Statistical Machine Translation with Rule based Machine Translation

○Jin'ichi Murakami, Masato Tokuhisa
(Tottori University, Japan)

Phrase Based Statistical Machine Translation

Problem: *N*-gram = Local language information

Proposed Method:

Two-stage machine translation

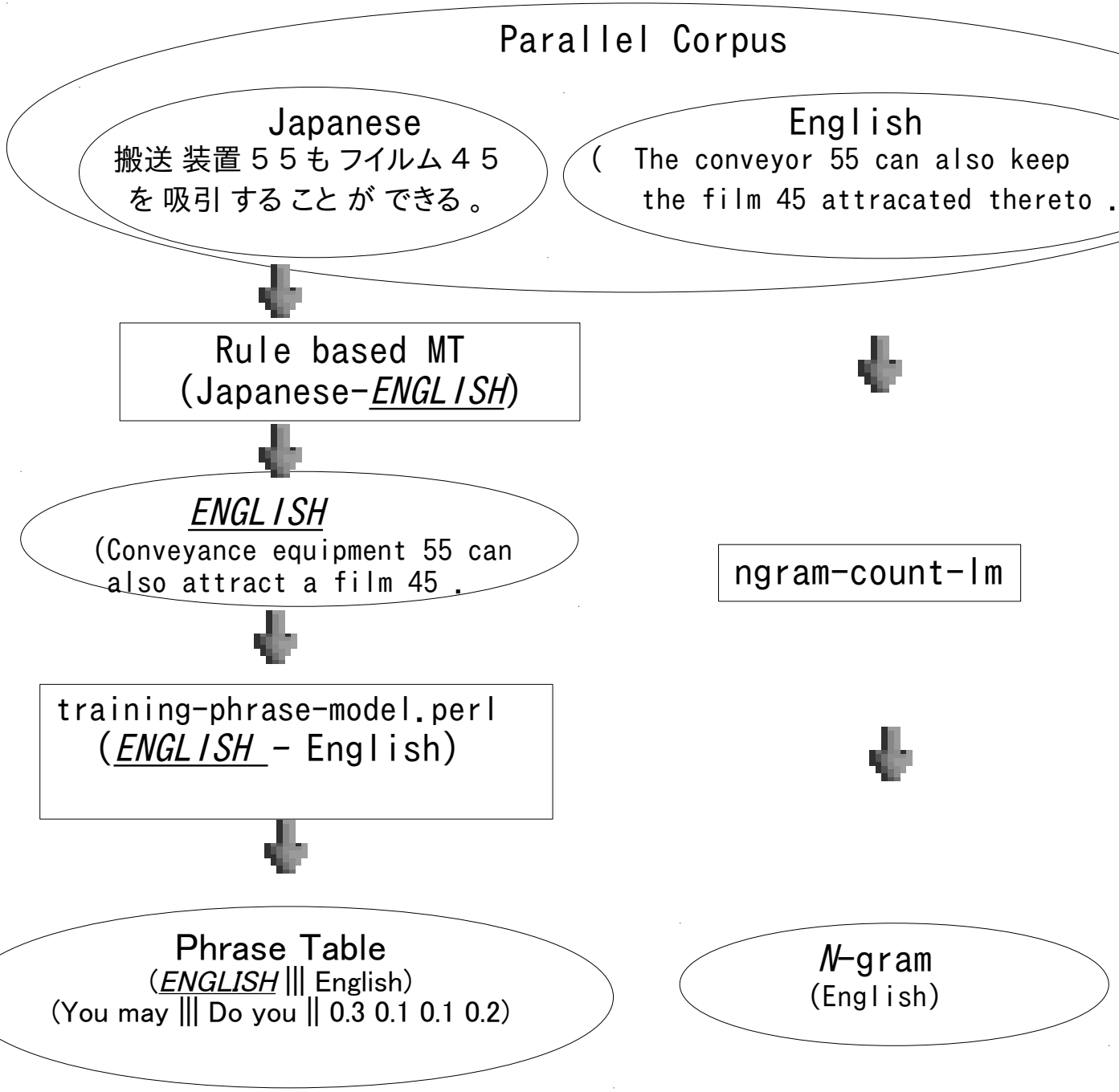
First stage : Rule-based MT

a) Few unknown words
b) Include grammatical information
c) Low levels of fluency and naturalness

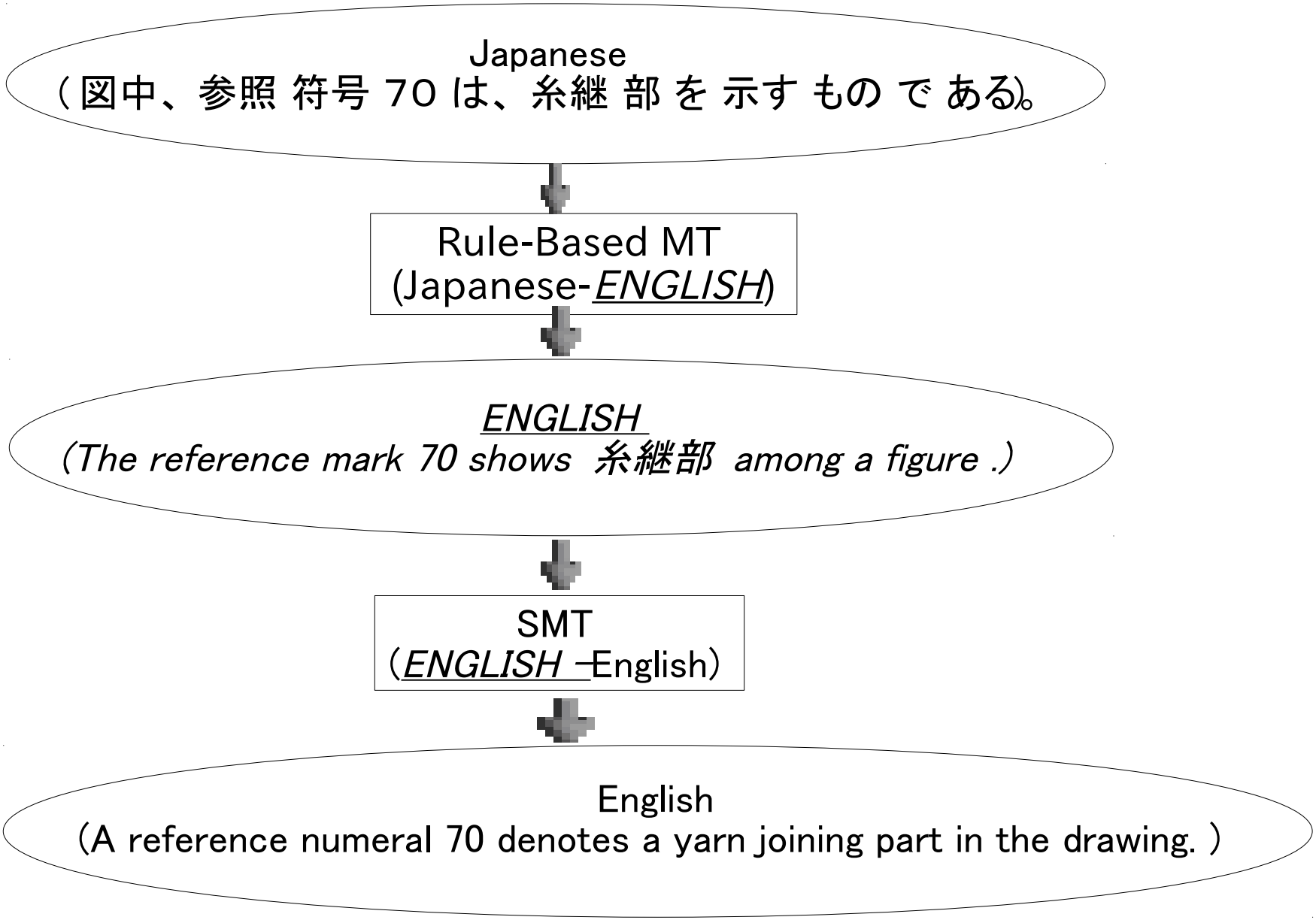
Second stage: Normal SMT

a) Revise the outputs of the first stage
b) Improve the naturalness and fluency

Training


Parallel Corpus
Japanese
搬送 装置 ５５ も フイルム ４５
を 吸引 する こと が できる 。
English
(The conveyor 55 can also keep
the film 45 attracated thereto .)
Rule based MT
(Japanese-ENGLISH)
ENGLISH
(Conveyance equipment 55 can
also attract a film 45 .
training-phrase-model.perl
(ENGLISH - English)
Phrase Table
(ENGLISH ||| English)
(You may ||| Do you || 0.3 0.1 0.1 0.2)
ngram-count-lm
N-gram
(English)

Decoding


Japanese
（図中、参照 符号 70 は、糸継 部 を 示す もの で ある。
Rule-Based MT
(Japanese-*ENGLISH*)
ENGLISH
(The reference mark 70 shows 糸継部 among a figure .)
SMT
(*ENGLISH* –English)
English
(A reference numeral 70 denotes a yarn joining part in the drawing.)

Experiments

Training Data :
NTCIR- 8 data (3,186,284 sentences)

Rule based MT:
A state-of-the-art trial Rule Based MT

SMT:
Moses (Phrase base)

Language Model : SRILM

5-gram

" cdiscount 0" (No smoothing)

SMT Decoder: Moses

(No parameter optimize and No reorder model)

moses.ini

ttable-limit 80 0

weight-d 0.1

weight-l 1.0

weight-t 0.5 0.0 0.5 0.1 0.0

weight-w -1

distortion-limit (-1 or 6)

Proposed Methods : distortion-limit 6

Standard SMT : distortion-limit -1

Output Sentences (JE)

Input:
また、サーモメータエンコーダ11は、入力信号IN0が「16」(10進数)の場合に例外処理を行い、剰余コードC0に「4」を出力し、マルチプルコードC1multipleに「12」を出力する。

Proposed (RBMT+SMT):
Further , the thermometer film encoder 1 1 conducts an exceptional process , when the input signal IN0 is " 1 6 " (decimal number) , it outputs " 4 " to the excessive code C0 and outputs " 1 2 " to multiple code C1 .

SMT(Moses):
he remainder is " 0 " , " 1 6 " and " 4 " in decimal number) , the output of the thermometer 1 2 multiple exception code (C1 code) encoder 1 0 1 is subjected to the multiple outputs the input signal IN . ¥¥

RBMT:
Moreover, the thermometer encoder 11 performs exception handling, when input signal IN0 is "16" (decimal number), it outputs "4" to the surplus code C0, and outputs "12" to multiple code C1multiple.

Reference:
Also , when input signal IN0 is 1 6 (decimal number) , thermometer encoder 1 1 performs exceptional treatment , outputs 4 to remainder code C0 , and outputs 1 2 to multiple code C1 multiple .

Output Sentences (JE)

Input:
図4に非磁性層を流れる電流に対する電気抵抗を縦軸にとってグラフに示している。

Proposed (RBMT+SMT):
The electric resistance to the current flowing in FIG . 4 via a non-magnetic layer is shown in the graph on the vertical axis .

SMT (Moses):
FIG . 4 is a graph showing the electric current flowing through the resistance nonmagnetic layer vertical axis .

RBMT:
The electrical resistance to the current which flows into Fig. 4 through a nonmagnetic layer is shown in the graph for the vertical axis.

Reference:
FIG . 4 is a graph showing an electric resistance to an electric current flowing in the nonmagnetic layer with the axis of ordinates representing the electric resistance .

Output Sentences (EJ)

Input:
When the sheathed wire is press-fitted into the slot portion 22 of each press-contacting piece portion 23 of the terminal 12 from the upper side, the insulating sheath of the sheathed wire is cut by the blade 21 of the slot portion 22, so that the internal conductor of the sheathed wire contacts the press-contacting piece portion 23.

Proposed (RBMT+SMT):
被覆 電線 を、上側 から 端子 １２ の 各 圧 接 片 ２３ の スロット ２２ に 圧入 する 際、被覆 電線 の 絶縁 被覆 は スロット 部 ２２ の ブレード ２１ で 切断 さ れ、被覆 電線 の 内部 導 電 体 は、圧 接 片 ２３ に 当 接 する。

SMT(Moses) :
また、圧 接 端子 １ の 各 圧 接 刃 ２１ の 開口 部 ２２ に 接触 し て 被覆 電線 の 被覆 電線 の 被覆 部 を 切断 し、絶縁 被覆 電線 を 圧入 する 溝部 ２２ の 片 ２３２ の 上面 から 内部 導体 片 ２３ で ある。

RBMT :
鞘 に 納め られ た ワイヤー が 上部 の 側 から の ターミナル １２ の 各 圧迫 と 接触 する 部分 部分 ２３ の スロット 部分 ２２ に 圧迫 入れ られる 場合、鞘 に 納め られ た ワイヤー の 絶縁 する さや は スロット 部分 ２２ の 葉 ２１ によって カット さ れる。その 結果、鞘 に 納め られ た ワイヤー の 内部 伝導 体 は 圧迫 と 接触 する 部分 部分 ２３ と 接触 する。

Reference:
そして、端子 １２ の 上方 から 圧 接 片 ２３ の スロット ２２ に 被覆 電線 が 圧入 さ れる と、被覆 電線 の 絶縁 被覆 が スロット ２２ の 刃 部 ２１ により 切 裂 さ れ て 被覆 電線 の 内部 の 導体 が 圧 接 片 ２３ に 接触 する。

Output Sentences (EJ)

Input:
As shown in FIG. 4, the valve overlap amount decreases as the valve working angle of the intake valve 30 decreases.

Proposed (RBMT+SMT):
図４に示すように、バルブオーバラップ量は、吸気弁３０の弁の角度が減少している。

SMT (Moses):
図４に示すように、バルブオーバーラップ量が減少し、吸気弁３の作用角が減少している。

RBMT:
図４に示されるように、バルブ・オーバーラップ量は、吸い込み弁３０減少のバルブを動かす角度につれて減少する。

Reference:
同図４に示すように、吸気バルブ３０のバルブ作動角が小さいときほど、バルブオーバーラップ量は小さい

Results of Experiments (Automatic Evaluation)

(#order / #system)

	Task	BLEU	NIST	RIBES
Proposed (RBMT+SMT)	JE	0.1996 (28/36)	6.1112 (32/36)	0.6932 (9/36)
RBMT	JE	0.2090 (26/36)	6.2831 (30/36)	0.6972 (8/36)
SMT(Moses)	JE	0.1436 (36/36)	4.926 (36/36)	0.6607 (20/36)
Proposed (RBMT+SMT)	EJ	0.2775 (21/32)	7.3284 (21/32)	0.7479 (4/32)
RBMT	EJ	0.2475 (25/32)	7.1413 (24/32)	0.6782 (23/32)
SMT(Moses)	EJ	0.0831 (32/32)	3.7711 (32/32)	0.5902 (32/32)

- (Mysterious Evaluation Results – Moses was too low)
- (RBMT+SMT > RBMT > SMT) ?
- #order: Low

Results of Experiments (Human Evaluation)

(#order / #system)

	Task	Adequacy	Acceptability	(tie)
Proposed (RBMT+SMT)	JE	2.73 (7/19)	0.4604 (8/14)	0.3312 (9/14)
Proposed (RBMT+SMT)	EJ	2.6 (9/17)	0.4318 (8/11)	0.2992 (5/11)

- Good results
- Order: middle
- Large difference (Human vs Automatic)

Results of Experiments with Parameter Tuning (JE)

Task	Tuning	BLEU	NIST	METEOR	TER	WER	RIBES	IMPACT
Proposed	○	**0.3598**	**8.1769**	**0.6676**	**0.5387**	**0.6436**	**0.7412**	**0.5654**
Proposed	×	0.2697	7.1982	0.6049	0.5666	0.6566	0.7240	0.5197
RBMT	×	0.2761	6.8759	0.6099	0.6172	0.7048	0.7114	0.5064
Moses	○	0.2886	7.1503	0.6567	0.6684	0.8307	0.6334	0.4527
Moses	×	0.2120	6.9635	0.5741	0.6431	0.7852	0.6727	0.4078

Good Results (Best Performance?)
Proposed > Moses > RBMT

Results of Experiments with Parameter Tuning (EJ)

Task	Tuning	BLEU	NIST	TER	WER	RIBES	IMPACT
Proposed	○	**0.3911**	**8.3941**	**0.4991**	**0.6184**	**0.6709**	**0.5753**
Proposed	×	0.3076	7.6219	0.5441	0.6492	0.6562	0.5326
RBMT	×	0.1998	5.4690	0.7274	0.8075	0.5632	0.4393
Moses	○	0.2408	6.4319	0.5441	0.6492	0.6563	0.4743
Moses	×	0.2531	7.1181	0.5968	0.7377	0.5532	0.4394

Good Results (Best Performance?)
Proposed > Moses > RBMT

Discussion:

Human evaluation:

Proposed > SMT

Proposed ≅ RBMT

(Proposed ≦ RBMT ?)

Human Evaluation vs Automatic Evaluation

No match

(BLUE,NIST,METEROR,TER,WER,RIBES,IMPACT)

Conclusion

Our System:

Two-stage machine translation system.

First stage : Rule-based MT

Second stage : SMT

Object:

Less ungrammatical sentences. (Compared to SMT)

Results:

Effective (Compared to SMT)

Future:

Automatic Evaluation vs Human Evaluation

RBMT ≦ Proposed ?