

# Statistical Pattern-Based Machine Translation with Statistical French-English Machine Translation

Jin'ichi Murakami, Takuya Nishimura, Masato Tokuhisa

Tottori University, Japan

# Problems of Phrase-Based SMT

Language model: *N*-gram

Local Language Information

No Grammatical Information

# Grammatical Information

## 1) Hierarchical Phrase-Based Machine Translation (JOSHUA)

Weak points:

Too many parameters

Unreliable score

Not always good performance

(Specially small training data)

## 2) Pattern-Based Machine Translation (PBMT)

Small number of Parameters

Trade-off Coverage and Translation Quality

Handmade (Too many costs)

# Proposed System: Two-stage MT System

First Stage : Automatic created PBMT System

- Low Coverage and High Quality
- Less Naturalness
- GIZA++, train-model.perl

Second Stage : Standard SMT System

- Add Naturalness
- MOSES

PBMT: preprocessing for SMT

# Related Work

## Two-stage MT system

M.Simard, et al: English-French  
SMT: 25.98 SYSTRAN+SMT :28.80

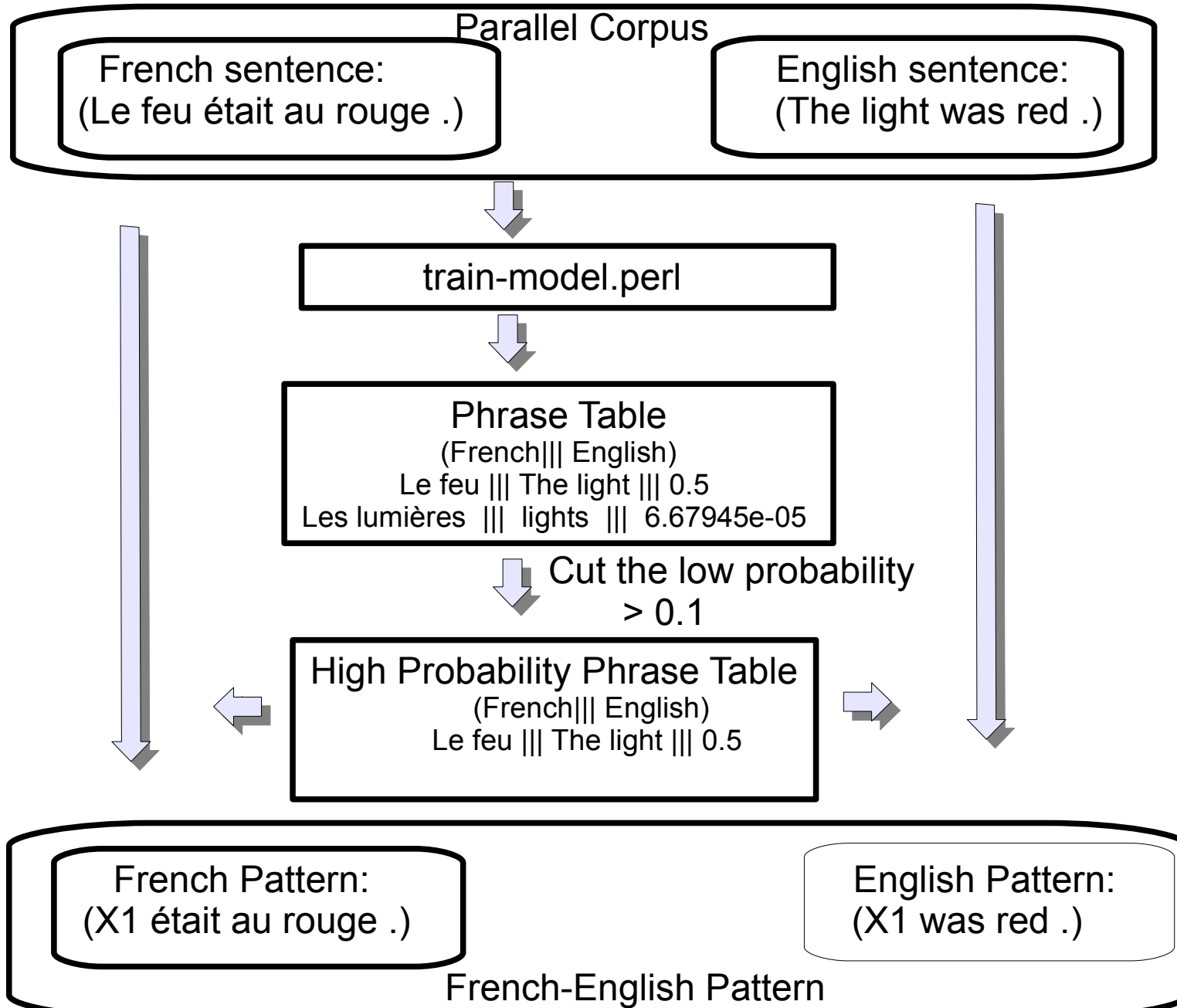
M.Simard, N.Ueffing, P.Isabelle, and R.Kuhn,  
"Rule-based translation with statistical phrase-based post-editing",  
in Second Workshop on SMT, pp.203-206, 2007.

Ehara: Japanese-English Patent task  
SMT: 28.21 RBMT+SMT: 29.21

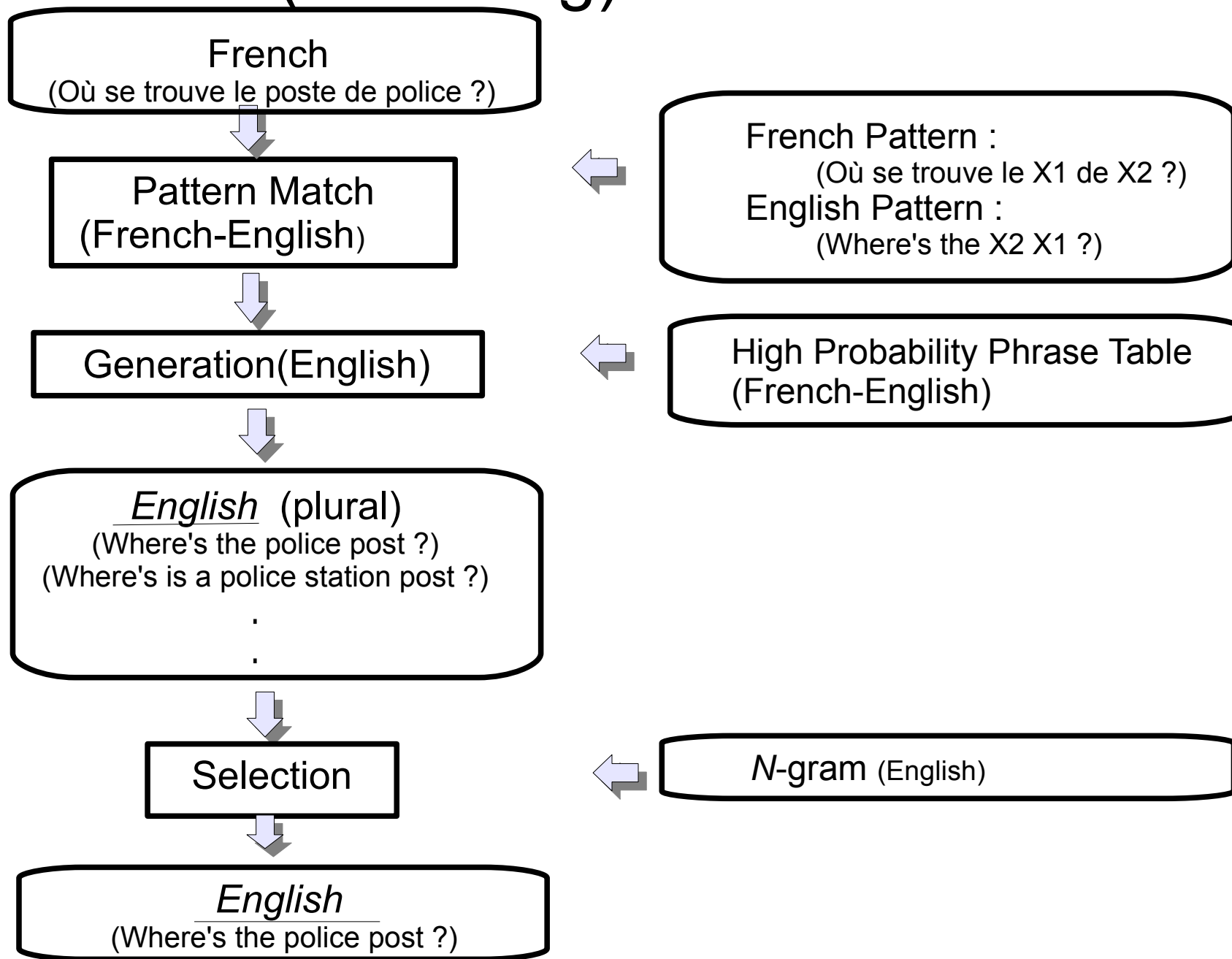
Terumasa EHARA, "Rule Based Machine Translation Combined with  
Statistical Post Editor for Japanese to English Patent Translation",  
Proceedings of Machine Translation Summit XI, Workshop on Patent  
Translation, pp.13-18, Sept., 2007.

SMT : post-processing for RBMT

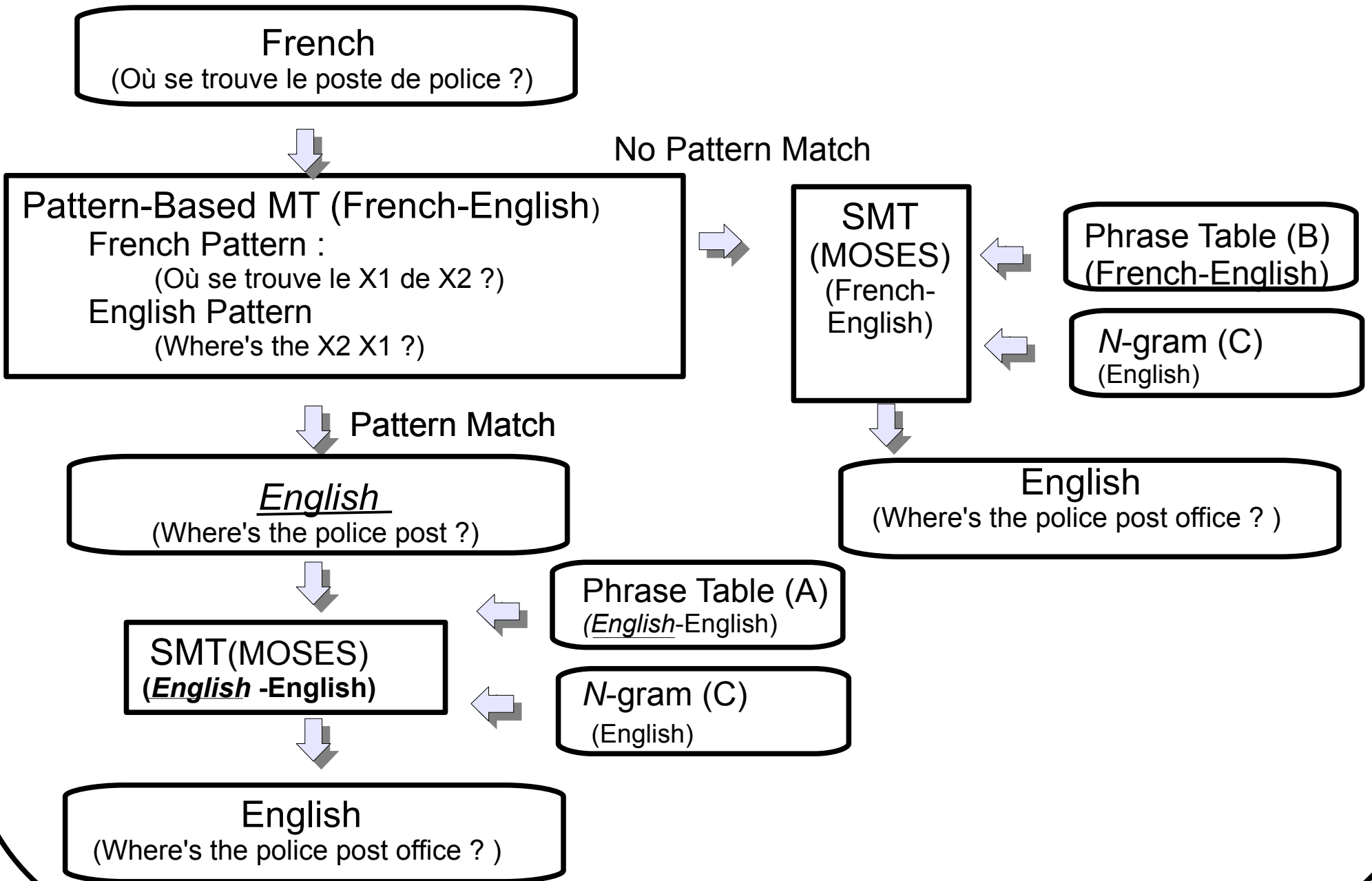
# Automatically Created Pattern-Based Machine Translation Making French-English Patterns



# Automatically Created Pattern-Based Machine Translation (Decoding)

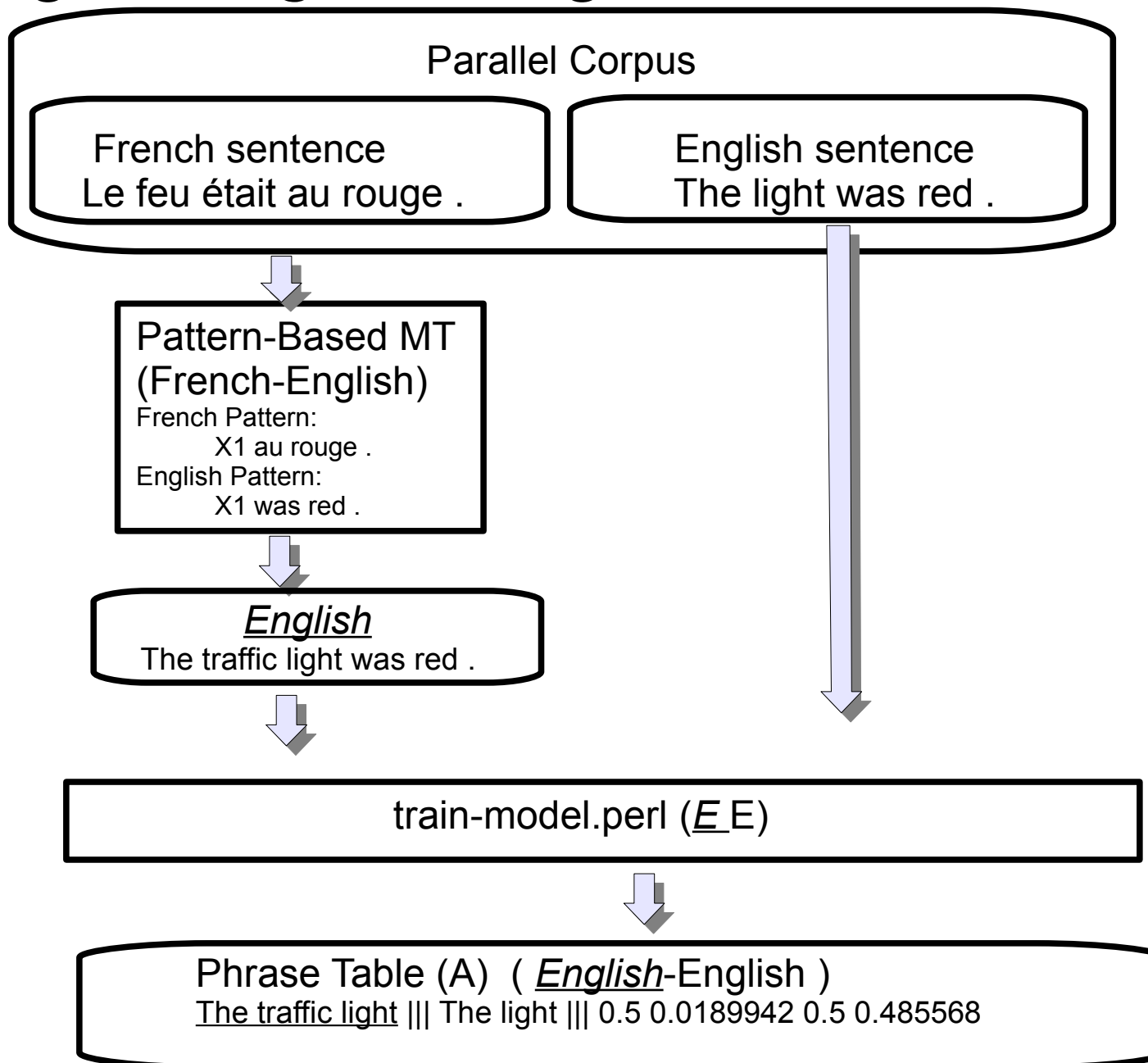


# Proposed System (Decoding)

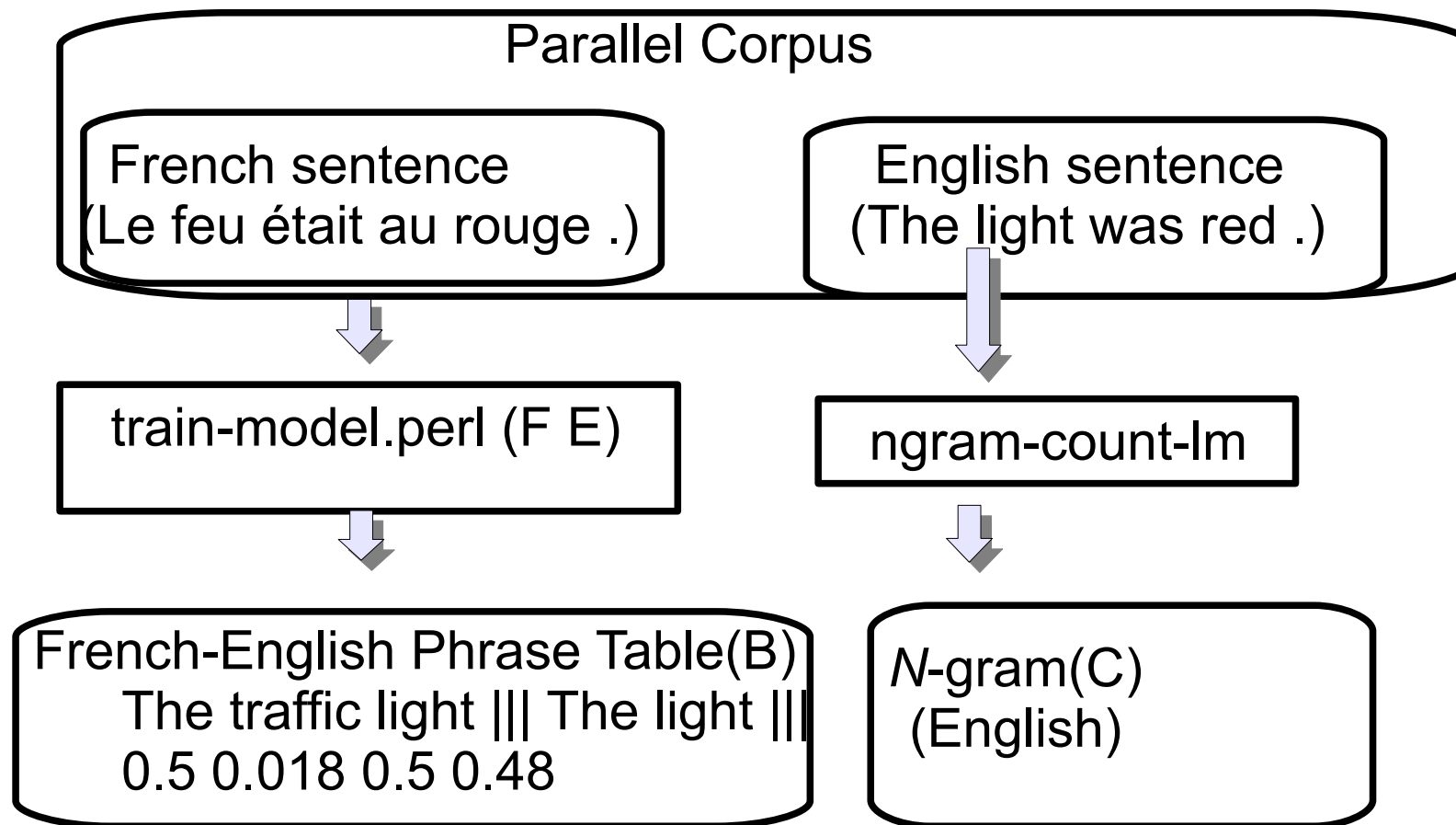




# Making of *English*-English Phrase Table (A)



# Making of French-English Phrase Table (B) and N-gram(C)



# Results of Experiments (BTEC-FE)

IWSLT10	BLEU	METEOR	WER	NIST
Proposed	0.5201	0.7916	0.3305	8.5812
MOSES	0.5077	0.7808	0.3365	8.4804
IWSLT09	BLEU	METEOR	WER	NIST
Proposed	0.5670	0.7844	0.3360	9.6467
MOSES	0.5504	0.7748	0.3541	9.4419

# Examples of Outputs (BTEC-FE)

2	Input	J'ai un rhume .
	Proposed	I have a cold .
	PBMT	I a your name cold .
	MOSES	I have a cold .
9	Input	Puis-je voir votre billet d'avion ?
	Proposed	Can I see your airline ticket ?
	PBMT	Can I see your ticket airline ?
	MOSES	Can I see your airline ticket ?
11	Input	Veillez me donner votre adresse .
	Proposed	Please me give your address .
	PBMT	Please me give your address .
	MOSES	Please give me your address .
12	Input	Je ne comprends pas .
	Proposed	I don't understand .
	PBMT	I don't understand don't .
	MOSES	I don't understand .
21	Input	Nous nous intéressons à la peinture .
	Proposed	We're interested in painting .
	PBMT	We're interested in paint .
	MOSES	We're interested in at the paint .
24	Input	C'est merveilleux .
	Proposed	That's wonderful .
	PBMT	It's wonderful much .
	MOSES	That's wonderful .
26	Input	Combien de temps allez-vous rester ?
	Proposed	How long will you be staying ?
	PBMT	How many long stay is it ?
	MOSES	How long will you be staying ?

28	Input	Avez-vous des pulls en cachemire ?
	Proposed	Do you have any sweater cashmere in ?
	PBMT	Do you have any sweater in cashmere ?
	MOSES	Do you have any pulls in of cashmere ?
30	Input	C'est notre limite .
	Proposed	It's our your limit .
	PBMT	It's our your name limit .
	MOSES	It's our latest .
32	Input	Je prends le vol dix pour Tokyo .
	Proposed	I'm taking flight for ten Tokyo .
	PBMT	I take to flight it ten for Tokyo .
	MOSES	I'm taking flight ten to Tokyo .
71	Input	Combien en tout ?
	Proposed	How much is in all ?
	PBMT	How much is in all ?
	MOSES	How much altogether ?
75	Input	Quel est le num 日本 ro de
		l'ambassade japonaise ?
	Proposed	What's Embassy Japanese number ?
	PBMT	What's Embassy Japanese number ?
	MOSES	What's the number of Japanese the Japanese embassy ?
356	Input	Je voudrais louer ce type de voiture pour une semaine .
	Proposed	I'd like to rent this type of car for a week .
	PBMT	I'd like to rent this type of car for a week .
	MOSES	I'd like to rent this kind of car for a week .

# Discussion 1

Analysis of Outputs:

Impossible to analyze these results  
(no French Speaker)

We feel more grammatically correct sentences

Proposed system affected PBMT.

Example No. 26 and No. 27

## Discussion 2

Many improvement points for PBMT.

Trade-off between a coverage and a translation quality

"High Probability Phrase Table"  
Threshold : 0.1 ( heuristic )

Many bugs in our system

# Conclusion

Our System : Two-stage MT system

Our goal : Less ungrammatical sentences

First stage : Automatically created PBMT system

Second stage : SMT

## Results

Proposed > Moses (BTEC-FE)

0.5201      0.5077 (BLEU)

## Future work

Our system:      7th (Total 9 systems)

# Additional Experiments two stage MT

IWSLT10						IWSLT09				
no_case+no_punc	BLEU	METEOR	WER	NIST		case+punc	BLEU	METEOR	WER	NIST
Proposed (=PBMT+MOSES)	0.5201	0.7916	0.3305	8.5812		Proposed (=PBMT+MOSES)	0.5894	0.8173	0.2932	9.2554
MOSES	0.5077	0.7808	0.3365	8.4804		MOSES	0.5793	0.8079	0.3065	9.1049
SYSTRAN+MOSES	0.5341	0.8001	0.3177	8.7090		SYSTRAN+MOSES	0.5985	0.8268	0.2814	9.3422
JOSHUA	0.4871	0.7648	0.3443	8.0456		JOSHUA	0.5696	0.7940	0.3096	8.8269
SYSTRAN+JOSHUA	0.5209	0.7892	0.3186	8.5260		SYSTRAN+JOSHUA	0.5850	0.8112	0.2923	9.0368
MOSES(no tuning)	0.4882	0.7700	0.3655	8.3291		MOSES (no tuning)	0.5574	0.7940	0.3275	8.8250
SYSTRAN+MOSES(no tuning)	0.5104	0.7909	0.3479	8.5575		SYSTRAN+MOSES (no tuning)	0.5809	0.8132	0.3071	9.1918
PBMT+MOSES (no tuning)	0.4991	0.7797	0.3600	8.4342		PBMT+MOSES (no tuning)	0.5670	0.8005	0.3235	9.0278
no_case+no_punc	BLEU	METEOR	WER	NIST		case+punc	BLEU	METEOR	WER	NIST
Proposed (=PBMT+MOSES)	0.4949	0.7606	0.3776	8.7709		Proposed (=PBMT+MOSES)	0.5670	0.7844	0.3360	9.6467
MOSES	0.4812	0.7488	0.3920	8.6752		MOSES	0.5504	0.7748	0.3541	9.4419
SYSTRAN+MOSES	0.5110	0.7707	0.3655	8.9056		SYSTRAN+MOSES	0.5742	0.7965	0.3198	9.7262
JOSHUA	0.4580	0.7292	0.4021	8.0676		JOSHUA	0.5438	0.7592	0.3587	9.1024
SYSTRAN+JOSHUA	0.4977	0.7604	0.3682	8.6502		SYSTRAN+JOSHUA	0.5603	0.7805	0.3289	9.3301
MOSES (no tuning)	0.4584	0.7351	0.4218	8.4744		MOSES(no tuning)	0.5301	0.7607	0.3784	9.1431
SYSTRAN+MOSES(no tuning)	0.4815	0.7610	0.3990	8.7320		SYSTRAN+MOSES(no tuning)	0.5494	0.7830	0.3486	9.5210
PBMT+MOSES (no tuning)	0.4711	0.7482	0.4110	8.6140		PBMT+MOSES (no tuning)	0.5381	0.7678	0.3684	9.3586

Best system : SYSTRAN + MOSES