Proceedings of the 15th Workshop on Building and Using Comparable Corpora (BUCC 2022) @LREC2022 , pages 61–66
Marseille, June 25, 2022

Fusion of linguistic, neural and sentence-transformer features for improved term alignment

Andraž Repar[1], Boshko Koloski[1], Matej Ulčar[2], Senja Pollak[1]
[1]Jožef Stefan Institute, Jožef Stefan International Postgraduate School
Jamova cesta 39, Ljubljana, Slovenia
[2]Faculty of Computer and Information Science, University of Ljubljana
Večna pot 113, Ljubljana, Slovenia
{andraz.repar,boshko.koloski,senja.pollak}@ijs.si, matej.ulcar@fri.uni-lj.si

Abstract

Crosslingual terminology alignment task has many practical applications. In this work, we propose an aligning method for the shared task of the 15th Workshop on Building and Using Comparable Corpora. Our method combines several different approaches into one cohesive machine learning model, based on SVM. From shared-task specific and external sources, we crafted four types of features: cognate-based, dictionary-based, embedding-based, and combined features, which combine aspects of the other three types. We added a post-processing re-scoring method, which reduces the effect of hubness, where some terms are nearest neighbours of many other terms. We achieved the average precision score of 0.833 on the English-French training set of the shared task.

Keywords: term alignment, cognates, embeddings, sentence-transformers

1. Introduction

Having the ability to align concepts between languages can provide significant benefits in many practical applications, such as aligning terms between languages in bilingual terminology, aligning keywords in news industry or using aligned concepts as seed data for other NLP tasks like multilingual vector space alignment.

In this paper, we present the experiments and their results on the data provided in the bilingual term alignment in comparable specialized corpora shared task organized as part of the 15th Workshop on Building and Using Comparable Corpora (the BUCC workshop). Given a pair of comparable corpora in two languages and a pair of term lists where terms originate in the two corpora, participants were required to produce lists of term pair candidates ranked by their alignment probability (i.e. terms closer to the top are more likely to be true alignments).

Our method involves a machine learning approach based on our work in (Repar et al., 2019) and (Repar et al., 2021) with additional improvements. Our system uses several external resources detailed in Section 3, all of which are publicly available online.

This paper is organized as follows: Section 1 introduces the topic, Section 2 provides the related work, Section 3 describes the methodology, Section 4 contains the results and Section 5 the conclusion.

2. Related work

Initial attempts at bilingual terminology extraction involved parallel input data (Kupiec, 1993; Daille et al., 1994; Gaussier, 1998), and the interest of the community continued until today. However, most parallel corpora are owned by private companies[1], such as language service providers, who consider them to be their intellectual property and are reluctant to share them publicly. For this reason (and in particular for language pairs not involving English) considerable efforts have also been invested into researching bilingual terminology extraction from comparable corpora (Fung and Yee, 1998; Rapp, 1999; Chiao and Zweigenbaum, 2002; Cao and Li, 2002; Daille and Morin, 2005; Morin et al., 2008; Vintar, 2010; Bouamor et al., 2013a; Bouamor et al., 2013b; Hazem and Morin, 2016; Hazem and Morin, 2017).

The approach designed by Aker et al. (2013) and replicated and adapted in Repar et al. (2019) served as the basis of our work. It was developed to align terminology between languages with the help of parallel corpora using machine-learning techniques. They use terms from the Eurovoc (Steinberger et al., 2002) thesaurus and train an SVM binary classifier (Joachims, 2002) (with a linear kernel and the trade-off between training error and margin parameter c = 10). The task of bilingual alignment is treated as a binary classification - each term from the source language S is paired with each term from the target language T and the classifier then decides whether the aligned pair is correct or incorrect. Aker et al. (2013) run their experiments on the 21 official EU languages covered by Eurovoc with English always being the source language (20 language pairs altogether). They evaluate the performance on a held-out term pair list from Eurovoc using recall, precision and F-measure for all 21 languages. Next, they

[1]However, some publicly available parallel corpora do exist. A good overview can be found at the OPUS web portal (Tiedemann, 2012).

propose an experimental setting for a simulation of a real-world scenario where they collect English-German comparable corpora of two domains (IT, automotive) from Wikipedia, perform monolingual term extraction using the system by Pinnis et al. (2012) followed by the bilingual alignment procedure described above and manually evaluate the results (using two evaluators). They report excellent performance on the held-out term list with many language pairs reaching 100% precision and the lowest recall being 65%. For Slovenian, which is of our main interest, the reported results were excellent with perfect or nearly perfect precision and good recall. The reported results of the manual evaluation phase were also good, with two evaluators agreeing that at least 81% of the extracted term pairs in the IT domain and at least 60% of the extracted term pairs in the automotive domain can be considered exact translations. Repar et al. (2019) tried to reproduce their approach and after initially having little success they were at the end able to achieve comparable results with precision exceeding 90% and recall over 50%.

Despite the problem of bilingual term alignment lending itself well to the binary classification task, there have been relatively few approaches utilizing machine learning. Similar to Aker et al. (2013), Baldwin and Tanaka (2004) generate corpus-based, dictionary-based and translation-based features and train an SVM classifier to rank the translation candidates. Note that they only focus on multi-word noun phrases (noun + noun). A similar approach, again focusing on noun phrases, is also described by Cao and Li (2002). Finally, Nassirudin and Purwarianti (2015) also reimplement Aker et al. (2013) for the Indonesian-Japanese language pair and further expand it with additional statistical features.

3. Methodology

Initial experiments were performed with cross-lingual embeddings (see Section 3.1) and sentence transformers (see Section 3.2). However, the results were lower than expected, which is why we adapted an approach described in Repar et al. (2021) by adding additional features based on the cross-lingual embedding and sentence transformer experiments.

3.1. Cross-lingual aligned embeddings

We used fastText Bojanowski et al. (2017) word embeddings for both involved languages. We constructed a bilingual English-French dictionary from Wiktionary entries, using the wikt2dict tool Acs (2014). The extracted dictionary has 204 341 entries. For the purpose of embedding alignment, we filtered it to keep only single-word entries, i.e. those that have a single word in both languages. After the filtering, we had 129 912 entries, of which 24 923 have an identical word in both languages (e.g. place names or chemicals) There's an average of 1.55 English translations for each French word, and 1.56 French translations for each English word. 23.4% of English words have multiple French translations, while 24.3% of French words have multiple English translations.

We then aligned the French and English word embeddings into a common vector space in a supervised manner, utilizing the bilingual dictionary. We used Vecmap Artetxe et al. (2018) tool, which aligns the vectors using the Moore-Penrose pseudo-inverse, which minimizes the sum of squared Euclidean distances. We extracted one vector for each term in each language. For multi-word terms we averaged the word vectors of all the words the term is composed of. Finally we use the cosine similarity score to find the most similar terms in language 1 for each term in language 2, and vice-versa. Using this approach, we achieve an average precision of 0.496 (for details, see Table 2).

3.2. Sentence-transformers features

We used the Sentence-Transformers Reimers and Gurevych (2019) model to embed the terms of the both languages. We utilized the implementations of *c19 python library* (Koloski et al., 2021) to obtain the embeddings [2]. The sentence-transformer architecture is designed to solve the task of sentence similarity, it leverages BERT tokens and via pooling it creates sentence-embeddings. The BERT Devlin et al. (2018) model uses tokens as input to it's transformer architecture, the BERT-tokenizer tokenizes the words in subwords. We consider using the sentence-transformers because of the sub-word information that is taken into account while learning the model. We feed the model with single or multi-word terms as "sentences" and obtain the sentence-embedding.

3.2.1. Terms as sentences evaluation methodology

For each term in each language respectively we obtain the sentence-embeddings. Next, for each term in English we rank all of the French terms with regards of cosine-similarity.

We consider using five different Language Models:

- XLM (Lample and Conneau, 2019)
- DistilBERT (Sanh et al., 2019)
- All-MPNet (Song et al., 2020)
- MiniLM (Wang et al., 2020)
- Roberta-Large (Liu et al., 2019)

The highest average precision of 0.680 among the five models was achieved with the *distilbert-base* model (for details, see Table 2).

3.3. Supervised machine learning approach

Since the results of the individual approaches described in the previous two sections were lower than expected, we further experimented with combining the individual models into a machine learning model. We reused and

[2] `https://github.com/bkolosk1/c19_rep`

adapted an approach described in Repar et al. (2021) by incorporating the cosine similarity values of the cross-lingual and sentence transformer models into features of the machine learning model.
This approach uses Eurovoc (Steinberger et al., 2002) terms, Giza++ dictionaries (generated from the DGT translation memory (Steinberger et al., 2013)) and word similarity information to generate features for an SVM binary classifier (Joachims, 2002) (with the trade-off between training error and margin parameter c = 10). The model is trained on a list of 7181 Eurovoc English-French term pairs as well as an additional 1.4 million incorrect term pairs generated by randomly pairing English and French Eurovoc terms to simulate real-world conditions. In addition to the binary classification, the model also provides confidence scores which are later used to rank aligned candidate pairs.
For each potential candidate pair, we calculate features of the following types:

- Cognate-based features
- Dictionary-based features
- Embedding-based features
- Combined features

As described in Repar et al. (2019) and Repar et al. (2021), cognate-based features take advantage of word similarity between languages (e.g. *democracy* in English and *démocratie* in French) and dictionary-based features are calculated using results of the Giza++ word alignment algorithm. Embedding-based features are calculated using cosine similarity scores described in Sections 3.1 and 3.2. For each model, we produce a list of word pairs with their cosine similarity scores. These scores are then used to generate embedding features by creating 3-tuples[3] of most similar - based on cosine similarity - source-to-target and target-to-source words, such as:

- xénophobie ['xenophobia', '0.744'], ['racism', '0.6797'], ['anti-semitism', '0.654']
- femme ['woman', '0.7896'], ['women', '0.73'], ['female', '0.722']

where the tuple contains the source language word along with their three most likely corresponding words in the target language and their cosine similarities. The 3-tuples of most similar words were used to construct additional features for the machine learning algorithm as indicated in 1. Finally, combined features combine some aspects of the first three feature types.

[3]This number was determined experimentally.

3.4. Post-process re-ordering

In post-processing we altered the confidence scores of some of the term-pairings. For some term x_1 from language 1, we wanted to ensure that the best performing aligned pair is as close to the top of the list as possible. For x_1, a large number of candidate terms from language 2 can have a high confidence score for a matching term and this might negatively affect the final average precision scores as defined in the shared task, since most terms would not have more than 2-3 correct alignments. Another term x_2 from language 1 might have a lower confidence score with every candidate term from language 2 than all the candidates for x_1. That is, there are such $x_1, x_2 \in L_1$, that $S(x_1, y) > \max_{y'}(S(x_2, y')), \forall y \in L_2$, where S is confidence/similarity score and L_1 and L_2 are languages 1 and 2, respectively. We therefore boosted the confidence scores of the top n candidates for each term by a constant c. Based on the performance on the training dataset, we chose the parameters $n = 1$ and $c = 1.0$.

4. Experimental setup

In step one, we trained the model on publicly available data (Eurovoc thesaurus, Giza++ word alignment lists trained on the DGT corpus and embedding and transformer models trained on the data provided within the BUCC shared task). In step two, we evaluated its performance on the term lists provided as part of the training package in the shared task. To do so, we generated all possible term pairs between the English and French term lists, calculated the features described in Table 1, produced predictions using the model trained in step one and evaluated them against the English-French term list provided as part of the shared task training data.

5. Results

We report results in Table 2. Using just individual language models described in Section 3.2, the best average precision (0.680) is achieved with the distilbert-base model. When we used the SVM approach described in Repar et al. (2021) (i.e. the *SVM old*, we reach an average precision of 0.712 and when we add additional features based on sentence transformer models we achieve an average precision of 0.833 (i.e. *SVM new*. The post-process re-ordering parameters *n* and *c* were as indicated in Section 3.4.

6. Conclusion

In this paper, we presented the results of our experiments for the shared task of the 15th Workshop on Building and Using Comparable Corpora. We first attempted to align terms using cross-lingual embedding and sentence transformer models, but the results were less than satisfactory. Next, we reused an existing machine learning approach and added additional features based on the cross-lingual embedding and sentence

Feature	Cat	Description
isFirstWordTranslated	Dict	Checks whether the first word of the source term is a translation of the first word in the target term (based on the Giza++ dictionary)
isLastWordTranslated	Dict	Checks whether the last word of the source term is a translation of the last word in the target term
percentageOfTranslatedWords	Dict	Ratio of source words that have a translation in the target term
percentageOfNotTranslatedWords	Dict	Ratio of source words that do not have a translation in the target term
longestTranslatedUnitInPercentage	Dict	Ratio of the longest contiguous sequence of source words which has a translation in the target term (compared to the source term length)
longestNotTranslatedUnitInPercentage	Dict	Ratio of the longest contiguous sequence of source words which do not have a translation in the target term (compared to the source term length)
Longest Common Subsequence Ratio	Cogn	Measures the longest common non-consecutive sequence of characters between two strings
Longest Common Substring Ratio	Cogn	Measures the longest common consecutive string (LCST) of characters that two strings have in common
Dice similarity	Cogn	2*LCST / (len(source) + len(target))
Needlemann-Wunsch distance	Cogn	LCST / min(len(source), len(target))
isFirstWordCognate	Cogn	A binary feature which returns True if the longest common consecutive string (LCST) of the first words in the source and target terms divided by the length of the longest of the two words is greater than or equal to a threshold value of 0.7 and both words are longer than 3 characters
isLastWordCognate	Cogn	A binary feature which returns True if the longest common consecutive string (LCST) of the last words in the source and target terms divided by the length of longest of the two words is greater than or equal to a threshold value of 0.7 and both words are longer than 3 characters
Normalized Levensthein distance (LD)	Cogn	1 - LD / max(len(source), len(target))
isFirstWordCovered	Comb	A binary feature indicating whether the first word in the source term has a translation or transliteration in the target term
isLastWordCovered	Comb	A binary feature indicating whether the last word in the source term has a translation or transliteration in the target term
percentageOfCoverage	Comb	Returns the percentage of source term words which have a translation or transliteration in the target term
percentageOfNonCoverage	Comb	Returns the percentage of source term words which have neither a translation nor transliteration in the target term
difBetweenCoverageAndNonCoverage	Comb	Returns the difference between the last two features
isFirstWordMatch	Emd	Checks whether the first word of the source term is the most likely translation of the first word in the target term (based on the aligned embeddings)
isLastWordMatch	Emd	Checks whether the last word of the source term is the most likely translation of the last word in the target term (based on the aligned embeddings)
percentageOfFirstMatchWords	Emb	Ratio of source words that have a first match (i.e. first position in the 3-tuple) in the target term
percentageOfNotFirstMatchWords	Emb	Ratio of source words that do not have a first match (i.e. first position in the 3-tuple) in the target term
longestFirstMatchUnitInPercentage	Emb	Ratio of the longest contiguous sequence of source words which has a first match (first position in the 3-tuple) in the target term (compared to the source term length)
longestNotFirstMatchUnitInPercentage	Emb	Ratio of the longest contiguous sequence of source words which do not have a first match (first position in the 3-tuple) in the target term (compared to the source term length)
isFirstWordTopnMatch	Emd	Checks whether the first word of the source term is in the 3-tuple of most likely translations of the first word in the target term (based on the aligned embeddings)
isLastWordTopnMatch	Emd	Checks whether the first word of the source term is not in the 3-tuple of most likely translations of the first word in the target term (based on the aligned embeddings)
percentageOfTopnMatchWords	Emb	Ratio of source words that have a match (i.e. any position in the 3-tuple) in the target term
percentageOfNotTopnMatchWords	Emb	Ratio of source words that do not have a match (i.e. any position in the 3-tuple) in the target term
longestTopnMatchUnitInPercentage	Emb	Ratio of the longest contiguous sequence of source words which has a match (any position in the 3-tuple) in the target term (compared to the source term length)
longestNotTopnMatchUnitInPercentage	Emb	Ratio of the longest contiguous sequence of source words which do not have a match (any position in the 3-tuple) in the target term (compared to the source term length)
isFirstWordCoveredEmbeddings	Comb	A binary feature indicating whether the first word in the source term has a match (any position in the 3-tuple) or transliteration in the target term
isLastWordCoveredEmbeddings	Comb	A binary feature indicating whether the last word in the source term has a match (any position in the 3-tuple) or transliteration in the target term
percentageOfCoverageEmbeddings	Comb	Returns the percentage of source term words which have a match (any position in the 3-tuple) or transliteration in the target term
percentageOfNonCoverageEmbeddings	Comb	Returns the percentage of source term words which do not have a match (any position in the 3-tuple) or transliteration in the target term
diffBetweenCoverageAnd-NonCoverageEmbeddings	Comb	Returns the difference between the last two features

Table 1: Features used in the experiments. Note that some features are used more than once because they are direction-dependent or used multiple times with different embedding or transformer models.

Model	Average precision
aligned fastText	0.496
distilbert-base	0.680
xlm-r	0.650
all-mpnet	0.616
all-MiniLM	0.621
roberta-large	0.523
SVM old	0.712
SVM new	**0.833**

Table 2: Results

transformer models. Using this model, we achieved the average precision of 0.833. Our experiments show that careful feature engineering could still produce better results than more novel deep learning approaches.

In terms of future work, there is still room for improvement which could be achieved by generating additional features using other transformer or embedding models. The system is also quite resource intensive — model training and prediction on the BUCC dataset took more than 24 hours. Finally, there is also room for a more systematic approach to the postprocess re-ranking step.

7. Acknowledgements

This work was supported by the Slovenian Research Agency (ARRS) grants for the core programme Knowledge technologies (P2-0103), as well as the European Union's Horizon 2020 research and innovation programme under grant agreement No 825153, project EMBEDDIA (Cross-Lingual Embeddings for Less-Represented Languages in European News Media).

8. Bibliographical References

Acs, J. (2014). Pivot-based multilingual dictionary building using Wiktionary. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation LREC*.

Aker, A., Paramita, M., and Gaizauskas, R. (2013). Extracting bilingual terminologies from comparable corpora. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 402–411.

Artetxe, M., Labaka, G., and Agirre, E. (2018). Generalizing and improving bilingual word embedding mappings with a multi-step framework of linear transformations. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Baldwin, T. and Tanaka, T. (2004). Translation by machine of complex nominals: Getting it right. In

Proceedings of the Workshop on Multiword Expressions: Integrating Processing, pages 24–31.

Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Bouamor, D., Popescu, A., Semmar, N., and Zweigenbaum, P. (2013a). Building specialized bilingual lexicons using large scale background knowledge. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 479–489, Seattle, Washington, USA, October. Association for Computational Linguistics.

Bouamor, D., Semmar, N., and Zweigenbaum, P. (2013b). Context vector disambiguation for bilingual lexicon extraction from comparable corpora. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 759–764.

Cao, Y. and Li, H. (2002). Base noun phrase translation using web data and the EM algorithm. In *Proceedings of the 19th International Conference on Computational Linguistics - Volume 1*, pages 1–7.

Chiao, Y.-C. and Zweigenbaum, P. (2002). Looking for candidate translational equivalents in specialized, comparable corpora. In *Proceedings of the 19th International Conference on Computational Linguistics - Volume 2*, pages 1–5.

Daille, B. and Morin, E. (2005). French-English terminology extraction from comparable corpora. In *Natural Language Processing – IJCNLP 2005*, pages 707–718.

Daille, B., Gaussier, E., and Langé, J.-M. (1994). Towards automatic extraction of monolingual and bilingual terminology. In *Proceedings of the 15th Conference on Computational Linguistics - Volume 1*, pages 515–521.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Fung, P. and Yee, L. Y. (1998). An IR approach for translating new words from nonparallel, comparable texts. In *Proceedings of the 17th International Conference on Computational Linguistics - Volume 1*, pages 414–420.

Gaussier, E. (1998). Flow network models for word alignment and terminology extraction from bilingual corpora. In *Proceedings of the 17th International Conference on Computational Linguistics - Volume 1*, pages 444–450.

Hazem, A. and Morin, E. (2016). Efficient data selection for bilingual terminology extraction from comparable corpora. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 3401–3411.

Hazem, A. and Morin, E. (2017). Bilingual word embeddings for bilingual terminology extraction from specialized comparable corpora. In *Proceedings of the 8th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 685–693.

Joachims, T. (2002). *Learning to classify text using support vector machines: Methods, theory and algorithms*. Kluwer Academic Publishers.

Koloski, B., Stepišnik-Perdih, T., Pollak, S., and Škrlj, B. (2021). Identification of covid-19 related fake news via neural stacking. In Tanmoy Chakraborty, et al., editors, *Combating Online Hostile Posts in Regional Languages during Emergency Situation*, pages 177–188, Cham. Springer International Publishing.

Kupiec, J. (1993). An algorithm for finding noun phrase correspondences in bilingual corpora. In *Proceedings of the 31st annual meeting on Association for Computational Linguistics*, pages 17–22.

Lample, G. and Conneau, A. (2019). Cross-lingual language model pretraining.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019). Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.

Morin, E., Daille, B., Takeuchi, K., and Kageura, K. (2008). Brains, not brawn: The use of smart comparable corpora in bilingual terminology mining. *ACM Trans. Speech Lang. Process.*, 7(1):1:1–1:23, October.

Nassirudin, M. and Purwarianti, A. (2015). Indonesian-Japanese term extraction from bilingual corpora using machine learning. In *Advanced Computer Science and Information Systems (ICACSIS), 2015 International Conference on*, pages 111–116.

Pinnis, M., Ljubešić, N., Stefanescu, D., Skadina, I., Tadić, M., and Gornostaya, T. (2012). Term extraction, tagging, and mapping tools for under-resourced languages. In *Proceedings of the 10th Conference on Terminology and Knowledge Engineering (TKE 2012), June*, pages 20–21.

Rapp, R. (1999). Automatic identification of word translations from unrelated english and german corpora. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics*, pages 519–526.

Reimers, N. and Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *CoRR*, abs/1908.10084.

Repar, A., Martinc, M., and Pollak, S. (2019). Reproduction, replication, analysis and adaptation of a term alignment approach. *Language Resources and Evaluation*, pages 1–34.

Repar, A., Martinc, M., Ulčar, M., and Pollak, S. (2021). Word-embedding based bilingual terminology alignment. *Electronic lexicography in the 21st century (eLex 2021) Post-editing lexicography*, page 98.

Sanh, V., Debut, L., Chaumond, J., and Wolf,

T. (2019). Distilbert, a distilled version of BERT: smaller, faster, cheaper and lighter. *CoRR*, abs/1910.01108.

Song, K., Tan, X., Qin, T., Lu, J., and Liu, T. (2020). Mpnet: Masked and permuted pre-training for language understanding. *CoRR*, abs/2004.09297.

Steinberger, R., Pouliquen, B., and Hagman, J. (2002). Cross-lingual document similarity calculation using the multilingual thesaurus eurovoc. *Computational Linguistics and Intelligent Text Processing*, pages 101–121.

Steinberger, R., Eisele, A., Klocek, S., Pilos, S., and Schlüter, P. (2013). DGT-TM: A freely available translation memory in 22 languages. In *Proceedings of the 8th international conference on Language Resources and Evaluation (LREC'2012)*.

Tiedemann, J. (2012). Parallel data, tools and interfaces in opus. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, may. European Language Resources Association (ELRA).

Vintar, Š. (2010). Bilingual term recognition revisited: The bag-of-equivalents term alignment approach and its evaluation. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 16(2):141–158.

Wang, W., Wei, F., Dong, L., Bao, H., Yang, N., and Zhou, M. (2020). Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers. In H. Larochelle, et al., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 5776–5788. Curran Associates, Inc.

9. Language Resource References