Statistical Machine Translation adding Pattern-based Machine Translation in Chinese-English Translation

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

Phrase Based Statistical Machine Translation

Problem

a) Small Database
   Unknown word

b) *N*-gram
   Local language information

Proposed Method:
Two-stage machine translation

First stage : Rule-based MT

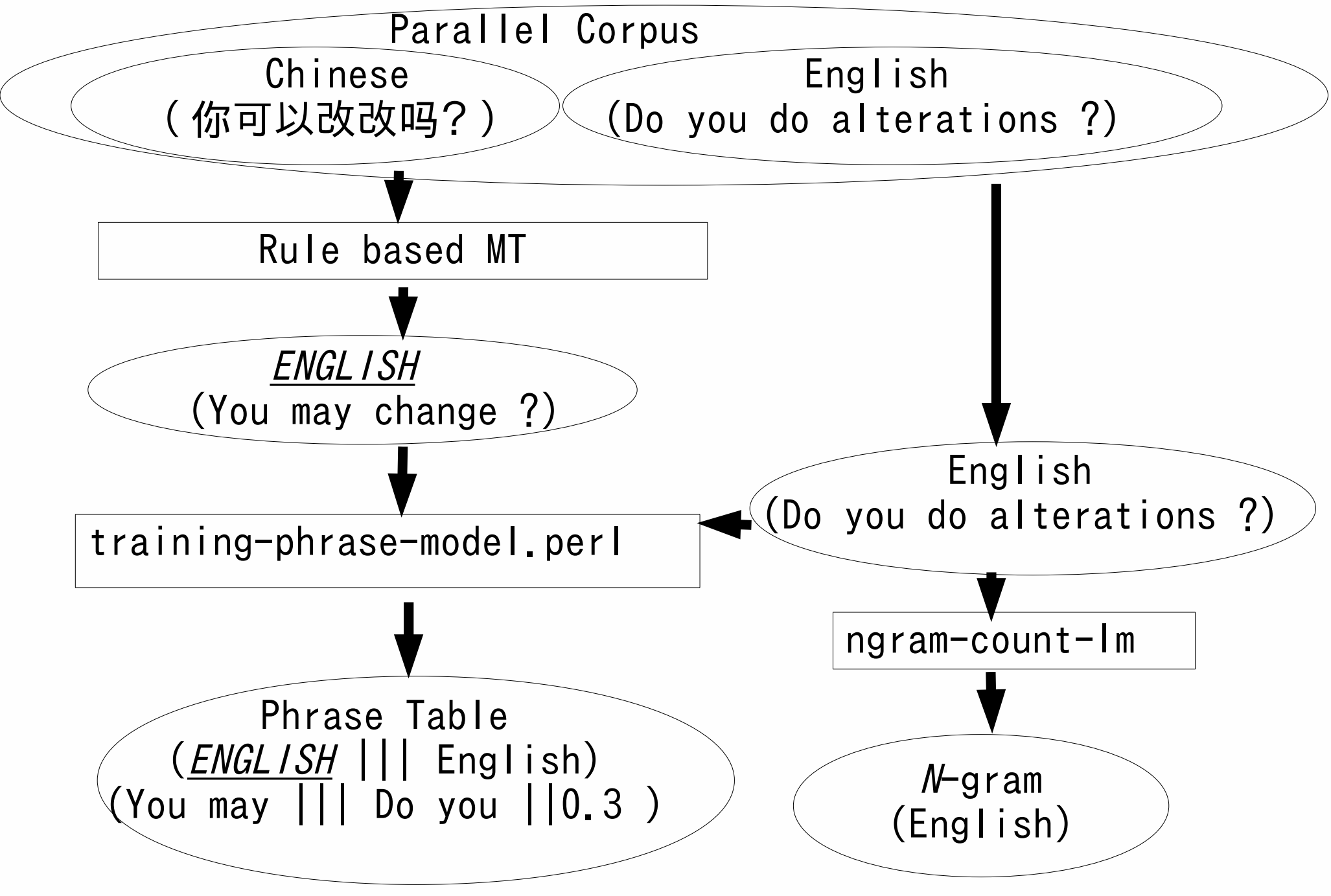
a) Few unknown words
b) Include grammatical information
c) Low levels of fluency and naturalness

Second stage: Normal SMT

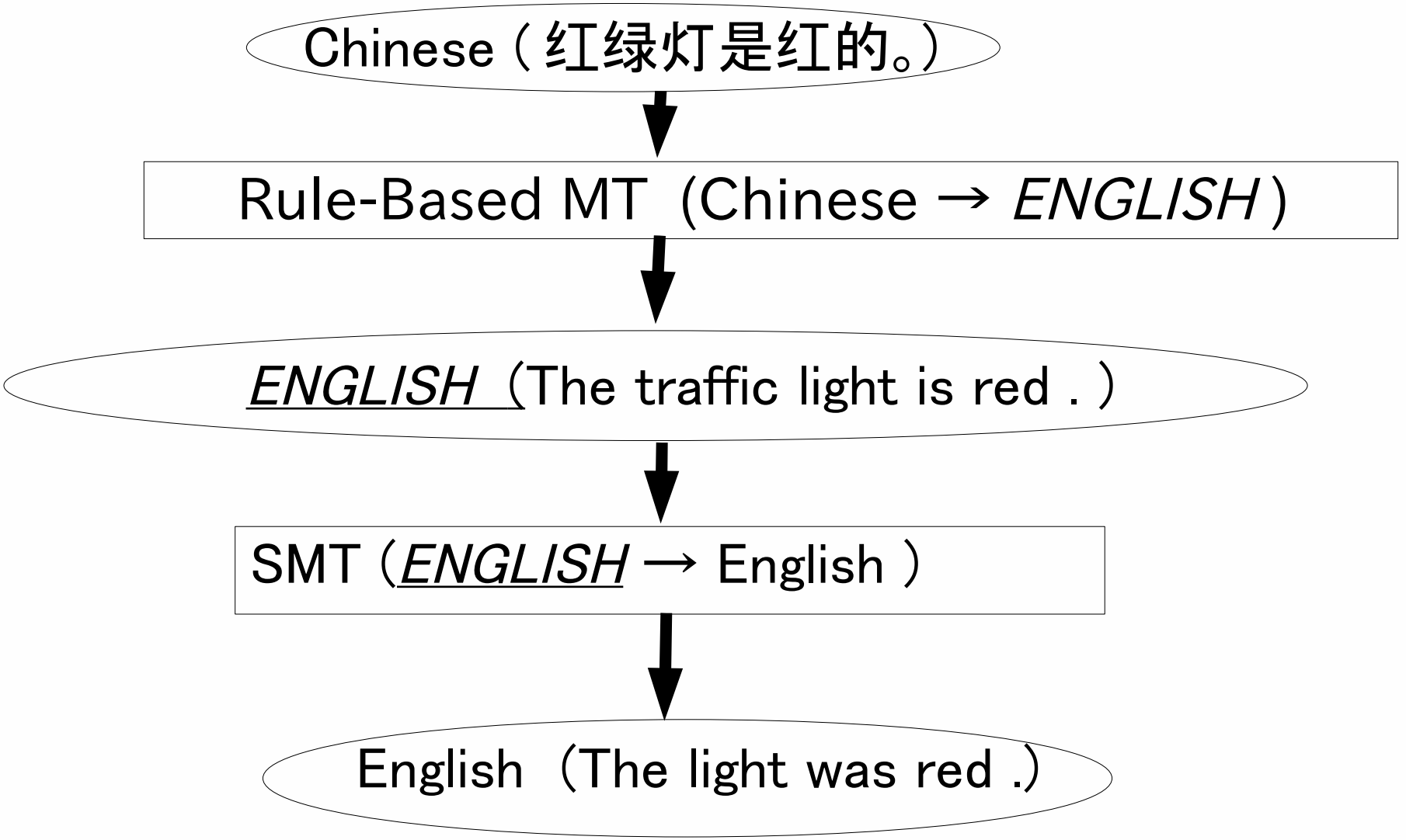
a) Revise the outputs of the first stage
b) Improve the naturalness and fluency

Chinese-English: SYSTRAN + Moses

Training


Parallel Corpus
Chinese
(你可以改改吗?)
English
(Do you do alterations ?)
Rule based MT
ENGLISH
(You may change ?)
English
(Do you do alterations ?)
training-phrase-model.perl
ngram-count-lm
Phrase Table
(ENGLISH ||| English)
(You may ||| Do you ||0.3)
N-gram
(English)

Decoding


Chinese（红绿灯是红的。）
Rule-Based MT （Chinese → *ENGLISH*）
ENGLISH（The traffic light is red .）
SMT（*ENGLISH* → English）
English（The light was red .）

Chinese : 不用 担心 那个 。
我 要 买 它 你 不 需要 把 它 包 起来 。
English : No worry about that .
I'll take it and you need not wrap it up .
SYSTARN : Does not need to worry that .
I must buy its you not to need to wrap it .

Chinese : 你 可以 改改 吗 ？
English : Do you do alterations ?
SYSTRAN : You may change ?

Chinese : 红绿灯 是 红 的 。
English : The light was red .
SYSTRAN : The traffic light is red .

Examples of phrase-tables (*ENGLISH* - English BTEC-CE)

Extremely appropriate . ||| It fits very well .|||
1 0.0037774 1 0.000165701
Extremely appropriate ||| It fits very well|||
1 0.00394828 1 0.000167943
Extremely attractive . ||| It is very beautiful .|||
1 0.00468009 0.5 0.000167226
Extremely attractive . ||| Very beautiful .|||
1 0.121764 0.5 0.0529012
Extremely attractive ||| It is very beautiful|||
1 0.00489181 0.5 0.000169488
Extremely attractive ||| Very beautiful|||
1 0.127273 0.5 0.053617
want to go to eat meal . ||| like to have dinner .|||
1 4.70488e-06 0.5 0.00340606
want to go to eat meal . ||| want to go to the restaurant .|||
1 1.02487e-05 0.5 4.7193e-06
want to go to eat meal ||| like to have dinner|||
0.333333 4.91772e-06 0.5 0.00345215
want to go to eat meal ||| want to go to the restaurant|||
1 1.07123e-05 0.5 4.78316e-06
want to go to eat ||| like to have|||
0.0222222 3.18012e-05 1 0.0191019

5-gram Language Model

SRILM " -ukndiscount -interpolate"

Decoder

Moses (No parameter optimize)

moses.ini

ttable-limit 40 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)

Standard SMT : distortion-limit -1
Proposed Methods (2-stage) : distortion-limit 6

Output Sentences (BTEC-CE)

01	Chinese	救命！
	Proposed	Save a life .
	Moses	Help .
	SYSTRAN	Save a life !
02	Chinese	请 把 房间 打扫 干净 。
	Proposed	Please remove clean my room .
	Moses	Please make up clean the room .
	SYSTRAN	Please clean the room cleanly .
03	Chinese	这 是 旅行 支票 购买 存根 。
	Proposed	Is this the purchase stub traveler's checks .
	Moses	Is this the stub purchase traveler's checks .
	SYSTRAN	This is the traller's check purchase checkstub .
04	Chinese	请 开 得 再 慢 一点 。
	Proposed	Please open slow another one .
	Moses	Is this the stub purchase traveler's checks .
	SYSTRAN	Please open again slow one .

Output Sentences (BTEC-CE)

Chinese	我 想 和 史密斯 先生 谈话 。
Proposed	I'd like to converse to Mr . Smith .
Moses	I'd like to talk to Mr . Smith , 谈话 .
SYSTRAN	I want to converse with Mr . Smith .

Chinese	有 套餐 的 菜单 吗 ？
Proposed	Do you have a menu set meal ?
Moses	Do you have set meals menu ?
SYSTRAN	Has the prix fixe the menu ?

Chinese	日本 的 十 几 岁 青少年 很 喜欢 玩 电子 游戏机 。
Proposed	The game a few years old young people like an electronic games mechanical well .
Moses	How many Japanese ten years old 青少年 electronic 游戏机 really like fun .
SYSTRAN	Japan's several year old young people like playing the electronic mechanical games very much .

Results of Experiments

TASK	BTEC_CE									
case+punc	bleu	meteor	f1	prec	recl	wer	per	ter	gtm	nist
Proposed	0.3151	0.6169	0.6569	0.6465	0.6676	0.5590	0.4760	48.0710	0.6478	6.3834
Moses	0.3311	0.6109	0.6610	0.6758	0.6468	0.5377	0.4567	44.8140	0.6423	6.1511
Systran	0.1070	0.4697	0.5619	0.5671	0.5567	0.7017	0.6182	60.0070	0.4863	3.9727
TASK	CT_CE									
case+punc	bleu	meteor	f1	prec	recl	wer	per	ter	gtm	nist
Proposed.CRR	0.2797	0.5971	0.6306	0.6092	0.6536	0.6590	0.5099	61.3850	0.6592	5.5309
Moses.CRR	0.2706	0.5881	0.6189	0.5945	0.6453	0.6712	0.5113	62.4990	0.6533	5.4633
Systran.CRR	0.0642	0.3953	0.4928	0.5051	0.4811	0.8046	0.6823	74.9560	0.4312	3.2979
Proposed.ASR.1	0.2482	0.5489	0.5910	0.5773	0.6053	0.6943	0.5456	64.8360	0.6136	5.0705
Moses.ASR.1	0.2650	0.5610	0.6000	0.5876	0.6128	0.6647	0.5220	62.0140	0.6307	5.2804
Systran.ASR.1	0.0602	0.3654	0.4644	0.4822	0.4479	0.8148	0.7018	76.1960	0.4009	2.9995
TASK	CT_EC									
case+punc	bleu	meteor	f1	prec	recl	wer	per	ter	gtm	nist
Proposed.CRR	0.2759	0.5328	0.5500	0.5150	0.5900	0.7421	0.5382	68.6970	0.6914	5.3888
Moses.CRR	0.3391	0.5744	0.6204	0.6430	0.5994	0.5942	0.4356	52.3780	0.6930	6.1764
Systran.CRR	0.2300	0.5063	0.5596	0.5599	0.5594	0.6993	0.4987	63.2230	0.6304	5.4766
Proposed.ASR.1	0.2214	0.4417	0.4516	0.4100	0.5025	0.8518	0.6447	80.8210	0.6399	4.5091
Moses.ASR.1	0.2853	0.5134	0.5604	0.5784	0.5435	0.6609	0.4986	59.2510	0.6331	5.4212
Systran.ASR.1	0.1902	0.4483	0.4986	0.4948	0.5025	0.7627	0.5683	70.5120	0.5689	4.6699

Discussion

<no native speakers>

Unknown Words

Proposed method :

very few unknown words

Grammatical Correctness

Proposed method:

more grammatically correct sentences.

However, the BLEU score was not so good?

Conclusion

Our System:

Two-stage machine translation system.

First stage : Rule-based machine translation

Second stage : SMT

Object:

Fewer unknown words &

Fewer ungrammatical sentences.

Results:

Not so good

Future:

a) Optimize parameters & reordering model

b) SYSTRAN ?

Results of Parameter Tunings

TASK	BTEC_CE									
case+punc	bleu	meteor	f1	prec	recl	wer	per	ter	gtm	nist
Proposed	0.3351	0.6256	0.6522	0.6301	0.6759	0.5704	0.4874	0.5048	0.6613	6.5972
Moses	0.3423	0.6135	0.6500	0.6463	0.6538	0.5436	0.4721	0.4674	0.6551	6.5624
Systran	0.1070	0.4697	0.5619	0.5671	0.5567	0.7017	0.6182	60.007	0.4863	3.9727