Lost in Morphology
Enhancing Bilingual Lexicon Induction with Lemmatisation

Michaela Denisová and Pavel Rychlý

Natural Language Processing Centre
Faculty of Informatics, Masaryk University
Botanická 68a, 602 00 Brno, Czech Republic
{449884,pary}@mail.muni.cz

Abstract. Bilingual Lexicon Induction (BLI) is a fundamental task in cross-lingual word embedding (CWE) evaluation, aimed at retrieving word translations from monolingual corpora in two languages. However, morphological complexity poses an intractable challenge, where translations deemed incorrect tend to be morphological variations of the correct ones. This study explores the role of lemmatisation in mitigating this issue by comparing two integration strategies: (1) pre-alignment lemmatisation, applied before training monolingual word embeddings (MWEs), and (2) post-retrieval lemmatisation, applied to retrieved target words. We conduct experiments using three state-of-the-art CWEs across a wide range of language pairs, comparing Slavonic and other language families, with varying morphological complexity. Our findings reveal notable differences between the two approaches: post-retrieval lemmatisation proves more beneficial for less morphologically complex language pairs, while pre-alignment lemmatisation performs well for those with moderate complexity, and for highly inflected languages, the choice of approach has minimal impact.

Keywords: Bilingual lexicon induction, Morphology, Lemmatisation, Evaluation.

1 Introduction

Bilingual Lexicon Induction (BLI) is an intrinsic evaluation task designed to identify and extract translations of individual words. This task has been widely adopted as a method for evaluating cross-lingual embeddings (CWEs), which aim to align two (or more) sets of independently trained monolingual word embeddings (MWEs) into a shared cross-lingual space, where semantically similar words are represented by closely aligned vectors. [19]

Given this characteristic, they have shown significant advantages in numerous NLP applications, such as machine translation [9,27,24], cross-lingual information retrieval [23], and language acquisition and learning [26].

One of the challenges in this task arises from the morphological complexity of languages. Many languages exhibit a certain degree of morphology richness,

A. Horák, P. Rychlý, A. Rambousek (eds.): Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2025, pp. 3–14, 2025.

Table 1: Three randomly selected examples of incorrect translations using Muse for the German-English language pair. The correct lemmas are highlighted in **bold**.

Source word	Top-5 nearest neighbours				
frieren	freezing	melt	freezes	thawing	**freeze**
laufen	running	swim	**run**	smoothly	going
schatten	shadows	**shadow**	silhouetted	glare	shadowed

where inflected word forms can be more frequent in the corpus than their corresponding base forms (lemmas). This poses a potential limitation for BLI since certain CWEs tend to favour inflected yet valid word forms over their respective lemmas. Table 1 illustrates this, showing how the Muse model retrieves inflected word forms before their lemmas for three German words. For example, given the German source word *frieren*, Muse first translates it to the morphologically related form *freezing* rather than the lemma *freeze*. According to the prior research, these cases are not always accounted for in gold-standard evaluation dictionaries, creating bias in evaluation. [13,14,8]

To address this issue, we propose integrating lemmatisation — the process of converting word forms into their respective lemmas — into the BLI pipeline in two ways. Since CWEs are rooted in MWEs [19], the first approach is to lemmatise the corpus used for MWEs' training, ensuring that CWEs are trained using a normalised lexical space. This method could lead to more stable alignments, as embeddings would represent lemmas rather than inflected variants. Alternatively, lemmatisation can be applied to the retrieved target words, enhancing the output by standardising inflected word forms after alignment. While the former approach ensures consistency in the embedding space, the latter provides flexibility in handling inflected word forms post hoc.

In this paper, we present the first empirical investigation of lemmatisation in BLI, focusing on which of these two integration strategies yields better BLI performance. We systematically compare the impact of pre-alignment and post-retrieval lemmatisation using three state-of-the-art CWEs with different degrees of supervision evaluated across a wide variety of language pairs, exhibiting different levels of morphology complexity, ranging from less morphologically complex word pairs such as English-French (en-fr), Italian-English (it-en), through medium morphologically rich, namely Estonian-Slovak (et-sk) and Czech-Slovak (cs-sk), to the most morphologically compound as German-Hungarian (de-hu) language pair. [20] We aim to determine the most effective method for enhancing BLI quality for languages with different degrees of morphological complexity. Moreover, we manually analyse the effects of lemmatisation on translation accuracy and semantic consistency.

Our contribution is threefold.

1. We present the first empirical comparison of the impact of pre-alignment and post-retrieval lemmatisation on BLI performance, assessing their effectiveness across different CWEs with varying degrees of supervision.

2. We conduct extensive experiments across multiple language pairs (en-fr, it-en, et-sk, cs-sk, de-hu), covering a spectrum of morphological complexity, to determine the role of morphology in BLI performance.
3. We provide a detailed manual error analysis, exploring the effects of lemmatisation as one of the tools for morphology-aware BLI and offering insights into its advantages and limitations.

This paper is structured as follows. In Section 2, we explain the background of the BLI task. In Section 3, we present the different aspects of the setup of our experiments in detail. In Section 4, we evaluate the baseline models exploiting two approaches of integrating lemmatisation, provide manual error analysis and discuss the results. In Section 5, we offer concluding remarks.

2 Background

The pioneering work introducing the embedding-based method evaluated on the BLI task was proposed by Mikolov et al. (2013) [16], spawning an immense number of articles continuing in the research. These proposed methods ranged from classical, state-of-the-art baseline models such as Muse [5], VecMap [3,2], RCLS [12] to current endeavours in integrating contextual embeddings and Large Language Models (LLMs) [22,15,11].

The morphological variations have been challenging in the BLI task. [19,25] Some methods offered to incorporate sub-word information (words' stems) to generate translations for less common word forms. Moreover, Resiandi (2023) [18] implemented character-level embeddings to capture morphological variations. From the earlier efforts, Yang et al. (2019) [25] introduced a morphology-aware alignment that ensures corresponding word forms align together.

However, there have only been a few discussions on the morphology involved in the evaluation. The reflection on models finding viable inflected word forms missing from the evaluation datasets and creating bias was first expressed in Kementchedjhieva et al. (2019) [13] Moreover, Czarnowska et al. (2019) [6] tried to address this issue, creating morphologically complete evaluation datasets for 5 Slavic (Polish, Czech, Russian, Slovak, and Ukrainian) and 5 Romance (French, Spanish, Italian, Portuguese, and Catalan) languages. This approach has faced criticism for its limited selection of language pairs and reliance solely on the top-1 retrieved target word candidate for evaluation. [14]

3 Experimental Setup

In this Section, we introduce four key aspects of this experiment: the evaluation and training datasets used, training details of the CWE models, the evaluation metrics employed during the evaluation, and the integrated lemmatisers.

3.1 Data

Training data. To train CWEs, we utilised pre-trained Sketch Engine MWEs [10] for French, English, Italian, Estonian, Slovak, Czech, German, and Hungarian. [1] All these embeddings were trained using the SkipGram model with dimension 128 across the web corpora. Embeddings for each language occur in two versions: a model calculated from the word forms using a raw corpus (for post-retrieval lemmatisation) and a model calculated from lemmas using a corpus with every word converted to its base form [10] (for pre-alignment lemmatisation).

To train CWEs in supervised mode, we employed Muse seed lexicons [5] for English-French and Italian-English, seed lexicons for German-Hungarian from Anastasopoulos and Neubig (2020). [1] For Estonian-Slovak, we constructed a seed lexicon from the Estonian-Slovak database published in Denisová (2021) [7], and for Czech-Slovak, we manually compiled a seed lexicon consisting of words that are similar in both languages. All of these seed lexicons contain around 5K source words.

Evaluation data. The evaluation data originated from sources similar to the seed lexicons described above. We evaluated English-French and Italian-English using the Muse dataset [5], German-Hungarian employing a dataset from Anastasopoulos and Neubig (2020) [1], Estonian-Slovak utilising a database from Denisová (2021) [7], and for Czech-Slovak, we manually compiled a dataset containing words that differ in both languages. All of the word pairs have been converted into their lemma forms and contain around 800 to 1.5K source words.

3.2 Training details of CWEs

For our comparison, we selected three state-of-the-art CWE methods, Muse, VecMap (VM), and RCLS. All three models are trained in a supervised mode (Muse-S, VM-S, RCLS), while only Muse and VM are in an unsupervised (Muse-U, VM-U) mode and a mode that relies on identical strings and numerals occurring in MWEs (Muse-I, VM-I).

The default settings closely followed the Muse training described in [5], RCLS setting in [12], and VM-S and VM-I in [2], and VM-U settings in [3]. The results are computed from the first 300K aligned embeddings.

3.3 Metrics

The most common evaluation metric in the BLI task is **precision** at k (P@k), which computes the ratio of true positives (TP) to the sum of true positives and false positives (FP). In other words, it is the ratio of the positive target words to the total number of target words found by the model (positive and negative). In this case, k represents the top target word candidates extracted for each source word. In this paper, $k = 1$.

In addition to that, we report on **recall** (R), which is calculated using the standard formula; later, we also refer to R@1, i.e., recall at rank 1.

3.4 Lemmatisers

Pre-alignment. The corpora are lemmatised using pipelines which Lexical Computing integrates when building corpora for Sketch Engine. [10]

Post-retrieval. We used four lemmatisers for four target languages, i.e., French, English, Slovak and Hungarian.

For French and Hungarian, we employed the Lefff [21] and HuSpaCy [17] lemmatisers, respectively. [2] For English, we utilised a lemmatiser integrated in the Python library NLTK. [4] For Slovak, we applied the lemmatiser Majka. [3]

4 Evaluation

Overall results comparing the effects of pre-alignment and post-retrieval lemmatisation across all language pairs, showing P@1 and R@1 are displayed in Tables 2 and 3, respectively.

Table 2: P@1 across models and language pairs using pre-alignment lemmatisation (pre) and post-retrieval lemmatisation (post).

P@1	it-en		en-fr		et-sk		cs-sk		de-hu	
	pre	post	pre	post	pre	post	pre	post	pre	post
Muse-S	50.41	**62.26**	44.26	**45.76**	36.93	**38.56**	**87.48**	84.91	**27.88**	23.11
Muse-I	49.58	**61.59**	44.66	**45.76**	37.09	**39.11**	**87.48**	84.91	22.61	**25.87**
Muse-U	49.27	**62.19**	44.5	**45.02**	35.87	**35.93**	**87.57**	85.26	23.86	**24.37**
VM-S	50.1	**61.29**	**45.94**	45.91	40.74	**42.92**	**87.74**	85.78	**29.14**	28.89
VM-I	46.16	**59.65**	**47.47**	45.98	38.38	**39.74**	**88.27**	85.44	21.6	**23.86**
VM-U	46.57	**59.28**	44.26	**45.61**	**38.61**	38.56	**88.01**	85.44	21.6	**23.61**
RCLS	51.34	**65.02**	45.61	**48.04**	**39.75**	38.2	**87.66**	84.91	**24.87**	22.86

When looking at Tables 2 and 3, we can observe that post-retrieval lemmatisation generally leads to better results in languages with low morphological complexity. This is especially evident in the Italian-English language pair, where the performance gap between pre-alignment and post-retrieval approaches consistently falls within a margin of approximately 10% to 20%. Furthermore, the

[2] Both published under the Apache License (https://www.apache.org/licenses/LICENSE-2.0).

[3] Published under the Creative Commons Attribution-ShareAlike 3.0 Unported License (https://creativecommons.org/licenses/by-sa/3.0/), GNU AFFERO GENERAL PUBLIC LICENSE, version 3 (https://www.gnu.org/licenses/agpl-3.0.html), and GNU Free Documentation License (https://www.gnu.org/licenses/fdl-1.3.html).

Table 3: R@1 across models and language pairs using pre-alignment lemmatisation (pre) and post-retrieval lemmatisation (post).

R@1	it-en		en-fr		et-sk		cs-sk		de-hu	
	pre	post	pre	post	pre	post	pre	post	pre	post
Muse-S	24.55	**42.19**	36.8	**41.4**	**32.33**	28.33	**67.26**	65.06	**12.86**	10.66
Muse-I	24.15	**41.73**	37.13	**41.4**	**32.46**	28.73	**67.26**	65.06	10.42	**11.93**
Muse-U	24.0	**42.14**	37.0	**40.73**	**31.4**	26.4	**67.33**	65.33	11.0	**11.23**
VM-S	24.4	**41.53**	38.13	**41.53**	**35.66**	31.53	**67.46**	65.73	**13.44**	13.32
VM-I	22.48	**40.42**	39.46	**41.6**	**33.6**	29.2	**67.86**	65.46	9.96	**11.0**
VM-U	22.68	**40.17**	36.8	**41.26**	**33.8**	28.33	**67.66**	65.46	9.96	**10.89**
RCLS	25.01	**44.06**	41.26	**43.46**	**34.8**	28.06	**67.4**	65.06	**11.47**	10.54

post-retrieval lemmatisation enhanced outcomes for morphologically complex languages in unsupervised settings.

On the other hand, for languages with mid to high morphological complexity, the pre-alignment lemmatisation surpassed the post-retrieval, particularly in scenarios involving supervision. In the Estonian-Slovak language pair, the pre-alignment lemmatisation performed well when calculating recall over precision.

To assess the reliability of these observed differences, we also conducted paired t-tests for statistical significance between pre-alignment (A) and post-retrieval (B) approaches. For each (*language* × *model*) pair, we computed the mean difference in P@1 (Δ = A – B) over 7 random source words. We estimated a 95% confidence interval CI by 1,000-iteration paired bootstrapping. Table 4 presents the results. Negative values indicate a drop in performance for the pre-alignment lemmatisation; positive values indicate an improvement. Statistically significant values are marked in bold, where the threshold is $p < 0.05$.

Given Table 4, most comparisons are statistically significant, favouring post-retrieval lemmatisation for morphologically low-complexity language pairs such as Italian–English and English–French, and pre-alignment lemmatisation for mid-complexity pairs such as Czech–Slovak and Estonian–Slovak. The exception is the most morphologically rich pair, German-Hungarian, where model performance on P@1 is largely statistically indistinguishable, and observed differences may be due to sampling noise.

Interestingly, in the Estonian–Slovak language pair, the statistical significance test consistently indicates better performance for the pre-alignment approach, despite the post-retrieval approach achieving higher P@1 scores in the majority of individual runs. This reflects a common pattern in significance testing: frequent but marginal wins may be outweighed by infrequent but stronger gains, leading to statistical significance in the opposite direction. It underscores why paired tests that account for both the direction and magnitude of differences are crucial for fair approach comparison.

These results suggest that languages with low morphology levels, like Italian-English and English-French, benefit from post-retrieval lemmatisation,

Table 4: Paired t-tests for P@1 (Δ = A – B) between pre-alignment (A) and post-retrieval (B) lemmatisation approaches. Each Δ is the average over 7 random-seed runs; the 95% confidence interval CI is computed from 1,000 paired bootstrap resamples of those seven differences. Values marked in **bold** are statistically significant, where $p < 0.05$.

P@1	it-en		en-fr		et-sk		cs-sk		de-hu	
	CI	Δ	CI	Δ	CI	Δ	CI	Δ	CI	Δ
MUSE-S	**[-0.074, -0.020]**	-0.046	**[-0.047, -0.006]**	-0.026	**[0.007, 0.054]**	+0.030	**[0.009, 0.047]**	+0.029	**[0.008, 0.089]**	+0.047
MUSE-I	**[-0.068, -0.016]**	-0.041	**[-0.039, -0.003]**	-0.022	[-0.003, 0.051]	+0.023	**[0.009, 0.047]**	+0.029	[-0.058, 0.013]	-0.024
MUSE-U	**[-0.075, -0.023]**	-0.050	**[-0.038, -0.002]**	-0.020	**[0.015, 0.066]**	+0.040	**[0.007, 0.045]**	+0.026	[-0.037, 0.034]	-0.000
VM-S	**[-0.068, -0.014]**	-0.039	**[-0.048, -0.002]**	-0.024	[-0.003, 0.047]	+0.022	**[0.005, 0.039]**	+0.023	[-0.034, 0.042]	+0.003
VM-I	**[-0.087, -0.035]**	-0.060	[-0.026, 0.010]	-0.008	**[0.005, 0.054]**	+0.029	**[0.012, 0.047]**	+0.030	[-0.047, 0.018]	-0.016
VM-U	**[-0.076, -0.025]**	-0.051	**[-0.052, -0.013]**	-0.032	**[0.014, 0.066]**	+0.040	**[0.010, 0.046]**	+0.028	[-0.045, 0.021]	-0.013
RCLS	**[-0.092, -0.039]**	-0.065	**[-0.104, -0.057]**	-0.080	**[0.047, 0.096]**	+0.071	**[0.011, 0.049]**	+0.031	[-0.013, 0.065]	+0.024

while pre-alignment lemmatisation boosts performance for mid-complex language pairs, such as Czech-Slovak and Estonian-Slovak. For the most morphologically rich language pairs, such as German-Hungarian, the choice of approach appears to have little impact, with both approaches equally boosting performance, and no statistically significant advantage is observed for either approach.

4.1 Error Analysis

In this part, to assess how our conclusions manifest in practice, we examined the retrieved target word that caused gaps between the pre-alignment and post-retrieval approach performances. Specifically, we looked at the differing targets retrieved by the models MUSE and VM trained in a supervised setting using the English-French, Czech-Slovak and German-Hungarian language pairs. The categorised errors are shown in Table 5. On top of that, Fig. 1 displays the share of errors in each category per language pair.

Given Table 5, the most common error for all three language pairs, in both approaches, was that the model suggested a different but still valid synonym than their counterpart (Type A). Sometimes the retrieved target-word candidate was absent from the evaluation dataset, e.g., *bojko*, or, in the case of the English–French pre-alignment approach, was archaic, e.g., *mégarde*. Another frequent source of error for both approaches in all languages involved selecting a target-word candidate with a lexical-semantic relationship rather than a direct translation (Type D), such as *tyler – jesse*, *dave*.

The errors typical for the post-retrieval approach in the Czech-Slovak language combination were caused by bad lemmatisation, where the lemmatiser

Table 5: Examples of the retrieved targets for source words (SRC) using pre-alignment and post-retrieval lemmatisation, showing types of errors in both approaches.

Type	SRC	Pre-alignment	Post-retrieval	EN	Explanation
A	strašpytel	strachopud	bojko	coward	synonym
	carpenter	menusier	charpentier	-	
	accidentally	mégarde	accidentellement	-	
	exzellent	elsőrangú	kiváló	excellent	
B	luštěnina	strukovina	-	legume	low frequency
	později	-	neskôr	later	
	wednesday	-	aprèm (eng. *afternoon*)	-	
	zimbabwe	-	-	-	
	zerstörer	-	-	destroyer	
C	fandit	fandiť	fandiť*	cheer on	bad
	nechat	nechať	nenechať (negation)	let	lemmatisation
	atomic	atomique	atomique**s**	-	
	admission	admission**s**	admission	-	
D	klokan	kengura	slon (eng. *elephant*)	kangaroo	related words
	patnáctý	pätnásty	dvanásty (eng. *twelfth*)	fifteenth	
	tyler	jesse	dave	-	
	alpha	alpha	béta	-	
E	léto	rok (eng. *year*)	leto	summer	polysemy
	pole	mât (eng. *bar*)	pôle	north pole vs ski pole	
F	advocacy	advocacy	plaidoyer	-	English words
	dream	dreams	yesterday	-	
G	anfangen	kedz	**el**kedz	to start	mixed affixes
	beratung	tanácsadás	tanácsadás**i**	advice	
	lernen	**meg**tanul	tanul	to learn	

added a negation or asterisk symbol to the target (Type C), e.g., *fandiť**, *nenechať*. In the English–French case, such issues occurred in both approaches, with plurals being particularly problematic, e.g., *atomiques*, *admissions*. Moreover, the lemmas of some words which typically occur in the inflected form rather than lemma in the corpora became out-of-vocabulary words due to the low frequency (Type B), e.g., *luštěnina* (more common luštěniny, eng. *legumes*). Similarly, the pre-alignment approach struggled with various out-of-vocabulary words that were excluded due to their low frequency, such as *později*, *wednesday*, *zerstörer*.

Less frequent errors in English-French and German-Hungarian involved the retrieval of mixed-language words, particularly from English (Type F), e.g., *advocacy*, *dream*. Furthermore, another error category (Type E) stems from

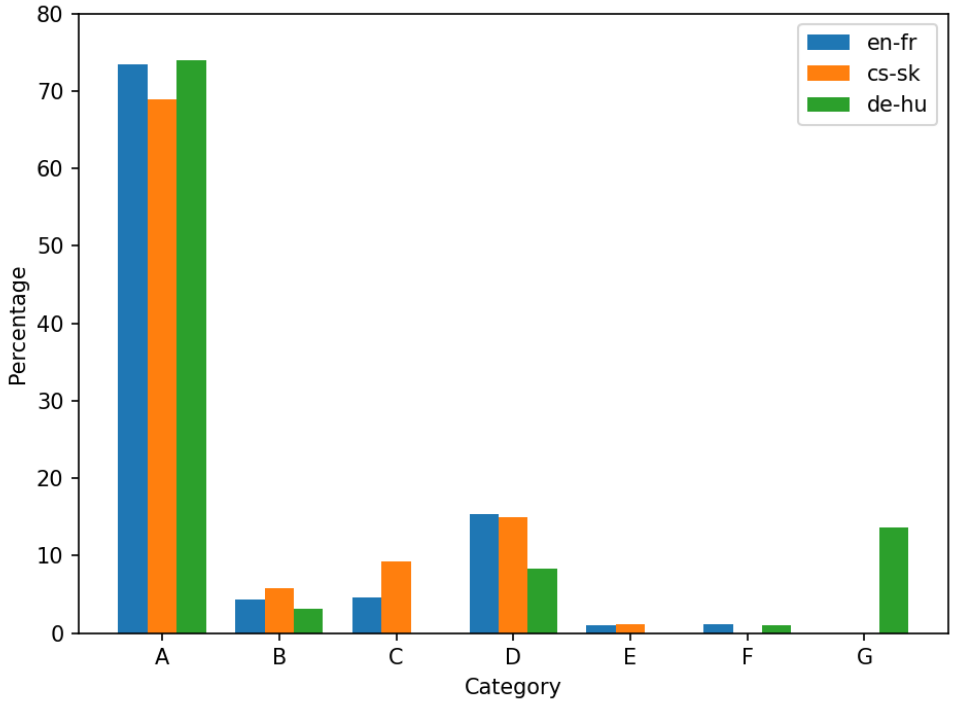

Fig. 1: The percentage of incorrect targets in each category type per language pair.

polysemy and sense ambiguity, where the system retrieves a valid target word corresponding to an unintended sense of the source word, missing from the evaluation dataset. For example, the English word *pole* could mean either *a long, thin stick* or *either of the two points at the most northern and most southern ends of the earth*. These errors are often amplified by lemmatisation, which collapses multiple inflected or context-specific forms into a single lemma, obscuring distinctions between senses. However, these cases were rare and observed only in the English–French and Czech–Slovak language pairs.

The majority of errors in German-Hungarian were caused by their complex morphology, often involving prefixes or suffixes, e.g., *tanul* (to learn something) and *meg-tanul* (to learn and complete the learning). While the base meaning is preserved, derivational or aspectual morphology causes mismatches with entries from the evaluation dataset. These errors are challenging for lemmatisation tools, since many affixed forms in Hungarian function as independent lemmas themselves, making it difficult to distinguish between inflectional variants and derivationally distinct but semantically related words.

5 Conclusion

In this paper, we have explored the impact of lemmatisation on BLI by comparing two integration strategies: pre-alignment and post-retrieval lemmatisation.

We have conducted an extensive evaluation across a wide range of languages with varying degrees of supervision. We have manually analysed the effects of lemmatisation and have offered insights into its advantages and limitations.

In conclusion, post-retrieval lemmatisation is more advantageous for less morphologically complex language pairs, whereas pre-alignment lemmatisation proves more effective for languages with mid-complex morphology. In languages with high morphological complexity, both approaches perform at a comparable level.

Acknowledgments. This work has been partly supported by the Ministry of Education, Youth and Sports of the Czech Republic within the LINDAT-CLARIAH-CZ project LM2023062.

References

1. Anastasopoulos, A., Neubig, G.: Should all cross-lingual embeddings speak English? In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 8658–8679. Association for Computational Linguistics (2020). https://doi.org/10.18653/v1/2020.acl-main.766
2. Artetxe, M., Labaka, G., Agirre, E.: Generalizing and improving bilingual word embedding mappings with a multi-step framework of linear transformations. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence. pp. 5012–5019 (2018). https://doi.org/10.1609/aaai.v32i1.11992
3. Artetxe, M., Labaka, G., Agirre, E.: A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 789–798. Association for Computational Linguistics (2018). https://doi.org/10.18653/v1/P18-1073
4. Bird, S., Klein, E., Loper, E.: Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc." (2009)
5. Conneau, A., Lample, G., Ranzato, M., Denoyer, L., J'egou, H.: Word translation without parallel data. ArXiv **abs/1710.04087** (2017). https://doi.org/10.48550/arXiv.1710.04087
6. Czarnowska, P., Ruder, S., Grave, E., Cotterell, R., Copestake, A.: Don't forget the long tail! a comprehensive analysis of morphological generalization in bilingual lexicon induction. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). pp. 974–983. Association for Computational Linguistics (2019). https://doi.org/10.18653/v1/D19-1090
7. Denisová, M.: Compiling an Estonian-Slovak dictionary with English as a binder. In: Proceedings of the eLex 2021 conference. pp. 107–120. Lexical Computing CZ, s.r.o. (2021)
8. Denisová, M., Rychlý, P.: Evaluation of the cross-lingual embedding models from the lexicographic perspective. In: Proceedings of the eLex 2023 conference. pp. 1–18. Lexical Computing CZ, s.r.o. (2023)
9. Duan, X., Ji, B., Jia, H., Tan, M., Zhang, M., Chen, B., Luo, W., Zhang, Y.: Bilingual dictionary based neural machine translation without using parallel sentences. In:

Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 1570–1579. Association for Computational Linguistics (2020). https://doi.org/10.18653/v1/2020.acl-main.143
10. Herman, O.: Precomputed word embeddings for 15+ languages. In: Proceedings of the Fifteenth Workshop on Recent Advances in Slavonic Natural Languages Processing (RASLAN). pp. 41–46 (2021)
11. Hu, L., Xu, Y.: DM-BLI: Dynamic multiple subspaces alignment for unsupervised bilingual lexicon induction. In: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 2041–2052. Association for Computational Linguistics (2024). https://doi.org/10.18653/v1/2024.acl-long.112
12. Joulin, A., Bojanowski, P., Mikolov, T., Jégou, H., Grave, E.: Loss in translation: Learning bilingual word mapping with a retrieval criterion. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. pp. 2979–2984. Association for Computational Linguistics (2018). https://doi.org/10.18653/v1/D18-1330
13. Kementchedjhieva, Y., Hartmann, M., Søgaard, A.: Lost in evaluation: Misleading benchmarks for bilingual dictionary induction. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). pp. 3336–3341. Association for Computational Linguistics (2019). https://doi.org/10.18653/v1/D19-1328
14. Laville, M., Morin, E., Langlais, P.: About evaluating bilingual lexicon induction. In: Rapp, R., Zweigenbaum, P., Sharoff, S. (eds.) Proceedings of the BUCC Workshop within LREC 2022. pp. 8–14. European Language Resources Association (2022), https://aclanthology.org/2022.bucc-1.2
15. Li, Y., Korhonen, A., Vulić, I.: On bilingual lexicon induction with large language models. In: Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. pp. 9577–9599. Association for Computational Linguistics (2023). https://doi.org/10.18653/v1/2023.emnlp-main.595
16. Mikolov, T., Le, Q.V., Sutskever, I.: Exploiting similarities among languages for machine translation. arXiv preprint arXiv:1309.4168 (2013), https://arxiv.org/abs/1309.4168
17. Orosz, G., Szabó, G., Berkecz, P., Szántó, Z., Farkas, R.: Advancing Hungarian Text Processing with HuSpaCy: Efficient and Accurate NLP Pipelines. In: Ekštein, K., Pártl, F., Konopík, M. (eds.) Text, Speech, and Dialogue. pp. 58–69. Springer Nature Switzerland, Cham (2023). https://doi.org/10.1007/978-3-031-40498-6_6
18. Resiandi, K., Murakami, Y., Nasution, A.H.: Neural network-based bilingual lexicon induction for indonesian ethnic languages. Applied Sciences **13**(15) (2023). https://doi.org/10.3390/app13158666
19. Ruder, S., Vulić, I., Søgaard, A.: A survey of cross-lingual word embedding models. The Journal of Artificial Intelligence Research **65**, 569–631 (2019). https://doi.org/10.1613/jair.1.11640
20. Sadeniemi, M., Kettunen, K., Lindh-Knuutila, T., Honkela, T.: Complexity of european union languages: A comparative approach. Journal of Quantitative Linguistics **15**(2), 185–211 (2008). https://doi.org/10.1080/09296170801961843
21. Sagot, B.: The lefff, a freely available and large-coverage morphological and syntactic lexicon for French. In: Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10). European Language Resources Association (ELRA) (2010)

22. Vulić, I., Glavaš, G., Liu, F., Collier, N., Ponti, E.M., Korhonen, A.: Probing cross-lingual lexical knowledge from multilingual sentence encoders. In: Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics. pp. 2089–2105. Association for Computational Linguistics (2023). https://doi.org/10.18653/v1/2023.eacl-main.153
23. Vulić, I., Moens, M.F.: Monolingual and cross-lingual information retrieval models based on (bilingual) word embeddings. Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval pp. 363–372 (2015). https://doi.org/10.1145/2766462.2767752
24. Wang, X., Ruder, S., Neubig, G.: Expanding pretrained models to thousands more languages via lexicon-based adaptation. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 863–877. Association for Computational Linguistics (2022). https://doi.org/10.18653/v1/2022.acl-long.61
25. Yang, P., Luo, F., Chen, P., Liu, T., Sun, X.: MAAM: A morphology-aware alignment model for unsupervised bilingual lexicon induction. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. pp. 3190–3196. Association for Computational Linguistics (2019). https://doi.org/10.18653/v1/P19-1308
26. Yuan, M., Zhang, M., Van Durme, B., Findlater, L., Boyd-Graber, J.: Interactive refinement of cross-lingual word embeddings. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 5984–5996. Association for Computational Linguistics (2020). https://doi.org/10.18653/v1/2020.emnlp-main.482
27. Zhou, Y., Geng, X., Shen, T., Zhang, W., Jiang, D.: Improving zero-shot cross-lingual transfer for multilingual question answering over knowledge graph. In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. pp. 5822–5834. Association for Computational Linguistics (2021). https://doi.org/10.18653/v1/2021.naacl-main.465