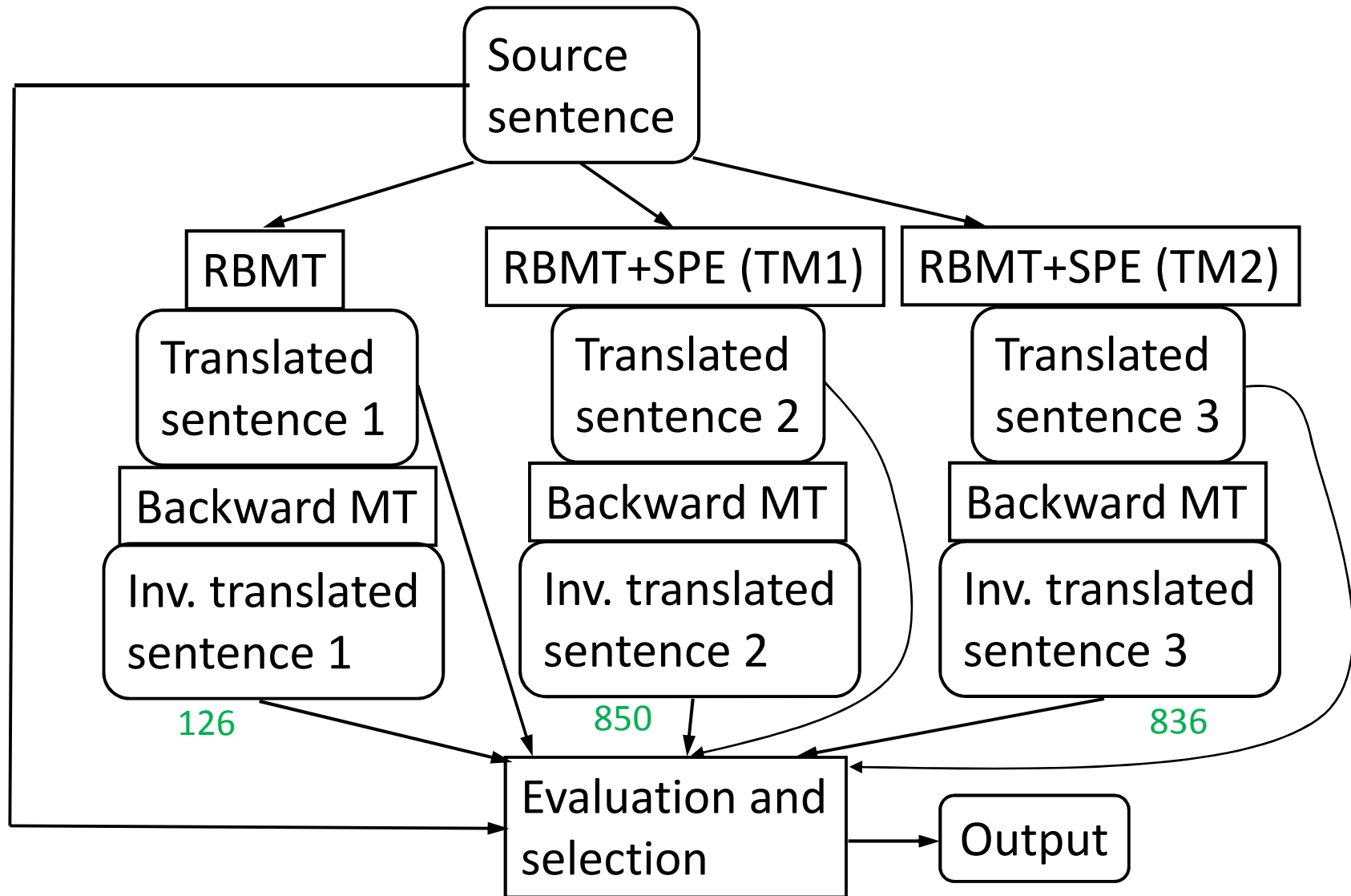
(ja-en task)



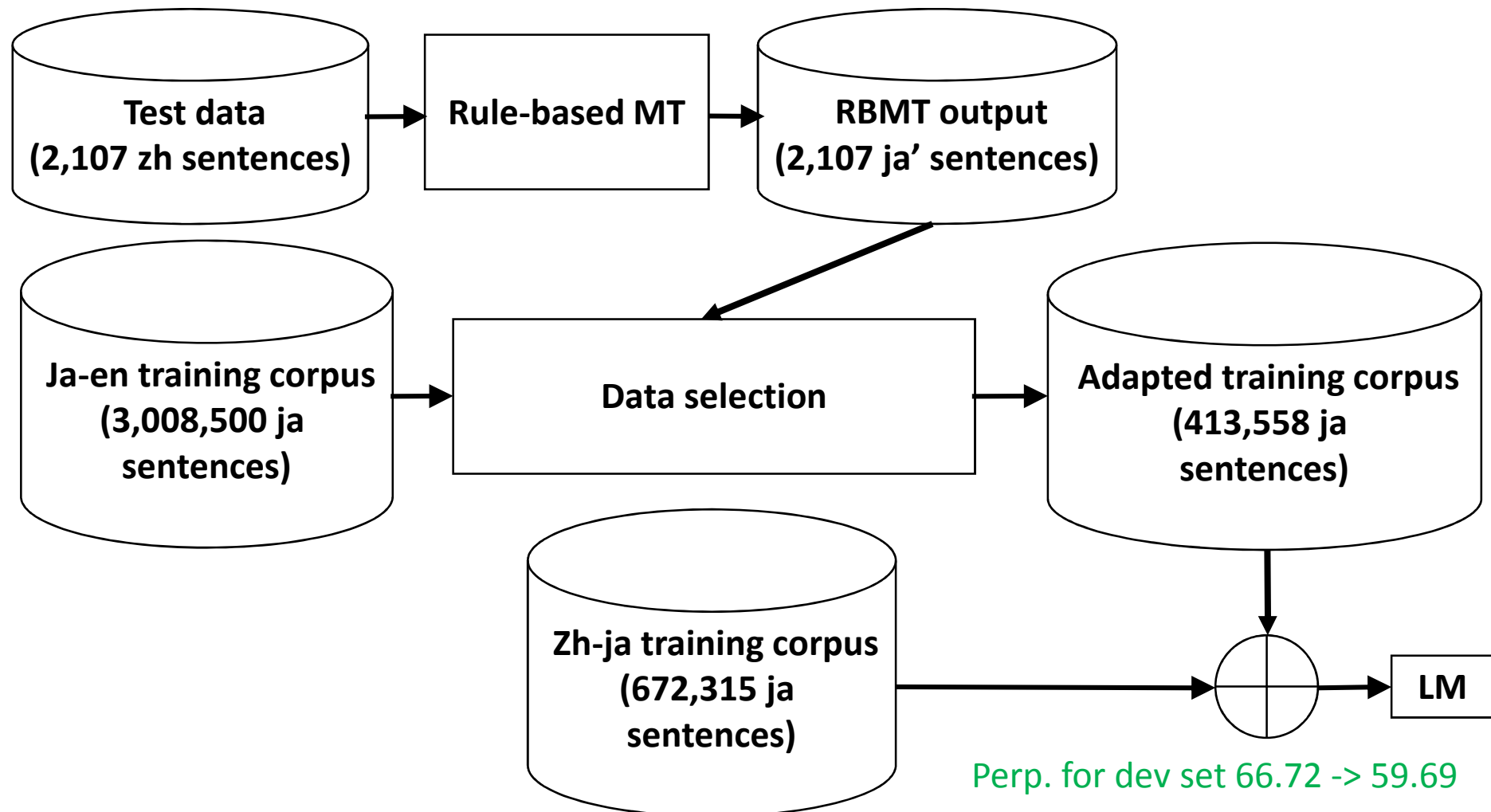
Translation model adaptation (2) (ja-en task)



Translation selection (ja-en task)



Language model adaptation (zh-ja task)



Evaluation results

Task	BLEU	RIBES	HUMAN
ja-en	19.86 (9/27)	0.7067 (5/27)	22.50 (7/16)
zh-ja	33.53 (12/19)	0.8114 (6/19)	15.00 (5/11)

Numbers in parentheses :

(EIWA' s rank / number of all evaluated systems
up to September 14th)

Example of EIWA' s win case

EIWA	この発見は角野の推測に新たな根拠を与えていると考えられる。
base line	この発見により与えられると考えられる対角野の推測のために新しい。
reference	このことは角野の推測に新たな根拠を与えるものと考えられる。
source	可以认为这个发现对角野的推测给予新的根据。

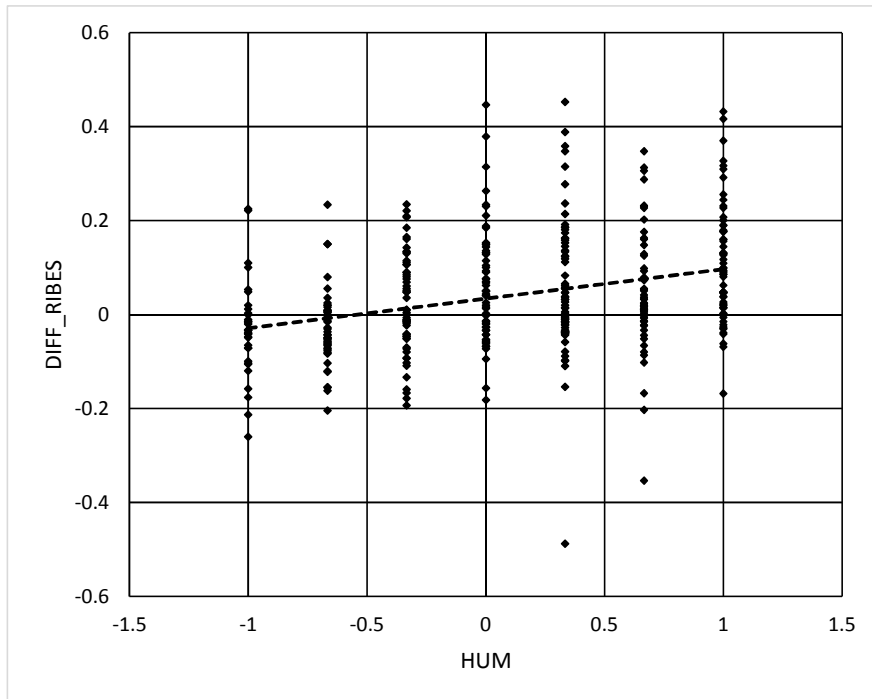
Example of EIWA' s loss case (1)

EIWA	灯油漏洩を防止処置を扱った研究は非常に多く、化学処理方法の検討中の段階であった。
base line	石油流出の防止処理対策についての研究は非常に多く、化学の防犯処理方法は研究段階にある。
reference	流出油の防除手法についてはかなり研究が進んでいるが、ケミカルの防除手法はまだ調査研究の途中段階である。
source	关于石油泄漏的防范处理措施的研究非常多，通过化学的防范处理方法正处在研究阶段。

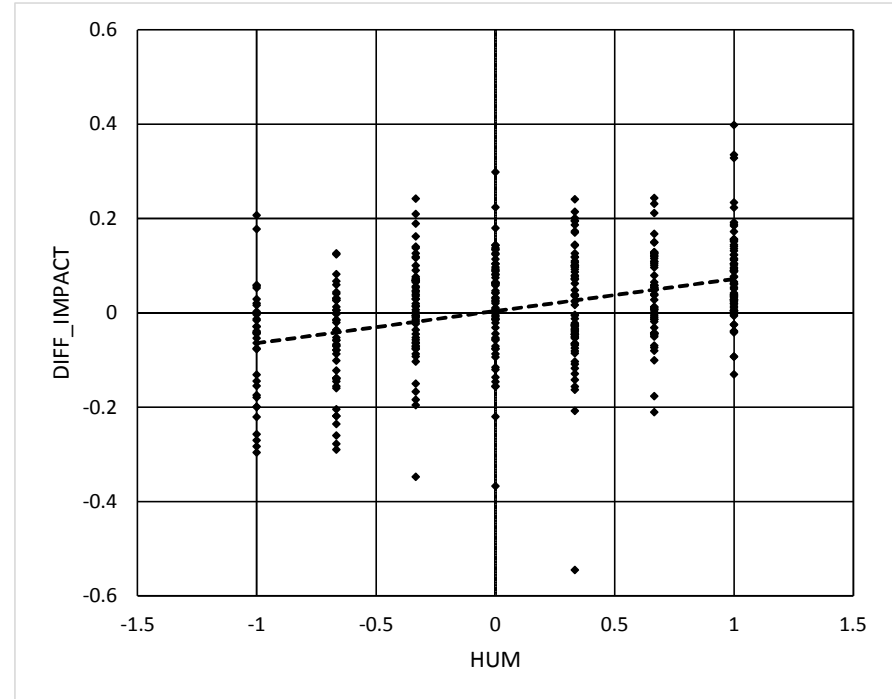
Example of EIWA' s loss case (2)

EIWA	着目し, 地盤材料として利用する人工石炭灰粒子として, すべての粒子の特性に基づいて, 粒子の形状, 物理的特性と単粒子の粉碎特性について検討を行った。
base line	個々の粒子の特性に基づいて, 地盤材料としての利用の人工石炭灰粒子に着目し, 粒子の形状, 物理特性および単個の粒子の破碎特性について検討した。
reference	造粒石炭灰を地盤材料として利用することに着目し, 粒子個々の特性に基づいて, 粒子の形状や物理的特性および単粒子破碎特性について検討した。
source	着眼于作为地基材料予以利用的人造煤灰颗粒, 基于每个粒子的特性, 对粒子的形状、物理特性以及单颗粒子的破碎特性进行了讨论。

Correlation between HUM and DIFF_RIBES and DIFF_IMPACT



$$\rho = 0.3618$$



$$\rho = 0.3686$$

HUM: average value of scores of three human annotators

DIFF_RIBES: difference of RIBES scores of EIWA's output and baseline output

DIFF_IMPACT: difference of IMPACT scores of EIWA's output and baseline output

HUM < 0 and DIFF_RIBES > 0 case

EIWA	水稻のカドミウムの高吸収特性を示す品種「暗号陽23号」を用いた。
base line	カドミウム高吸収特性の品種「密陽23号」を示すイネを採用した。
reference	イネはカドミウム高吸収特性を示す品種「密陽23号」を用いた。
source	稻子采用了显示镉高吸收特性的品种“密阳23号”。
	(a) HUM=-1, DIFF_RIBES=0.222

HUM > 0 and DIFF_RIBES < 0 case

EIWA	表3に人工株式市場の各エージェントの予測の木で用いる記号を構成することを示した。
base line	人工株式市場を構成する各エージェントの予測の木に用いる記号を表3に示す。
reference	人工株式市場を構成する各エージェントの予測の木に用いる記号を表3に示す。
source	表3显示构成人工股份市场的各代理商的预测树中使用的记号。
	(c) HUM=0.3333, DIFF_RIBES=-0.4873

Conclusion and future works

- System architecture, experimental setting and evaluation results of EIWA are described.
- We are in the middle position in human and automatic evaluation.
- To improve parsing accuracy in the RBMT part.
- To combine rule-based technique and statistical technique more tightly beyond RBMT plus SPE method.