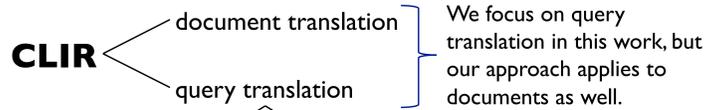


# Looking Inside the Box: Context-Sensitive Translation for CLIR

Ferhan Ture, Jimmy Lin, and Douglas W. Oard

University of Maryland

{fture | jimmylin | oard}@umd.edu



**token-to-token bilingual mapping (token2token)**

- + alternative translations and corresponding probabilities
- ignores useful contextual clues

**machine translation (MT)**

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

## State-of-the-art approaches in statistical MT

- model translation with synchronous context-free grammars (SCFGs)
- allow reordering of non-consecutive phrases
- can handle 1-to-many and many-to-1 translations

**SCFG** contains rules in the format:

$$[X] \mid \mid \mid \alpha \mid \mid \mid \beta \mid \mid \mid \Pr(\beta|\alpha)$$

$\alpha$  is a source-language phrase  
 $\beta$  is its translation in the target language

## Decoding

- searching for top  $N$  translations of a given sentence
- takes < 1 sec with modern decoders

$$t^* = \underset{t \in \text{Cand}(s)}{\text{argmax}} [\text{score}(t|s)] = \underset{r \in D(s \rightarrow t)}{\text{argmax}} \prod \ell(r) \text{LM}(t)$$

MT decoding task

$\text{score}(t|s)$  = decoder score of  $s$  being translated as  $t$

$\text{Cand}(s)$  = set of candidate translations of  $s$

$D(s \rightarrow t)$  = derivation of rules that generate  $s$  on LHS and  $t$  on RHS

$\ell(r)$  = likelihood of rule  $r$

$\text{LM}(t)$  = language model score of  $t$

## Example derivation

$s$  = "maternal leave in Europe"

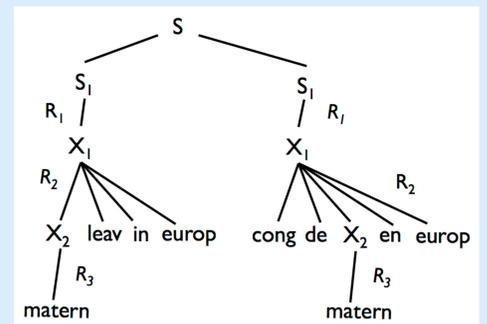
$R_1$  : [S] | | | [X] | | | [X] | | | 1.0

$R_2$  : [X] | | | [X,1] leave in europ | | | cong de

[X,1] en europ | | | 1.0

$R_3$  : [X] | | | matern | | | matern | | | 0.69

$t$  = "congé de maternité en Europe"

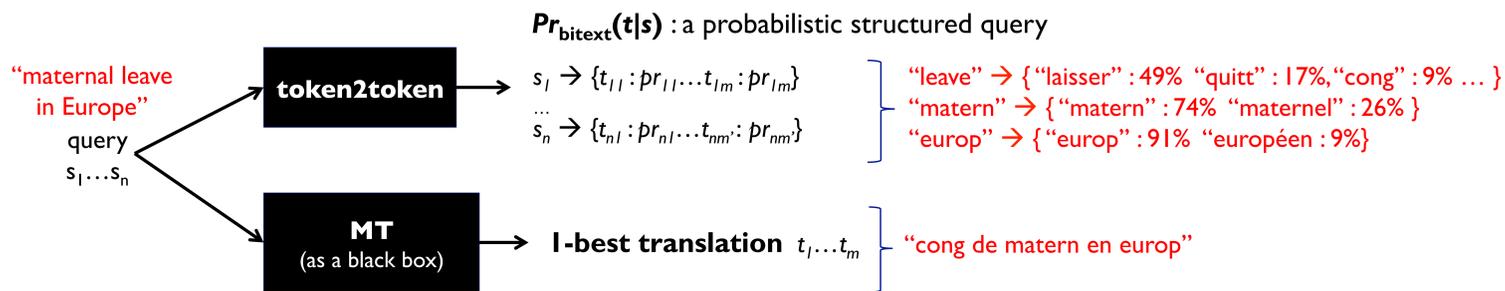


## Token-to-token translation in CLIR

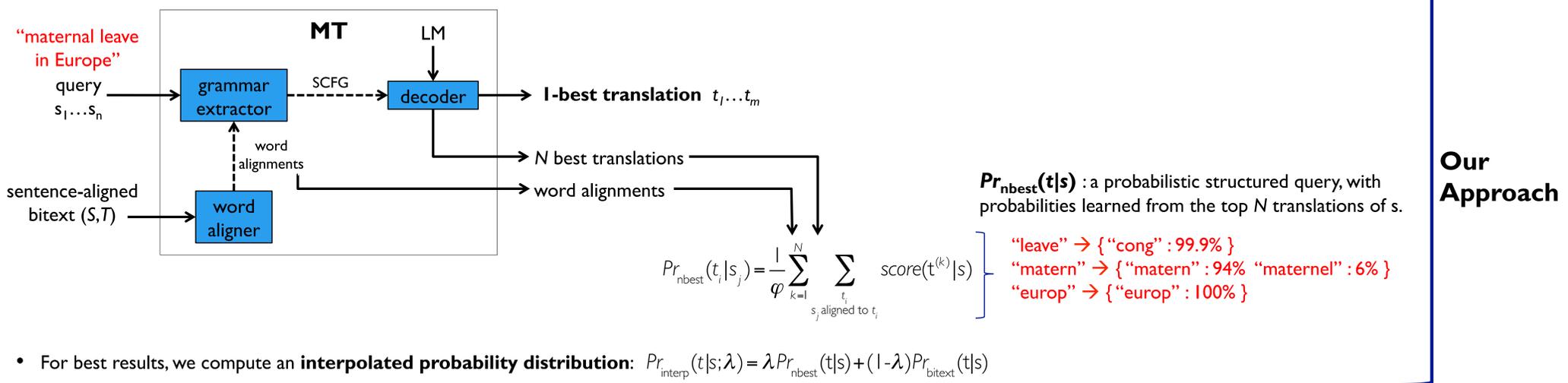
Compute a term and document frequency for each target-language term:

$$tf(s_j, d) = \sum_{\{t_i | Pr_{\text{bitext}}(t_i | s_j) > 0.005\}} tf(t_i, d) Pr_{\text{bitext}}(t_i | s_j)$$

- same formula for document frequency
- tf-idf model with bm25 scoring

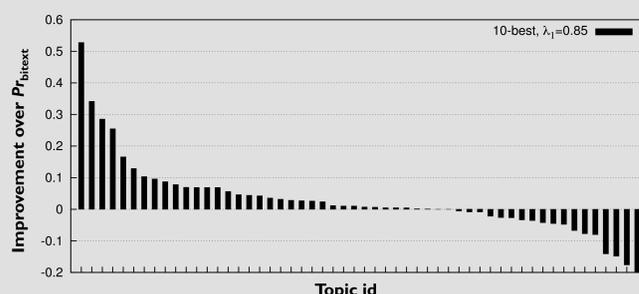
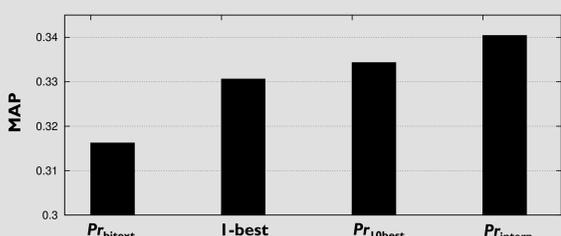


- CLIR literature → MT systems have been treated as a "black box" → (Magdy et al, 2011) show improvements when this is changed.
- Our approach → Looking inside the box:



## Effectiveness

Experimental results on TREC-5/6 (54 English topics and 164,778 Chinese documents):



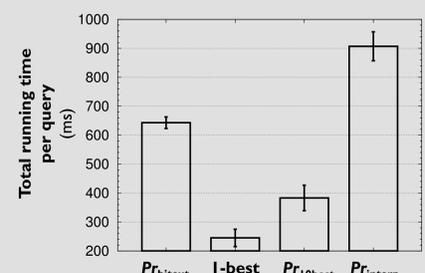
- Differences are not statistically significant.
- Results are consistent with theoretical expectations:
  - 10-best > 1-best
  - 10-best > bitext
  - interp > 10-best
- Grid search to find best-performing interpolation:  $\lambda_1 = 0.85$ , MAP = 0.3404
- $Pr_{\text{interp}}$  outperforms  $Pr_{\text{bitext}}$  in 36 out of 54 topics.

## Efficiency

Processes in all CLIR models:  
word-aligning bitext to train the translation model  
 + initialization + query generation + document retrieval

For MT-based CLIR approaches:

- extra processes: grammar extraction + decoding
- compact representation → retrieval takes less time



**10-best MT approach saves over 40% time!**