



# Flat vs. Hierarchical Phrase-based Translation Models for Cross-Language Information Retrieval

Ferhan Ture<sup>1,2</sup> and Jimmy Lin<sup>1</sup>

<sup>1</sup>University of Maryland, <sup>2</sup>BBN Technologies

fture@bbn.com jimmylin@umd.edu

## Cross-Language Information Retrieval (CLIR)

1. Document translation
  2. Query translation
- We focus on query translation in this work, but our approach applies to documents as well.

**Token-to-Token Bilingual Mapping**  
+ represents alternatives probabilistically  
- ignores useful context

**Machine Translation**  
+ good use of context  
- standard system output provides only 1-best

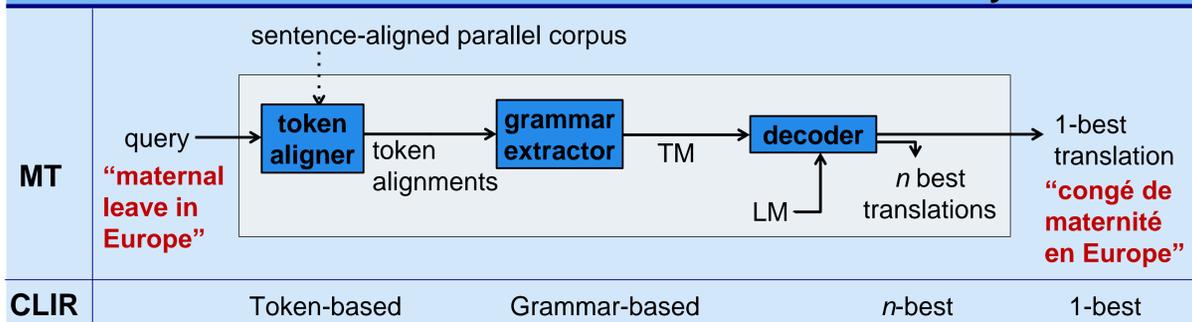
## Statistical Machine Translation (MT)

**Statistical Machine Translation** combines a language model (LM) and translation model (TM) to model translation of text from *source language* into *target language*.

Two types of TMs have become very successful:

- *Hierarchical* grammars allow reordering of non-consecutive phrases with variables
- *Flat* grammars are restricted to phrases that consist of consecutive text
- Both can handle 1-to-many and many-to-1 translations.

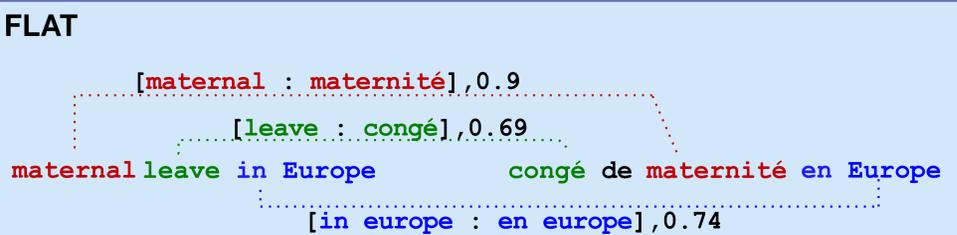
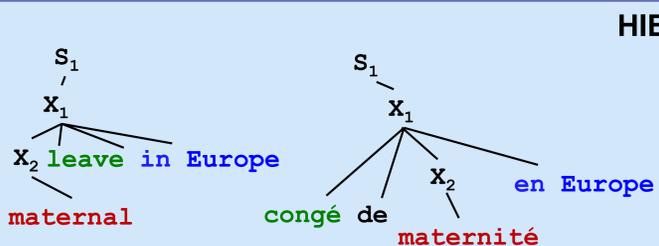
## MT for Query Translation in CLIR



- **Previous work** in [Ture et al., COLING'12]  
→ Using MT for CLIR: Introduced *token-based*, *grammar-based*, *n-best* and *interpolated* query translation approaches.
- **This paper:** MT systems mostly differ in *how they represent* the TM.  
→ How does different translation modeling strategies affect CLIR?

## Hierarchical vs. Flat Translation Models for Query Translation

A Translation Model contains rules in the format  $[ \alpha : \beta ] , \Pr(\beta|\alpha)$  where  $\alpha$  is a source-language phrase,  $\beta$  is its translation in the target language



$S \rightarrow [X : X] , 1.0$   
 $X \rightarrow [X_1 \text{ leave in europe} : \text{congé de } X_1 \text{ en europe}] , 0.9$   
 $X \rightarrow [\text{maternal} : \text{maternité}] , 0.9$   
 $X \rightarrow [X_1 \text{ leave} : \text{congé de } X_1] , 0.74$   
 $X \rightarrow [\text{leave} : \text{congé}] , 0.17$   
 $X \rightarrow [\text{leave} : \text{laisser}] , 0.49$   
 ...

$[\text{maternal} : \text{maternité}] , 0.9$   
 $[\text{in europe} : \text{en europe}] , 0.74$   
 $[\text{maternity leave} : \text{congé de maternité}] , 0.74$   
 $[\text{leave} : \text{congé}] , 0.17$   
 $[\text{leave} : \text{laisser}] , 0.49$   
 ...  
 $[\text{paternity leave} : \text{congé de paternité}] , 0.70$

## Multiple Alignment Heuristics

When a source token is aligned to multiple target tokens in a rule, apply one of the following heuristics:

Heuristic	one-to-one	one-to-many	one-to-none
Description	Treat each alignment independently	Combine target tokens into a multi-token phrase	Ignore 1-to-many alignments
Example rule	[ brand : marque de fabrique ] , 0.34		
Translations learned	(brand,marque) += 0.34 (brand,fabrique) += 0.34	(brand,marque de fabrique) += 0.34	-

## Experimental Results

Lang	TM	Token	Grammar			1-best		10-best			Interpolated		
			many	one	none	many/one	none	many	one	none	many	one	none
AR	hier	0.271	0.293	0.282	0.302	0.249	0.250	0.255	0.249	0.248	0.293*†	0.282‡	0.302*‡
	flat		0.274	0.266	0.273	0.249	0.232	0.264	0.254	0.249	0.280†	0.274	0.276
ZH	hier	0.150	0.182	0.188	0.170	0.155	0.155	0.159	0.159	0.159	0.192*†	0.193*	0.182*
	flat		0.156	0.151	0.167	0.155	0.146	0.169	0.163	0.163	0.183*†	0.177*	0.188*
FR	hier	0.262	0.297	0.288	0.292	0.276	0.235	0.307	0.304	0.295	0.318*†	0.314*	0.315*
	flat		0.264	0.257	0.262	0.297	0.242	0.289	0.300	0.282	0.307*	0.301	0.300

\* and † denote result is statistically significantly better (95%) than *token* and *1-best*, respectively.  
‡ denotes difference between *hierarchical* and *flat* MT is statistically significant (95%).

## Conclusions:

- **Hierarchical vs. Flat**  
→ Hierarchical > flat for grammar-based CLIR  
→ Hierarchical ~ flat for 10-best CLIR  
→ Flat much slower for grammar-based
- **Heuristics**  
→ *one-to-many* wins 18 out of 24 cases  
→ No difference in efficiency