Multi-Pass Decoding With Complex Feature Guidance for Statistical Machine Translation

Benjamin Marie[1,2] *& Aurélien Max*[1,3]

(1) LIMSI-CNRS, (2) Lingua et Machina, (3) Université Paris-Sud

— LINGUA ET MACHINA —
L'ÉCRIT MULTILINGUE DANS L'ENTREPRISE

HIGHLIGHTS

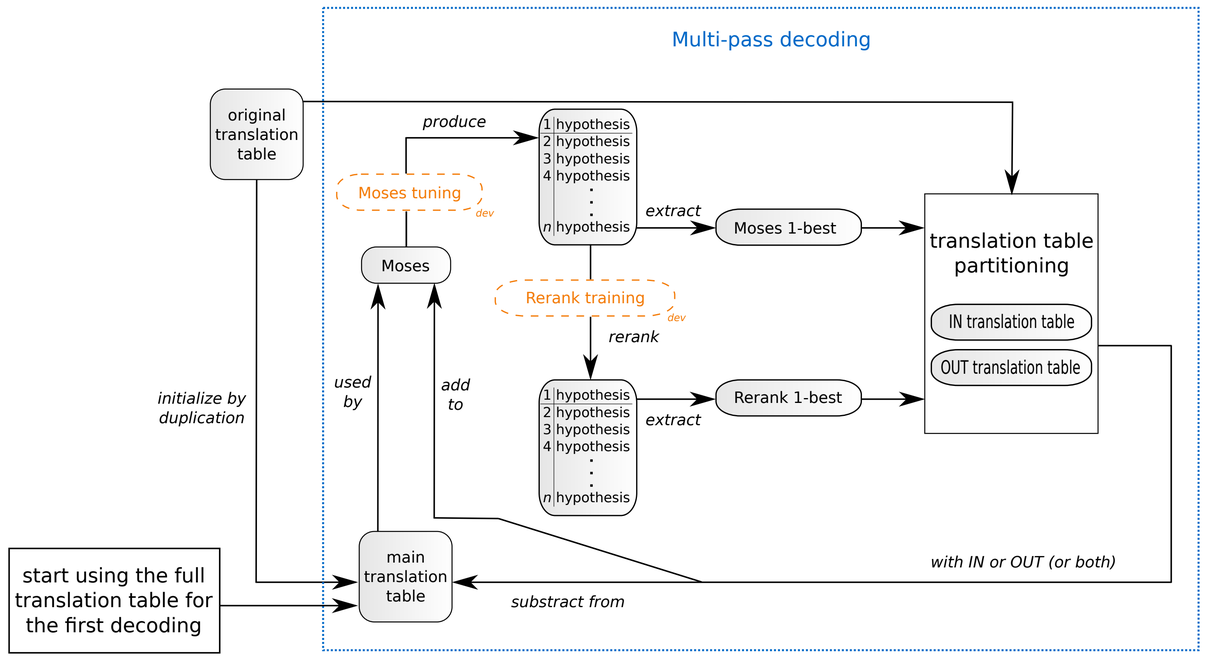
Using complex features in SMT

- some complex features are not easily integrable during decoding:
  - need of a complete hypothesis (e.g syntactic features)
  - computational cost (e.g neural network models)
  - not available during a first decoding (e.g word posteriors)
- such features are usually used through single-pass reranking

Main ideas

- exploit a reranking pass result during decoding
- isolate in separate translation tables the possibly misused bi-phrases to better optimize their feature weights

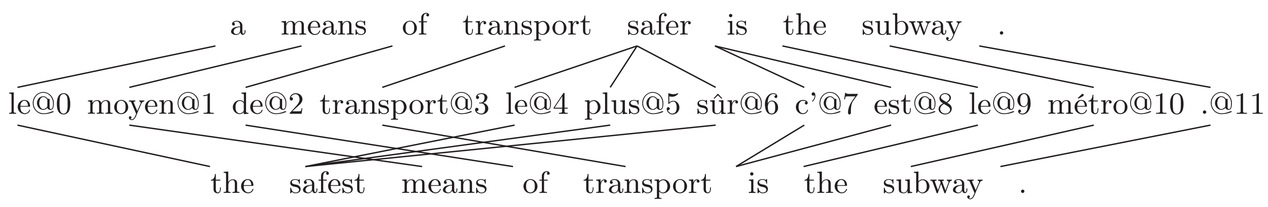
THE MULTI-PASS DECODING PROCEDURE



IN AND OUT TRANSLATION TABLES

- `IN`: contains bi-phrases of the reranking 1-best not in the decoder 1-best
- `OUT`: contains bi-phrases of the decoder 1-best not in the reranking 1-best
- extracted bi-phrases are removed from the original translation table
- `IN` or `OUT` (or both) are added to the next decoding pass

TRANSLATION TABLE PARTITIONING WITH LOCATED TOKENS



source	OUT	IN
le@0	a	the
le@0 moyen@1	a means	
le@0 moyen@1 de@2 transport@3	a means of transport	
le@4 plus@5 sûr@6	safer	safest

EXPERIMENTAL SETTINGS

- `Moses` tuned with `kb-mira`
- reranking system (`Rerank`) trained with `kb-mira` using the `Moses` 1,000-best
- complex features used during reranking:

Features	medical En→Fr	medical Fr→En	news En→Fr	news Fr→En
`Moses`	38.8	37.1	31.1	28.6
+ **MosesNorm**	38.9	37.2	31.1	28.7
+ **NeuralNet**	41.9	38.9	32.5	29.8
+ **POSLM**	39.2	37.7	31.1	28.9
+ **WPP**	39.1	37.1	31.2	28.6
+ **POSTagRatio**	38.9	37.3	31.1	28.8
+ **Syntax**	38.8	37.2	31.2	28.9
+ **IBM1**	39.1	37.2	30.9	28.8
`Rerank`	42.8	40.1	32.5	29.9

MULTI-PASS DECODING RESULTS FOR ALL CONFIGURATIONS

Configuration		medical En→Fr test	medical En→Fr # iter.	medical Fr→En test	medical Fr→En # iter.	news En→Fr test	news En→Fr # iter.	news Fr→En test	news Fr→En # iter.
baseline	`Moses`	38.8	-	37.1	-	31.1	-	28.6	-
	`Rerank`	42.8		40.1		32.5		29.9	
`OUT`	`Moses`	41.8	4	38.7	3	31.8	1	29.2	1
	`Rerank`	43.8		40.5		32.9		30.3	
`IN`	`Moses`	43.2	4	39.9	3	32.4	2	29.3	2
	`Rerank`	**44.2**		**41.0**		**33.2**		**30.4**	
`IN` and `OUT`	`Moses`	42.4	4	38.7	3	32.1	2	29.2	2
	`Rerank`	43.5		40.6		32.9		**30.4**	

- `IN` alone gives the best results
- "`IN` and `OUT`" performs between the configurations using `IN` and `OUT` alone
- most configurations converge quickly (between 1 and 4 iterations)

CONCLUSION & FUTURE WORK

Conclusion

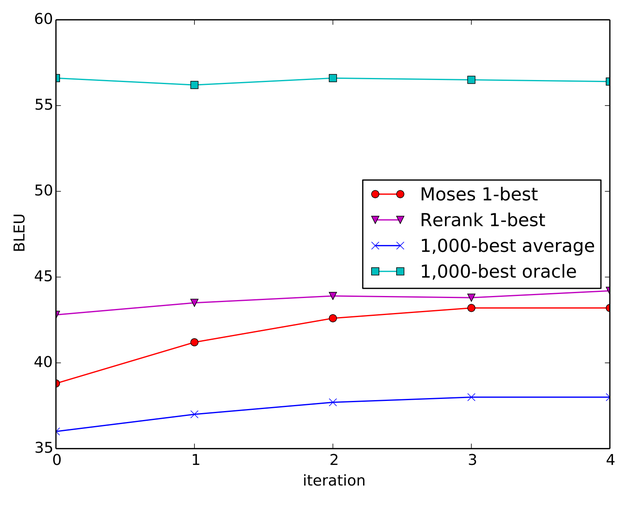
- strong and consistent improvements for all configurations
- simple criterion to iteratively partition the translation table
- makes a better use of complex features than single-pass reranking

Future work

- add features to the new translation tables to use more informations about the reranking result during decoding
- use the result of a rewriting system (Marie and Max, 2014) to guide the decoder
- enhance the diversity in the n-best list (Chatterjee and Cancedda, 2010; Gimpel et al., 2013) to train a better reranking system
- add more complex features to the reranking pass

INCREASING OF THE n-BEST QUALITY

- analysis for the `IN` configuration (`medical` En→Fr):



- quickly reduces the gap between `Moses` and `Rerank` BLEU scores
- 1,000-best average BLEU increases over the iterations