Supplementary Material

for

Translation Induction on Indian Language Corpora using Translingual Themes from Other Languages

1 Shared Data Sets and Code

1.1 Translation Correspondences

We applied the method to 30 language pairs using three auxiliary languages, and to 12 language pairs (involving Hindi) using two auxiliary languages—resulting in 42 translation correspondences.[1] The MRR was above 0.33 on the test set in most cases, which means that the correct translation was in the top 3 for most language pairs. Hence, with reasonable confidence, we can expect that the correct translation will be among the top 10 for the rest of the vocabulary. We are releasing the following data for every source-target language pair:

- For each word s in the source vocabulary:
  - Compute the score $S_A(t|s)$ for every word t in the target vocabulary.
  - Sort the target vocabulary in the descending order of these scores.
  - Save s along with the top 10 target words, and their scores.

Applications

1. There exist no have dictionaries for most of the language pairs we considered. The released TCs can be used by lexicographers to quickly bootstrap a dictionary in each of these language pairs.
2. We provide scores along with the translation candidates, so that the TCs can be used to probabilistic dictionaries for NLP applications in these language pairs. In addition, new algorithms or resources that might be developed in future can be directly applied to the TCs to improve the ranking of the correct translation.

[1] $L1 - L2$ and $L2 - L1$ are considered different pairs.

1.2 Corpora

We extracted, cleaned and preprocessed 35 [2] comparable corpora (see main paper, and Section 2 below) and 21 gold standard translation pair sets (see Section 2.1) from Wikipedia. The processing steps followed are described in the main paper. We are publicly releasing this data. This will enable researchers to quickly bootstrap CC-TCI research in Indian languages, and also to reproduce our results.

1.3 Software

WikiTSu We built a system *WikiTSu*for suggesting target Wikipedia titles given a source word, which internally uses the methods described in the paper for model estimation and title suggestion. We are releasing the source code (in Python 2.6) for the system publicly for use in Wikipedia-based systems, and also for researchers to easily try our methods on other data sets/languages.

WiCCX We built an in-house tool *WiCCX*for automating the generation of preprocessed and algorithm-ready comparable corpora from Wikipedia dumps. We are releasing the source code (in Python 2.6) [3] for the tool publicly to enable researchers to quickly assemble comparable corpora in their favorite languages for research on comparable corpora.

2 Wikipedia Comparable Corpus

Preprocessing. The Wikipedia XML dumps[4] for the 9 languages—Bengali (bn), Hindi (hi), Kannada (kn), Malayalam (ml), Marathi (mr), Tamil (ta), Telugu (te), English (en) and French (fr)—are from the period May–Nov, 2012. These dumps were processed using *WiCCX*, an in-house tool for extracting and preparing comparable corpora from Wikipedia dumps. The main tasks achieved by the tool are: (1) Create the largest possible comparable corpora for a language pair using Wikipedia `langlinks`. This resulted in 35 comparable corpora. (2) Convert all documents to plain text (removing Wikipedia markup). Remove all punctuation, numbers, and characters from other scripts. Segment the text into words based on whitespace. (3) Remove stopwords. Many languages do not have stopword lists, so we use the following criteria for identifying "useful" words. For each corpus, remove words with corpus frequency[5] $cf < 20$ and document frequency $df > 0.1 \times D$, where D = number of documents. This eliminates very rare words and most function words (such as *the, of, in,* etc.). Order the

[2] $\binom{7}{3} = 21$ corpora for the Indian languages used in the experiments, and $7 \times 2 = 14$ corpora between the Indian languages and English/French.

[3] The tool internally uses the `gwtwiki` API, which is publicly available, in one of the processing steps.

[4] `http://dumps.wikimedia.org/`

[5] The number of occurrences of the word in the corpus.

remaining words in decreasing order of average document count $adc = \frac{cf}{df}$, and take the top V words[6] as the vocabulary. (4) Discarded documents with fewer than 30 tokens, or with fewer than 10 unique tokens.[7] Some statistics about the corpus are given in Table 1.

A note on corpus quality. Table 1 shows a disparity in the size of the corpora for different language pairs. For example, many language pairs such as bn-kn have fewer than 1000 document pairs. To some extent, the problem of document size variability has been addressed in the preprocessing phase, as can be seen from the average document size in each language. However, the degree of "comparability" is quite low for several corpora, measured in terms of average *document size ratio*, the ratio of the size of the larger document to the size of the smaller document within a document pair. For example, in the bn-kn corpus, given two documents in a pair, one document is nearly 8 times the size of the other, on average, making the documents less comparable to each other. The presence of this kind of skew [8] adversely affects the quality of the models learned using any method.

2.1 Gold standard for evaluation

The *WiCCX* tool also compiles translation pairs using `langlinks` between article titles—an approach discussed in earlier work [1]. The main steps, for each language pair, are: (1) Compile the largest possible set of translation pairs that can be obtained using the Wikipedia `langlinks` between article titles. (2) Discard titles containing more than one word (splitting on whitespace), and titles containing numbers or punctuation. This gives a gold set. (3) Create reduced gold sets for each auxiliary language set by removing words that are not present in the auxiliary corpora. Thus, for each language pair, we obtained several gold sets $G(A)$ depending on the choice of the auxiliary language set A.

The sizes of the reduced gold sets for each language pair, for different choices of the auxiliary language set A, are shown in Table 2.

3 Evaluation measures

Given a test set in languages L_1 and L_2, for each L_1 word in the test set, each method was used to generate a ranked list of candidate L_2 words. Similarly, L_1 candidates were generated for L_2 words. Each ranked list was evaluated for mean reciprocal rank (MRR) [2] and "Presence-at-k" (Pres@k) for k = 1 and 5. Let

[6] We tried with V = 2000, 4000, 5000, 6000, 8000, and 10000, and found the best results for V=5000.

[7] This resulted in a reduction in vocabulary size of around 303 words, on average.

[8] An aligned CC is skewed if the documents in a pair have drastically different sizes.

$tr(w)$ be the translation of w in the gold set. Given a ranked list generated for w,

$$RR(w) = \frac{1}{\text{Rank of } tr(w) \text{ in the list}} ,$$

$$Pres@k(w) = \begin{cases} 1 & \text{if } tr(w) \text{ is present in} \\ & \text{the top-}k \text{ in the list} \\ 0 & \text{otherwise} \end{cases}$$

Reciprocal rank measures the overall ability of a method to rank translations higher. Pres@1 measures the success of a method at suggesting translations. Pres@5 measures the usefulness of the method in certain settings, e.g. when a human can review the first 5 candidates and pick the best one.[9] The performance values were averaged over all words in the test set, and again averaged over all folds in the cross-validation. The final scores are reported in the paper.

4 Scoring function and baseline

The TI+Cue method as proposed in [3] yielded the scoring function S_{raw} (as described in the paper) which was used by AUX-COMB, and is described below.

Let the source and target vocabulary sizes be V_S and V_T respectively, and the number of topics be K. Topic modeling gives us a model ϕ such that, for every source word $s \in \{1 \dots V_S\}$ and topic $\tau \in \{1 \dots K\}$, we have $p(s|\tau) = \phi^S_{\tau s}$. During the modeling, each token in the corpus is labeled with a topic. Let $N^S_{\tau s}$ be the number of times s was labeled τ. Define similar expressions for each target word t. The scoring function $S_{raw}(t|s)$ is defined as follows. Let

$$Cue(t|s) = \sum_{\tau=1}^{K} \phi^T_{\tau t} \frac{\phi^S_{\tau s}}{Z_s},$$

$$\left(TI^S_s\right)_\tau = \frac{N^S_{\tau s}}{\sum_{s' \in \{1 \dots V_S\}} N^S_{\tau s}} \log\left(\frac{K}{1+|\tau' : N^S_{\tau' s} > 0|}\right),$$

$$TI(t|s) = \frac{TI^S_s \cdot TI^T_t}{|TI^S_s||TI^T_t|} \; ; \text{ then}$$

$$S_{raw}(t|s) = TI(t|s) + \alpha \; Cue(t|s)$$

where Z_s is a normalization factor, TI^T_t is defined similar to TI^S_s, $\cdot$ is the dot product operator, $|\ |$ is the Euclidean norm, and $\alpha = 10$ was set empirically.

5 Results for CC-TCI

In this section, we report the results for CC-TCI for different auxiliary language sets and different algorithm settings. In each table, there are separate sections

[9] We chose Pres@k instead of Precision-at-k (Prec@k) since our gold set has exactly 1 translation per word, so that measuring precision was not very meaningful. Note that Pres@1 is the same as Prec@1.

for each evaluation measure—MRR, Pres@1 and Pres@5. The number in row L_1 and column L_2 is the performance measured when identifying translations for L_1 words in L_2.

5.1 Experiments using $G(en)$

For the gold set $G(en)$, the results are as below:

- Performance of TI+Cue: Table 3
- Performance of AUX-COMB using grid search: Table 4
- Percentage improvement for AUX-COMB using grid search: Table 5
- Performance of AUX-COMB using EM: Table 6
- Percentage improvement for AUX-COMB using EM: Table 7

5.2 Experiments using $G(fr)$

For the gold set $G(fr)$, the results are as below:

- Performance of TI+Cue: Table 8
- Performance of AUX-COMB using grid search: Table 9
- Percentage improvement for AUX-COMB using grid search: Table 10
- Performance of AUX-COMB using EM: Table 11
- Percentage improvement for AUX-COMB using EM: Table 12

5.3 Experiments using $G(en, fr)$

For the gold set $G(en, fr)$, the results are as below:

- Performance of TI+Cue: Table 13
- Performance of AUX-COMB using grid search: Table 14
- Percentage improvement for AUX-COMB using grid search: Table 15
- Performance of AUX-COMB using EM: Table 16
- Percentage improvement for AUX-COMB using EM: Table 17

Grid search better than EM. Looking at the results so far, we found that grid search was doing better than EM consistently. Hence, for the remaining experiments, we used only grid search.

5.4 Experiments using $G(hi)$

For the gold set $G(hi)$, the results are as below:

- Performance of TI+Cue: Table 18
- Performance of AUX-COMB using grid search: Table 19
- Percentage improvement for AUX-COMB using grid search: Table 20

5.5 Experiments using $G(en, fr, hi)$

For the gold set $G(en, fr, hi)$, the results are as below:

- Performance of TI+Cue: Table 21
- Performance of AUX-COMB using grid search: Table 22
- Percentage improvement for AUX-COMB using grid search: Table 23

References

1. Erdmann, M., Nakayama, K., Hara, T., Nishio, S.: Improving the extraction of bilingual terminology from wikipedia. ACM Trans. Multimedia Comput. Commun. Appl. (2009)
2. Voorhees, E.M., et al.: The trec-8 question answering track report. In: TREC. (1999)
3. Vulić, I., De Smet, W., Moens, M.: Identifying word translations from comparable corpora using latent topic models. In: ACL-HLT. (2011)

Table 1. Number of document pairs (before and after preprocessing), average document size, and average document size ratio (see Section 2) for corpora of different language pairs L_1-L_2.

L_1-L_2	No. of doc pairs		Avg doc size		Avg doc size ratio
	before prepro	after prepro	L_1	L_2	
bn-en	18721	2438	134 ± 188	166 ± 157	3.4 ± 3.3
bn-fr	12663	2684	135 ± 204	202 ± 253	3.7 ± 4.5
bn-hi	7082	1831	197 ± 304	330 ± 484	5.2 ± 7.6
bn-kn	3023	752	276 ± 431	555 ± 535	7.7 ± 8.7
bn-ml	4851	1676	205 ± 311	178 ± 286	3.9 ± 5.9
bn-mr	5979	1329	236 ± 374	197 ± 372	4.1 ± 6.5
bn-ta	5726	1590	199 ± 314	259 ± 436	5.0 ± 7.7
bn-te	3103	1103	244 ± 381	416 ± 554	6.6 ± 8.8
en-hi	22870	2027	149 ± 134	300 ± 317	3.9 ± 3.6
en-kn	6249	2066	237 ± 211	330 ± 321	2.3 ± 2.0
en-ml	19232	1608	166 ± 160	123 ± 169	3.5 ± 3.8
en-mr	17500	2853	180 ± 160	151 ± 310	3.5 ± 3.8
en-ta	32633	1374	148 ± 131	241 ± 299	3.9 ± 4.0
en-te	10120	2578	199 ± 190	336 ± 384	3.4 ± 3.3
fr-hi	16338	2218	188 ± 245	293 ± 332	4.4 ± 4.7
fr-kn	4819	1542	264 ± 314	389 ± 334	3.6 ± 4.9
fr-ml	11952	1883	184 ± 230	129 ± 194	3.6 ± 4.4
fr-mr	11916	2689	223 ± 311	144 ± 274	4.0 ± 5.3
fr-ta	16771	1589	169 ± 204	242 ± 317	4.2 ± 4.5
fr-te	6087	1801	231 ± 303	411 ± 441	4.6 ± 5.6
hi-kn	4201	1350	461 ± 468	365 ± 334	3.2 ± 4.1
hi-ml	6610	1885	255 ± 362	153 ± 232	4.8 ± 7.2
hi-mr	7069	1751	296 ± 473	207 ± 743	4.7 ± 7.6
hi-ta	8423	1667	331 ± 407	282 ± 370	4.0 ± 5.8
hi-te	5052	1764	368 ± 409	349 ± 381	3.0 ± 4.8
kn-ml	3216	952	443 ± 468	207 ± 350	7.2 ± 9.3
kn-mr	2944	605	545 ± 606	213 ± 358	8.6 ± 10.7
kn-ta	3927	962	400 ± 370	441 ± 505	4.5 ± 6.0
kn-te	2980	1043	473 ± 449	539 ± 543	3.0 ± 4.9
ml-mr	4581	1472	171 ± 279	181 ± 305	3.6 ± 4.8
ml-ta	7540	1733	144 ± 231	209 ± 317	4.3 ± 6.2
ml-te	3963	1380	186 ± 315	353 ± 518	6.4 ± 9.5
mr-ta	6245	1826	175 ± 290	204 ± 380	4.0 ± 6.4
mr-te	3247	953	204 ± 345	399 ± 580	6.7 ± 10.1
ta-te	4795	1315	317 ± 451	343 ± 428	4.7 ± 7.2

Table 2. Gold standard translation pair set sizes for different language pairs: the full set (column 2), keeping only translations occurring in the corresponding comparable corpus (column 3), and the reduced sets for different choices of the auxiliary language set A (columns 4-8).

$L_1 - L_2$	Full	Corp. restr.	G(en)	G(fr)	G(en,fr)	G(hi)	G(en, fr,hi)
bnhi	2043	378	107	139	97	–	–
bnkn	1087	182	52	71	51	89	48
bnml	1945	259	110	174	106	205	93
bnmr	1783	353	162	252	157	282	150
bnta	1858	316	73	116	69	175	61
bnte	1237	260	86	126	82	142	75
hikn	1361	216	102	124	92	–	–
himl	2522	250	77	112	70	–	–
himr	2281	398	95	136	88	–	–
hita	2608	293	113	147	104	–	–
hite	1873	322	146	188	137	–	–
knml	1274	148	44	58	42	80	37
knmr	1119	172	57	73	51	97	48
knta	1310	188	76	94	72	118	60
knte	1049	223	105	122	96	138	81
mlmr	1961	240	95	144	90	179	83
mlta	2697	222	69	102	57	152	50
mlte	1751	208	62	86	54	112	49
mrta	2092	306	79	113	71	159	61
mrte	1372	253	91	118	83	134	78
tate	1725	263	77	110	70	139	60

Table 3. Performance of TI+Cue on $G(en)$.

MRR	bn	hi	kn	ml	mr	ta	te
bn	–	0.3174	0.1842	0.2422	0.2439	0.2923	0.2271
hi	0.284	–	0.2837	0.2408	0.3145	0.283	0.2942
kn	0.2113	0.2966	–	0.1273	0.165	0.2342	0.2313
ml	0.25	0.3228	0.1522	–	0.2226	0.2416	0.2381
mr	0.223	0.349	0.1403	0.1876	–	0.2832	0.2488
ta	0.2731	0.3232	0.241	0.2472	0.2511	–	0.2483
te	0.2506	0.2943	0.1748	0.3543	0.2318	0.2571	–
Pres@1	bn	hi	kn	ml	mr	ta	te
bn	–	0.1866	0.0583	0.1057	0.1311	0.1879	0.1174
hi	0.1881	–	0.1452	0.1474	0.2008	0.1562	0.1469
kn	0.1333	0.1371	–	0.025	0.0588	0.1056	0.0969
ml	0.1157	0.1982	0.025	–	0.1236	0.1379	0.1091
mr	0.1131	0.209	0.0471	0.0745	–	0.1923	0.1176
ta	0.1606	0.1726	0.1417	0.1448	0.1487	–	0.0973
te	0.1326	0.1373	0.0492	0.2318	0.1235	0.1216	–
Pres@5	bn	hi	kn	ml	mr	ta	te
bn	–	0.4642	0.275	0.3886	0.3721	0.4061	0.313
hi	0.3328	–	0.4823	0.2782	0.4655	0.4356	0.4303
kn	0.3083	0.4935	–	0.2	0.2294	0.3472	0.32
ml	0.42	0.4479	0.225	–	0.3255	0.3621	0.3955
mr	0.3967	0.5376	0.1882	0.3127	–	0.3077	0.3765
ta	0.403	0.5137	0.3333	0.3483	0.3641	–	0.4649
te	0.4022	0.4814	0.2815	0.4636	0.3353	0.4135	–

Table 4. Performance of AUX-COMB using grid search on $G(en)$.

MRR	bn	hi	kn	ml	mr	ta	te
bn	–	0.3966	0.3506	0.2926	0.2694	0.3746	0.3134
hi	0.3064	–	0.2999	0.2988	0.3932	0.3565	0.3711
kn	0.3276	0.3752	–	0.2442	0.2616	0.3518	0.3838
ml	0.2808	0.3844	0.2092	–	0.2621	0.2511	0.3254
mr	0.2702	0.4518	0.2328	0.2621	–	0.323	0.3075
ta	0.2962	0.3536	0.2643	0.2591	0.2707	–	0.3013
te	0.3245	0.403	0.3176	0.4235	0.3104	0.3676	–
Pres@1	bn	hi	kn	ml	mr	ta	te
bn	–	0.2567	0.25	0.1543	0.15	0.2333	0.1761
hi	0.197	–	0.1419	0.18	0.283	0.2123	0.2208
kn	0.1917	0.2048	–	0.125	0.1118	0.1944	0.2092
ml	0.1357	0.2339	0.075	–	0.1491	0.1103	0.1409
mr	0.1434	0.3045	0.0765	0.12	–	0.1821	0.1471
ta	0.1636	0.189	0.1333	0.1276	0.1487	–	0.1514
te	0.1935	0.2393	0.1492	0.3045	0.1529	0.227	–
Pres@5	bn	hi	kn	ml	mr	ta	te
bn	–	0.591	0.4583	0.4529	0.4074	0.5909	0.4413
hi	0.4015	–	0.5016	0.4035	0.5002	0.5151	0.546
kn	0.5667	0.5935	–	0.3	0.4824	0.5139	0.6123
ml	0.4643	0.5768	0.275	–	0.3727	0.431	0.6
mr	0.4	0.6368	0.3765	0.42	–	0.5051	0.4784
ta	0.4667	0.5699	0.4222	0.3862	0.4205	–	0.5378
te	0.4717	0.6168	0.5415	0.5727	0.4902	0.5	–

Table 5. Percentage improvement for AUX-COMB using grid search on $G(en)$.

MRR	bn	hi	kn	ml	mr	ta	te
bn	–	24.95	90.34	20.81	10.46	28.16	38.00
hi	7.89	–	5.71	24.09	25.02	25.97	26.14
kn	55.04	26.50	–	91.83	58.55	50.21	65.93
ml	12.32	19.08	37.45	–	17.74	3.93	36.67
mr	21.17	29.46	65.93	39.71	–	14.05	23.59
ta	8.46	9.41	9.67	4.81	7.81	–	21.35
te	29.49	36.94	81.69	19.53	33.91	42.98	–
Pres@1	bn	hi	kn	ml	mr	ta	te
bn	–	37.57	328.82	45.98	14.42	24.16	50.00
hi	4.73	–	-2.27	22.12	40.94	35.92	50.31
kn	43.81	49.38	–	400.00	90.14	84.09	115.89
ml	17.29	18.01	200.00	–	20.63	-20.01	29.15
mr	26.79	45.69	62.42	61.07	–	-5.30	25.09
ta	1.87	9.50	-5.93	-11.88	0.00	–	55.60
te	45.93	74.29	203.25	31.36	23.81	86.68	–
Pres@5	bn	hi	kn	ml	mr	ta	te
bn	–	27.32	66.65	16.55	9.49	45.51	40.99
hi	20.64	–	4.00	45.04	7.45	18.25	26.89
kn	83.81	20.26	–	50.00	110.29	48.01	91.34
ml	10.55	28.78	22.22	–	14.50	19.03	51.71
mr	0.83	18.45	100.05	34.31	–	64.15	27.07
ta	15.81	10.94	26.67	10.88	15.49	–	15.68
te	17.28	28.13	92.36	23.53	46.20	20.92	–

Table 6. Performance of AUX-COMB using EM on $G(en)$.

MRR	bn	hi	kn	ml	mr	ta	te
bn	–	0.3426	0.2424	0.2503	0.2598	0.2936	0.2442
hi	0.2935	–	0.2823	0.2494	0.317	0.2862	0.2952
kn	0.2848	0.2987	–	0.1287	0.1901	0.2394	0.2377
ml	0.2669	0.3365	0.1524	–	0.2349	0.2471	0.2469
mr	0.2366	0.3703	0.1723	0.1969	–	0.2901	0.2694
ta	0.2734	0.3262	0.2458	0.2546	0.2533	–	0.2511
te	0.2601	0.2982	0.1841	0.3579	0.2648	0.2624	–
Pres@1	bn	hi	kn	ml	mr	ta	te
bn	–	0.2164	0.1333	0.1143	0.1344	0.1879	0.1217
hi	0.2	–	0.1403	0.1474	0.2008	0.1562	0.1469
kn	0.1917	0.1371	–	0.025	0.0765	0.1111	0.0969
ml	0.1343	0.2065	0.025	–	0.1309	0.1414	0.1091
mr	0.1197	0.2346	0.0647	0.0745	–	0.1923	0.1353
ta	0.1606	0.1726	0.1472	0.1448	0.1487	–	0.0973
te	0.1391	0.142	0.0508	0.2318	0.151	0.127	–
Pres@5	bn	hi	kn	ml	mr	ta	te
bn	–	0.503	0.35	0.3971	0.4295	0.3939	0.3457
hi	0.3373	–	0.4839	0.2999	0.4709	0.4411	0.4379
kn	0.3917	0.4984	–	0.2	0.3235	0.3472	0.3385
ml	0.4343	0.4752	0.25	–	0.3473	0.3759	0.3909
mr	0.3984	0.5486	0.2824	0.3327	–	0.3564	0.398
ta	0.3939	0.5082	0.3333	0.3586	0.3795	–	0.4784
te	0.4022	0.4862	0.2969	0.4909	0.3804	0.4027	–

Table 7. Percentage improvement for AUX-COMB using EM on $G(en)$.

MRR	bn	hi	kn	ml	mr	ta	te
bn	–	7.94	31.60	3.34	6.52	0.44	7.53
hi	3.35	–	-0.49	3.57	0.79	1.13	0.34
kn	34.78	0.71	–	1.10	15.21	2.22	2.77
ml	6.76	4.24	0.13	–	5.53	2.28	3.70
mr	6.10	6.10	22.81	4.96	–	2.44	8.28
ta	0.11	0.93	1.99	2.99	0.88	–	1.13
te	3.79	1.33	5.32	1.02	14.24	2.06	–
Pres@1	bn	hi	kn	ml	mr	ta	te
bn	–	15.97	128.64	8.14	2.52	0.00	3.66
hi	6.33	–	-3.37	0.00	0.00	0.00	0.00
kn	43.81	0.00	–	0.00	30.10	5.21	0.00
ml	16.08	4.19	0.00	–	5.91	2.54	0.00
mr	5.84	12.25	37.37	0.00	–	0.00	15.05
ta	0.00	0.00	3.88	0.00	0.00	–	0.00
te	4.90	3.42	3.25	0.00	22.27	4.44	–
Pres@5	bn	hi	kn	ml	mr	ta	te
bn	–	8.36	27.27	2.19	15.43	-3.00	10.45
hi	1.35	–	0.33	7.80	1.16	1.26	1.77
kn	27.05	0.99	–	0.00	41.02	0.00	5.78
ml	3.40	6.10	11.11	–	6.70	3.81	-1.16
mr	0.43	2.05	50.05	6.40	–	15.83	5.71
ta	-2.26	-1.07	0.00	2.96	4.23	–	2.90
te	0.00	1.00	5.47	5.89	13.45	-2.61	–

Table 8. Performance of TI+Cue on $G(fr)$.

MRR	bn	hi	kn	ml	mr	ta	te
bn	–	0.3302	0.1634	0.2074	0.1899	0.2383	0.1937
hi	0.263	–	0.2775	0.218	0.2804	0.2719	0.2724
kn	0.1793	0.3007	–	0.1903	0.1715	0.2248	0.2306
ml	0.2163	0.3012	0.1811	–	0.2104	0.2358	0.2082
mr	0.1892	0.3099	0.1718	0.1931	–	0.2671	0.2211
ta	0.2501	0.3132	0.2423	0.2368	0.2689	–	0.2459
te	0.2251	0.2793	0.1747	0.3258	0.1941	0.2276	–
Pres@1	bn	hi	kn	ml	mr	ta	te
bn	–	0.2081	0.0516	0.0881	0.092	0.1289	0.086
hi	0.1596	–	0.1417	0.1168	0.1667	0.1458	0.1241
kn	0.1032	0.1488	–	0.0944	0.0697	0.0907	0.0927
ml	0.103	0.1668	0.0556	–	0.1106	0.1097	0.1022
mr	0.0929	0.1783	0.0758	0.0769	–	0.1548	0.1
ta	0.1421	0.1598	0.1278	0.1357	0.1516	–	0.11
te	0.1035	0.1309	0.0524	0.1891	0.0846	0.0843	–
Pres@5	bn	hi	kn	ml	mr	ta	te
bn	–	0.4707	0.2581	0.3119	0.2858	0.3421	0.2686
hi	0.3384	–	0.455	0.2742	0.4367	0.428	0.4016
kn	0.2452	0.4826	–	0.2889	0.2697	0.3685	0.3244
ml	0.3373	0.4437	0.3111	–	0.3077	0.371	0.3109
mr	0.3222	0.4829	0.2667	0.3192	–	0.3562	0.3423
ta	0.3513	0.5206	0.3611	0.3458	0.3926	–	0.4514
te	0.3663	0.4532	0.2744	0.4652	0.2795	0.39	–

Table 9. Performance of AUX-COMB using grid search on $G(fr)$.

MRR	bn	hi	kn	ml	mr	ta	te
bn	–	0.4375	0.2615	0.2789	0.2347	0.2958	0.247
hi	0.3192	–	0.3411	0.2864	0.3046	0.3222	0.3261
kn	0.2566	0.3604	–	0.3238	0.2278	0.341	0.3334
ml	0.2645	0.3497	0.2813	–	0.2337	0.3025	0.2984
mr	0.2522	0.3923	0.2561	0.237	–	0.2944	0.3021
ta	0.3342	0.3495	0.3028	0.2874	0.2891	–	0.2907
te	0.2711	0.33	0.2782	0.3282	0.2413	0.3097	–
Pres@1	bn	hi	kn	ml	mr	ta	te
bn	–	0.3212	0.1613	0.1366	0.1269	0.1539	0.136
hi	0.1939	–	0.209	0.1898	0.1836	0.1813	0.1844
kn	0.1774	0.1956	–	0.2111	0.1	0.1815	0.1829
ml	0.1224	0.1912	0.1556	–	0.1288	0.1758	0.1609
mr	0.142	0.2548	0.1394	0.099	–	0.1452	0.1769
ta	0.2303	0.1907	0.1796	0.1551	0.1611	–	0.1329
te	0.1593	0.1784	0.1378	0.2043	0.1244	0.1786	–
Pres@5	bn	hi	kn	ml	mr	ta	te
bn	–	0.5899	0.3839	0.4403	0.3462	0.4724	0.3593
hi	0.4626	–	0.4633	0.3615	0.4451	0.4794	0.4832
kn	0.3226	0.5534	–	0.4444	0.3485	0.5463	0.4976
ml	0.4351	0.5166	0.3778	–	0.3308	0.4661	0.4717
mr	0.391	0.5743	0.3576	0.399	–	0.4753	0.4282
ta	0.4263	0.543	0.4426	0.446	0.4036	–	0.4857
te	0.3779	0.5152	0.4256	0.4304	0.3885	0.4743	–

Table 10. Percentage improvement for AUX-COMB using grid search on $G(fr)$.

MRR	bn	hi	kn	ml	mr	ta	te
bn	–	32.50	60.04	34.47	23.59	24.13	27.52
hi	21.37	–	22.92	31.38	8.63	18.50	19.71
kn	43.11	19.85	–	70.15	32.83	51.69	44.58
ml	22.28	16.10	55.33	–	11.07	28.29	43.32
mr	33.30	26.59	49.07	22.73	–	10.22	36.64
ta	33.63	11.59	24.97	21.37	7.51	–	18.22
te	20.44	18.15	59.24	0.74	24.32	36.07	–
Pres@1	bn	hi	kn	ml	mr	ta	te
bn	–	54.35	212.60	55.05	37.93	19.39	58.14
hi	21.49	–	47.49	62.50	10.14	24.35	48.59
kn	71.90	31.45	–	123.62	43.47	100.11	97.30
ml	18.83	14.63	179.86	–	16.46	60.26	57.44
mr	52.85	42.91	83.91	28.74	–	-6.20	76.90
ta	62.07	19.34	40.53	14.30	6.27	–	20.82
te	53.91	36.29	162.98	8.04	47.04	111.86	–
Pres@5	bn	hi	kn	ml	mr	ta	te
bn	–	25.32	48.74	41.17	21.13	38.09	33.77
hi	36.70	–	1.82	31.84	1.92	12.01	20.32
kn	31.57	14.67	–	53.82	29.22	48.25	53.39
ml	28.99	16.43	21.44	–	7.51	25.63	51.72
mr	21.35	18.93	34.08	25.00	–	33.44	25.09
ta	21.35	4.30	22.57	28.98	2.80	–	7.60
te	3.17	13.68	55.10	-7.48	39.00	21.62	–

Table 11. Performance of AUX-COMB using EM on $G(fr)$.

MRR	bn	hi	kn	ml	mr	ta	te
bn	–	0.3493	0.1978	0.2177	0.2177	0.2471	0.2075
hi	0.2635	–	0.2701	0.2319	0.2823	0.2745	0.274
kn	0.2359	0.3031	–	0.201	0.1982	0.2384	0.233
ml	0.2222	0.3062	0.2051	–	0.2207	0.2456	0.2193
mr	0.2248	0.3275	0.1962	0.2093	–	0.2778	0.2547
ta	0.2652	0.3165	0.2442	0.2442	0.2707	–	0.2466
te	0.2366	0.2844	0.1773	0.328	0.2221	0.2336	–
Pres@1	bn	hi	kn	ml	mr	ta	te
bn	–	0.2293	0.0806	0.0955	0.1061	0.1316	0.0965
hi	0.1576	–	0.126	0.1294	0.1667	0.1458	0.1241
kn	0.1419	0.1488	–	0.0944	0.097	0.1093	0.0927
ml	0.1067	0.1668	0.0833	–	0.1173	0.1194	0.1022
mr	0.1179	0.1858	0.0909	0.0856	–	0.1575	0.1192
ta	0.15	0.1598	0.1278	0.1374	0.1516	–	0.11
te	0.1023	0.1364	0.0524	0.1804	0.1141	0.0943	–
Pres@5	bn	hi	kn	ml	mr	ta	te
bn	–	0.4879	0.3161	0.3336	0.3269	0.3539	0.3081
hi	0.3424	–	0.461	0.2952	0.4273	0.428	0.4071
kn	0.3129	0.4862	–	0.2889	0.297	0.3685	0.3305
ml	0.341	0.4509	0.3833	–	0.3231	0.3855	0.3457
mr	0.3524	0.5286	0.3364	0.3423	–	0.3836	0.3833
ta	0.3789	0.5252	0.3611	0.3636	0.3899	–	0.4514
te	0.3942	0.4601	0.278	0.4804	0.3551	0.39	–

Table 12. Percentage improvement for AUX-COMB using EM on $G(fr)$.

MRR	bn	hi	kn	ml	mr	ta	te
bn	–	5.78	21.05	4.97	14.64	3.69	7.12
hi	0.19	–	-2.67	6.38	0.68	0.96	0.59
kn	31.57	0.80	–	5.62	15.57	6.05	1.04
ml	2.73	1.66	13.25	–	4.90	4.16	5.33
mr	18.82	5.68	14.20	8.39	–	4.01	15.20
ta	6.04	1.05	0.78	3.13	0.67	–	0.28
te	5.11	1.83	1.49	0.68	14.43	2.64	–
Pres@1	bn	hi	kn	ml	mr	ta	te
bn	–	10.19	56.20	8.40	15.33	2.09	12.21
hi	-1.25	–	-11.08	10.79	0.00	0.00	0.00
kn	37.50	0.00	–	0.00	39.17	20.51	0.00
ml	3.59	0.00	49.82	–	6.06	8.84	0.00
mr	26.91	4.21	19.92	11.31	–	1.74	19.20
ta	5.56	0.00	0.00	1.25	0.00	–	0.00
te	-1.16	4.20	0.00	-4.60	34.87	11.86	–
Pres@5	bn	hi	kn	ml	mr	ta	te
bn	–	3.65	22.47	6.96	14.38	3.45	14.71
hi	1.18	–	1.32	7.66	-2.15	0.00	1.37
kn	27.61	0.75	–	0.00	10.12	0.00	1.88
ml	1.10	1.62	23.21	–	5.00	3.91	11.19
mr	9.37	9.46	26.13	7.24	–	7.69	11.98
ta	7.86	0.88	0.00	5.15	-0.69	–	0.00
te	7.62	1.52	1.31	3.27	27.05	0.00	–

Table 13. Performance of TI+Cue on $G(en, fr)$.

MRR	bn	hi	kn	ml	mr	ta	te
bn	–	0.3381	0.1815	0.2456	0.243	0.2703	0.2599
hi	0.2832	–	0.3006	0.2951	0.3091	0.2836	0.2974
kn	0.2099	0.3231	–	0.2315	0.1945	0.2531	0.24
ml	0.249	0.3426	0.1891	–	0.2132	0.3181	0.2496
mr	0.2299	0.3609	0.2019	0.1782	–	0.3004	0.2715
ta	0.2697	0.3335	0.2722	0.2949	0.295	–	0.2514
te	0.2756	0.2957	0.1866	0.3683	0.2232	0.2702	–
Pres@1	bn	hi	kn	ml	mr	ta	te
bn	–	0.2081	0.0563	0.1127	0.1287	0.1647	0.1489
hi	0.1806	–	0.1544	0.1864	0.1957	0.158	0.1443
kn	0.1187	0.1632	–	0.1429	0.0813	0.1081	0.0918
ml	0.1197	0.2082	0.0571	–	0.1182	0.2136	0.1158
mr	0.1213	0.2279	0.0875	0.0618	–	0.2083	0.125
ta	0.1559	0.1797	0.1649	0.2	0.1917	–	0.0857
te	0.1489	0.1372	0.059	0.2263	0.1083	0.12	–
Pres@5	bn	hi	kn	ml	mr	ta	te
bn	–	0.479	0.2875	0.3648	0.3738	0.3941	0.3426
hi	0.3371	–	0.514	0.3524	0.4524	0.4478	0.4447
kn	0.2875	0.5228	–	0.3429	0.2875	0.4	0.3475
ml	0.4113	0.4684	0.2714	–	0.2927	0.4636	0.4158
mr	0.4	0.5344	0.2938	0.3109	–	0.3389	0.4292
ta	0.3941	0.542	0.3892	0.3682	0.4083	–	0.4743
te	0.4468	0.4792	0.2902	0.4947	0.3292	0.4429	–

Table 14. Performance of AUX-COMB using grid search on $G(en, fr)$.

MRR	bn	hi	kn	ml	mr	ta	te
bn	–	0.46	0.3493	0.3502	0.3085	0.3826	0.3818
hi	0.3539	–	0.3961	0.3805	0.4165	0.3572	0.4277
kn	0.3219	0.4112	–	0.4221	0.3686	0.3848	0.4674
ml	0.2838	0.4208	0.2869	–	0.2628	0.3778	0.4194
mr	0.3037	0.4896	0.3956	0.3275	–	0.3379	0.3865
ta	0.3766	0.3928	0.3354	0.3406	0.3514	–	0.3655
te	0.3682	0.4087	0.419	0.406	0.3811	0.4198	–
Pres@1	bn	hi	kn	ml	mr	ta	te
bn	–	0.321	0.2125	0.1958	0.1885	0.2382	0.2319
hi	0.2323	–	0.2596	0.258	0.3118	0.1855	0.2698
kn	0.1812	0.2368	–	0.2714	0.1875	0.1973	0.3016
ml	0.1338	0.2718	0.1286	–	0.1345	0.25	0.2526
mr	0.1811	0.3428	0.2625	0.1855	–	0.1778	0.2187
ta	0.2647	0.2174	0.2	0.2045	0.2194	–	0.2029
te	0.2404	0.2395	0.2689	0.2474	0.2479	0.2743	–
Pres@5	bn	hi	kn	ml	mr	ta	te
bn	–	0.6435	0.4875	0.5338	0.4352	0.5706	0.5362
hi	0.5048	–	0.5368	0.4813	0.5209	0.5522	0.6156
kn	0.525	0.6649	–	0.5571	0.625	0.6189	0.6754
ml	0.4606	0.6218	0.4286	–	0.3873	0.5455	0.6632
mr	0.4279	0.7356	0.5687	0.5036	–	0.5667	0.5646
ta	0.4882	0.6159	0.4622	0.5091	0.4833	–	0.5457
te	0.5085	0.6272	0.6197	0.6053	0.575	0.5971	–

Table 15. Percentage improvement for AUX-COMB using grid search on $G(en, fr)$.

MRR	bn	hi	kn	ml	mr	ta	te
bn	–	36.05	92.45	42.59	26.95	41.55	46.90
hi	24.96	–	31.77	28.94	34.75	25.95	43.81
kn	53.36	27.27	–	82.33	89.51	52.03	94.75
ml	13.98	22.83	51.72	–	23.26	18.77	68.03
mr	32.10	35.66	95.94	83.78	–	12.48	42.36
ta	39.64	17.78	23.22	15.50	19.12	–	45.39
te	33.60	38.21	124.54	10.24	70.74	55.37	–
Pres@1	bn	hi	kn	ml	mr	ta	te
bn	–	54.25	277.44	73.74	46.46	44.63	55.74
hi	28.63	–	68.13	38.41	59.33	17.41	86.97
kn	52.65	45.10	–	89.92	130.63	82.52	228.54
ml	11.78	30.55	125.22	–	13.79	17.04	118.13
mr	49.30	50.42	200.00	200.16	–	-14.64	74.96
ta	69.79	20.98	21.29	2.25	14.45	–	136.76
te	61.45	74.56	355.76	9.32	128.90	128.58	–
Pres@5	bn	hi	kn	ml	mr	ta	te
bn	–	34.34	69.57	46.33	16.43	44.79	56.51
hi	49.75	–	4.44	36.58	15.14	23.31	38.43
kn	82.61	27.18	–	62.47	117.39	54.73	94.36
ml	11.99	32.75	57.92	–	32.32	17.67	59.50
mr	6.98	37.65	93.57	61.98	–	67.22	31.55
ta	23.88	13.63	18.76	38.27	18.37	–	15.05
te	13.81	30.88	113.54	22.36	74.67	34.82	–

Table 16. Performance of AUX-COMB using EM on $G(en, fr)$.

MRR	bn	hi	kn	ml	mr	ta	te
bn	–	0.4126	0.3179	0.2631	0.3067	0.2765	0.3253
hi	0.3136	–	0.2954	0.3128	0.3437	0.2927	0.2979
kn	0.3512	0.326	–	0.272	0.3251	0.291	0.2566
ml	0.2612	0.3553	0.2393	–	0.2187	0.3329	0.2608
mr	0.2971	0.412	0.3402	0.2043	–	0.3133	0.3535
ta	0.2761	0.3404	0.2827	0.3147	0.3002	–	0.2514
te	0.3535	0.3018	0.2019	0.3732	0.2821	0.2701	–
Pres@1	bn	hi	kn	ml	mr	ta	te
bn	–	0.2823	0.1875	0.1239	0.182	0.1647	0.1979
hi	0.2032	–	0.1439	0.2121	0.2299	0.1594	0.1433
kn	0.2313	0.1632	–	0.1571	0.1875	0.1595	0.1
ml	0.1268	0.2169	0.1	–	0.1182	0.2318	0.1158
mr	0.1795	0.2721	0.1812	0.0782	–	0.2083	0.2
ta	0.1559	0.1797	0.1757	0.2045	0.1917	–	0.0857
te	0.2277	0.1452	0.0623	0.2158	0.1562	0.12	–
Pres@5	bn	hi	kn	ml	mr	ta	te
bn	–	0.5758	0.4625	0.3972	0.4615	0.4235	0.4617
hi	0.4129	–	0.5123	0.3639	0.4676	0.4652	0.4506
kn	0.575	0.5263	–	0.3857	0.5062	0.4622	0.3984
ml	0.4338	0.5145	0.4286	–	0.3127	0.4636	0.4368
mr	0.4344	0.6223	0.55	0.3527	–	0.3944	0.5083
ta	0.4088	0.5522	0.3784	0.4545	0.4333	–	0.4743
te	0.4936	0.4812	0.3262	0.5316	0.4812	0.4429	–

Table 17. Percentage improvement for AUX-COMB using EM on $G(en, fr)$.

MRR	bn	hi	kn	ml	mr	ta	te
bn	–	22.03	75.15	7.13	26.21	2.29	25.16
hi	10.73	–	-1.73	6.00	11.19	3.21	0.17
kn	67.32	0.90	–	17.49	67.15	14.97	6.92
ml	4.90	3.71	26.55	–	2.58	4.65	4.49
mr	29.23	14.16	68.50	14.65	–	4.29	30.20
ta	2.37	2.07	3.86	6.71	1.76	–	0.00
te	28.27	2.06	8.20	1.33	26.39	-0.04	–
Pres@1	bn	hi	kn	ml	mr	ta	te
bn	–	35.66	233.04	9.94	41.41	0.00	32.91
hi	12.51	–	-6.80	13.79	17.48	0.89	-0.69
kn	94.86	0.00	–	9.94	130.63	47.55	8.93
ml	5.93	4.18	75.13	–	0.00	8.52	0.00
mr	47.98	19.39	107.09	26.54	–	0.00	60.00
ta	0.00	0.00	6.55	2.25	0.00	–	0.00
te	52.92	5.83	5.59	-4.64	44.23	0.00	–
Pres@5	bn	hi	kn	ml	mr	ta	te
bn	–	20.21	60.87	8.88	23.46	7.46	34.76
hi	22.49	–	-0.33	3.26	3.36	3.89	1.33
kn	100.00	0.67	–	12.48	76.07	15.55	14.65
ml	5.47	9.84	57.92	–	6.83	0.00	5.05
mr	8.60	16.45	87.20	13.44	–	16.38	18.43
ta	3.73	1.88	-2.77	23.44	6.12	–	0.00
te	10.47	0.42	12.41	7.46	46.17	0.00	–

Table 18. Performance of TI+Cue on $G(hi)$.

MRR	bn	kn	ml	mr	ta	te
bn	–	0.1248	0.1829	0.1913	0.2207	0.2002
kn	0.1524	–	0.1828	0.1554	0.1809	0.2126
ml	0.194	0.1341	–	0.1892	0.2064	0.1827
mr	0.1809	0.1319	0.1675	–	0.2434	0.2071
ta	0.2339	0.1947	0.2128	0.2271	–	0.2158
te	0.226	0.177	0.2414	0.1824	0.1977	–
Pres@1	bn	kn	ml	mr	ta	te
bn	–	0.0367	0.077	0.0955	0.1148	0.0902
kn	0.0796	–	0.0875	0.0684	0.059	0.0816
ml	0.083	0.045	–	0.0899	0.0938	0.084
mr	0.0926	0.0404	0.0662	–	0.1261	0.0968
ta	0.1259	0.0795	0.1045	0.1082	–	0.0838
te	0.1167	0.0592	0.1231	0.0782	0.0717	–
Pres@5	bn	kn	ml	mr	ta	te
bn	–	0.1898	0.2703	0.2917	0.3193	0.2784
kn	0.2286	–	0.285	0.2228	0.3064	0.3204
ml	0.3327	0.2025	–	0.3036	0.3009	0.277
mr	0.2917	0.2246	0.2734	–	0.3218	0.3117
ta	0.3452	0.3218	0.3429	0.3415	–	0.3808
te	0.3324	0.2724	0.3874	0.2633	0.3212	–

Table 19. Performance of AUX-COMB using grid search on $G(hi)$.

MRR	bn	kn	ml	mr	ta	te
bn	–	0.2019	0.2306	0.236	0.2732	0.2509
kn	0.1978	–	0.2462	0.1847	0.2534	0.3284
ml	0.2177	0.2311	–	0.2376	0.2974	0.2435
mr	0.2083	0.2257	0.2238	–	0.3024	0.2854
ta	0.28	0.2417	0.2418	0.2713	–	0.2913
te	0.2728	0.3129	0.2887	0.2799	0.3063	–
Pres@1	bn	kn	ml	mr	ta	te
bn	–	0.1122	0.1048	0.1339	0.1526	0.1304
kn	0.1245	–	0.145	0.0702	0.1192	0.198
ml	0.0976	0.1025	–	0.1396	0.1679	0.1273
mr	0.1025	0.1175	0.1165	–	0.1672	0.166
ta	0.1659	0.0872	0.1179	0.1429	–	0.1364
te	0.1667	0.1806	0.1777	0.152	0.1626	–
Pres@5	bn	kn	ml	mr	ta	te
bn	–	0.2816	0.3412	0.3277	0.4015	0.3971
kn	0.249	–	0.3575	0.3158	0.3897	0.4531
ml	0.3521	0.38	–	0.3432	0.4536	0.3886
mr	0.3202	0.3123	0.3252	–	0.4706	0.3883
ta	0.4126	0.4115	0.3723	0.4285	–	0.4798
te	0.3667	0.4816	0.4097	0.4314	0.4707	–

Table 20. Percentage improvement for AUX-COMB using grid search on $G(hi)$.

MRR	bn	kn	ml	mr	ta	te
bn	–	61.78	26.08	23.37	23.79	25.32
kn	29.79	–	34.68	18.85	40.08	54.47
ml	12.22	72.33	–	25.58	44.09	33.28
mr	15.15	71.11	33.61	–	24.24	37.81
ta	19.71	24.14	13.63	19.46	–	34.99
te	20.71	76.78	19.59	53.45	54.93	–
Pres@1	bn	kn	ml	mr	ta	te
bn	–	205.72	36.10	40.21	32.93	44.57
kn	56.41	–	65.71	2.63	102.03	142.65
ml	17.59	127.78	–	55.28	79.00	51.55
mr	10.69	190.84	75.98	–	32.59	71.49
ta	31.77	9.69	12.82	32.07	–	62.77
te	42.84	205.07	44.35	94.37	126.78	–
Pres@5	bn	kn	ml	mr	ta	te
bn	–	48.37	26.23	12.34	25.74	42.64
kn	8.92	–	25.44	41.74	27.19	41.42
ml	5.83	87.65	–	13.04	50.75	40.29
mr	9.77	39.05	18.95	–	46.24	24.57
ta	19.52	27.87	8.57	25.48	–	26.00
te	10.32	76.80	5.76	63.84	46.54	–

Table 21. Performance of TI+Cue on $G(en, fr, hi)$.

MRR	bn	kn	ml	mr	ta	te
bn	–	0.1712	0.2565	0.2527	0.3201	0.2395
kn	0.1925	–	0.1817	0.2138	0.2259	0.2823
ml	0.2764	0.1646	–	0.2342	0.3394	0.278
mr	0.2376	0.1668	0.1907	–	0.2868	0.2708
ta	0.3271	0.2542	0.3228	0.2716	–	0.2707
te	0.2584	0.2145	0.3889	0.2332	0.2877	–
Pres@1	bn	kn	ml	mr	ta	te
bn	–	0.0391	0.1235	0.136	0.2139	0.122
kn	0.1043	–	0.0833	0.113	0.1	0.1268
ml	0.1309	0.0583	–	0.1362	0.224	0.1208
mr	0.1296	0.0609	0.0724	–	0.175	0.1302
ta	0.1889	0.1429	0.212	0.1611	–	0.0914
te	0.126	0.075	0.2375	0.1151	0.1143	–
Pres@5	bn	kn	ml	mr	ta	te
bn	–	0.2913	0.3897	0.3912	0.4389	0.334
kn	0.2783	–	0.275	0.287	0.3343	0.4125
ml	0.4926	0.2333	–	0.3172	0.48	0.4875
mr	0.4096	0.2478	0.3276	–	0.3583	0.417
ta	0.4944	0.38	0.444	0.3778	–	0.5229
te	0.436	0.3554	0.5167	0.3453	0.5	–

Table 22. Performance of AUX-COMB using grid search on $G(en, fr, hi)$.

MRR	bn	kn	ml	mr	ta	te
bn	–	0.3256	0.3842	0.2976	0.4329	0.3753
kn	0.2701	–	0.3629	0.3636	0.3878	0.5013
ml	0.312	0.3369	–	0.3354	0.4069	0.4327
mr	0.3066	0.3582	0.338	–	0.4162	0.4136
ta	0.4491	0.3306	0.3739	0.3383	–	0.508
te	0.3407	0.4454	0.4534	0.3932	0.4336	–
Pres@1	bn	kn	ml	mr	ta	te
bn	–	0.2174	0.2441	0.1784	0.2611	0.238
kn	0.1565	–	0.2	0.2043	0.2143	0.3375
ml	0.1471	0.1667	–	0.2103	0.284	0.2458
mr	0.1912	0.2217	0.1879	–	0.2611	0.2453
ta	0.3056	0.1857	0.236	0.2139	–	0.3371
te	0.216	0.3036	0.3042	0.2377	0.2657	–
Pres@5	bn	kn	ml	mr	ta	te
bn	–	0.4348	0.5471	0.4384	0.6556	0.53
kn	0.3913	–	0.5417	0.5739	0.6229	0.7179
ml	0.5191	0.5667	–	0.4966	0.556	0.6917
mr	0.4312	0.5435	0.5345	–	0.5917	0.6
ta	0.6417	0.4971	0.52	0.4889	–	0.7543
te	0.482	0.6304	0.6	0.6	0.6314	–

Table 23. Percentage improvement for AUX-COMB using grid search on $G(en, fr, hi)$.

MRR	bn	kn	ml	mr	ta	te
bn	–	90.19	49.79	17.77	35.24	56.70
kn	40.31	–	99.72	70.07	71.67	77.58
ml	12.88	104.68	–	43.21	19.89	55.65
mr	29.04	114.75	77.24	–	45.12	52.73
ta	37.30	30.06	15.83	24.56	–	87.66
te	31.85	107.65	16.59	68.61	50.71	–
Pres@1	bn	kn	ml	mr	ta	te
bn	–	456.01	97.65	31.18	22.07	95.08
kn	50.05	–	140.10	80.80	114.30	166.17
ml	12.38	185.93	–	54.41	26.79	103.48
mr	47.53	264.04	159.53	–	49.20	88.40
ta	61.78	29.95	11.32	32.77	–	268.82
te	71.43	304.80	28.08	106.52	132.46	–
Pres@5	bn	kn	ml	mr	ta	te
bn	–	49.26	40.39	12.07	49.37	58.68
kn	40.60	–	96.98	99.97	86.33	74.04
ml	5.38	142.91	–	56.56	15.83	41.89
mr	5.27	119.33	63.16	–	65.14	43.88
ta	29.79	30.82	17.12	29.41	–	44.25
te	10.55	77.38	16.12	73.76	26.28	–