

Reranking Translation Candidates Produced by Several Bilingual Word Similarity Sources



Laurent Jakubina and Philippe Langlais
RALI – Université de Montréal, Québec, Canada
rali.iro.umontreal.ca



Université
de Montréal

Approach: Reranked Combination with Features

Weighting [EN] : pondération [FR]

	Rapp		Miko		Faru
maximise	0.069	pondération	0.753	pondération	0.428
quantifié	0.066	pondéré	0.688	calculé	0.389
adaptatif	0.065	dépendra	0.648	quantification	0.387
pondération	0.066	calculé	0.622	luminance	0.336
pondéré	0.063	maximise	0.62	pondéré	0.319
calculé	0.054	quantifié	0.58	quantifié	0.307
dépendra	0.054	luminance	0.567	maximise	0.268
liquidité	0.052	dénominateur	0.55	inférieure	0.267

RERANKED COMBINATION (with FEATURES)

Results : **pondération (0.020)** ; maximise (0.0027) ; *pondéré (0.0021)* ;
quantifié (0,0019) ; calculé (0,0010) ; ... ; dépendra (...); ...

Protocol and Data

Comparable Corpus

- English and French Wikipedia from June 2013

3 Test Sets (1000 terms and their translation)

- Wiki_{>25} : English words seen at least 26 times in Wikipedia
- Wiki_{≤25} : English words seen below 26 times in Wikipedia
- Euro_{5-6k} : Same as (Mikolov and al., 2013b)

Reranker

- RankLib
- Trained to optimize P@1
- 3-fold cross validation
- 700 terms for training ; 300 for testing
- 3 sets of features: rank (2), string (5), frequency (4)

Metrics

- Accuracy (precision) at rank 1, 5 and 20

Experiments

	INDIVIDUAL			1-RERANKED			n-RERANKED		
	@1	@5	@20	@1	@5	@20	@1	@5	@20
Wiki _{>25}									oracle: 69.3
Rapp	20.0	33.0	43.0	36.3 ^{2.5}	48.8 ^{1.9}	53.8 ^{1.9}	base	34.3 ^{1.9}	47.6 ^{1.4} 58.8 ^{0.8}
Miko	17.0	32.6	41.6	38.1 ^{1.9}	49.0 ^{1.5}	54.3 ^{1.3}	R+M	43.3 ^{2.9}	58.4 ^{1.4} 62.4 ^{3.1}
Faru	13.3	26.0	33.3	34.3 ^{1.5}	44.0 ^{2.6}	47.9 ^{2.1}	R+M+F	45.6^{2.2}	59.6^{1.1} 64.0^{1.8}
Wiki _{≤25}									oracle: 28.6
Rapp	2.6	4.3	7.3	8.6 ^{1.2}	9.4 ^{0.8}	10.2 ^{1.0}	base	10.7 ^{0.6}	15.9 ^{1.2} 21.8 ^{0.7}
Miko	1.6	4.6	10.6	16.6 ^{2.2}	19.0 ^{1.5}	20.1 ^{1.4}	R+M	18.9 ^{2.01}	22.0 ^{1.3} 23.6 ^{2.2}
Faru	1.6	2.6	5.0	7.9 ^{2.2}	8.7 ^{2.5}	8.9 ^{2.7}	R+M+F	21.3^{1.86}	24.4^{1.7} 25.7^{1.9}
Euro _{5-6k}									oracle: 84.4
Rapp	16.6	31.8	41.2	34.6 ^{5.7}	48.6 ^{1.2}	51.9 ^{1.2}	base	33.6 ^{1.2}	59.3 ^{1.4} 71.7 ^{2.5}
Miko	42.0	59.0	67.8	47.0 ^{2.3}	68.1 ^{2.7}	73.0 ^{1.7}	R+M	49.5^{3.7}	68.7^{1.5} 76.1 ^{1.0}
Faru	30.6	47.7	59.8	41.2 ^{3.9}	58.0 ^{3.5}	66.0 ^{3.5}	R+M+F	47.6 ^{2.3}	68.5 ^{2.0} 76.2^{1.2}

Individual Approaches

- Rapp : In-house implementation of (Rapp, 1995)
- Miko : (monolingual) embeddings trained on Wikipedia (Mikolov and al., 2013b) ; projection learned (Dinu and Baroni, 2015)
- Faru : bilingual embeddings with CCA (Faruqui and Dyer, 2014)
- Best meta-parameters on each data set

Reranking Individual Approaches

- Best performance with *Random Forest*
- 15% gains on average

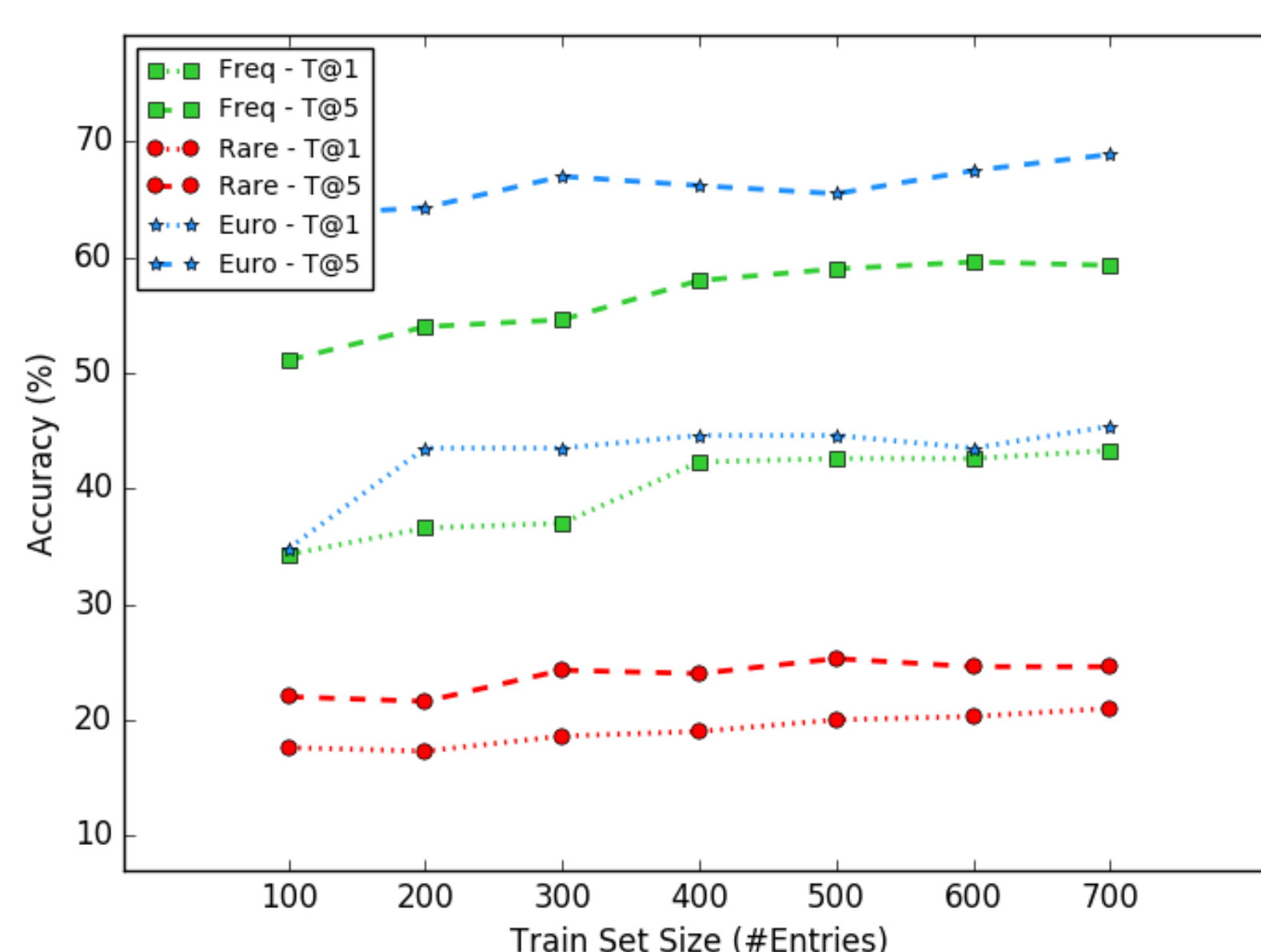
Combining by Reranking

- Systematic gains for the 3 test sets
- Close to oracle performances

Analysis

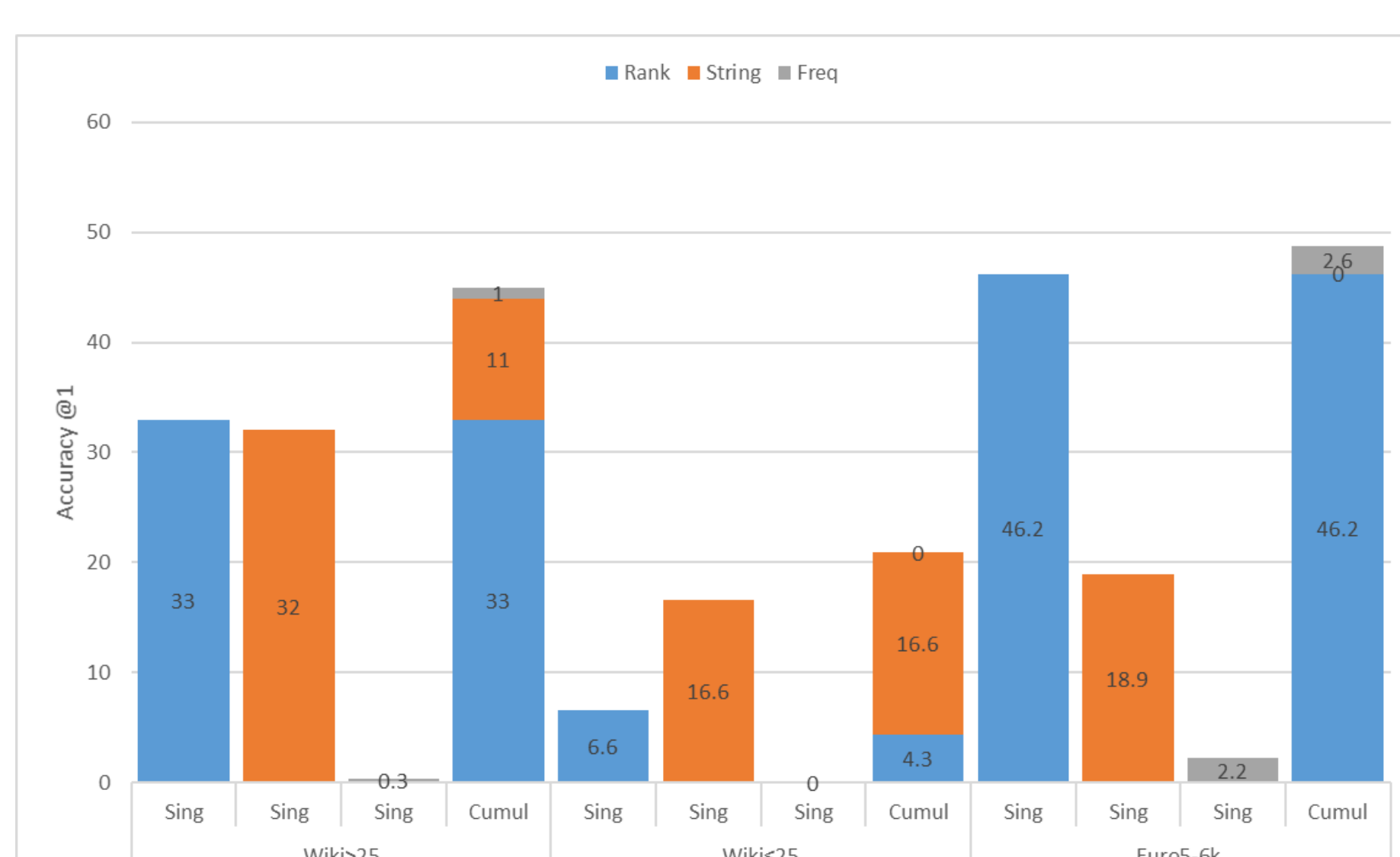
Training Size

- Test = 300 test words (fixed).
- More training examples
→ better results.
- But 200 training examples
seem good enough.



Feature Selection

- Overall: rank >> string >> frequency
- Using all features is preferable.



Reranker Analysis

- Always keeps solutions proposed at @1 by (at least) two approaches.
- Rare that the reranker picks the reference solution at @1 if no approach did it.
- Still the average rank of the reference solution decreases.

	Wiki _{>25}	Wiki _{≤25}	Euro _{5-6k}
Rapp	12.7	19.6	16.2
Miko	16.3	30.0	7.5
Faru	20.4	35.5	11.3
<i>list-oracle</i>	12.3	9.1	7.1
reranker	5.6	4.0	4.9

Average rank of the reference translation

Error Analysis

- Manual inspection of the first candidate by our best reranker for the first 100 test forms for which the candidate translation differs from the reference one.
- % of JUNK errors is much higher on Wiki_{≤25}
→ another illustration of the bias in favor of frequent terms.

	Wiki _{>25}	Wiki _{≤25}	Euro _{5-6k}
MORPHO	18	3	26
RELATED	16	4	23
synonyms	15	1	19
antonyms	1	2	2
hyponym			1
cohyponym		1	1
POLYSEMY	4	0	5
LOOSLY	14	15	20
ENGLISH	21	6	7
JUNK	27	72	19