Bilingual Terminology Alignment Using Contextualized Embeddings

Setha Imene
Research Center on Scientific
and Technical Information CERIST.
Algiers. Algeria
sethaimene1@gmail.com

Aliane Hassina
Research Center on Scientific
and Technical Information CERIST.
Algiers. Algeria
ahassina4@gmail.com

Abstract

Terminology Alignment faces big challenges in NLP because of the dynamic nature of terms. Fortunately, over these last few years, Deep Learning models have shown very good progress with several NLP tasks such as multilingual data resourcing, glossary building, terminology understanding. . . etc. In this work, we propose a new method for terminology alignment from a comparable corpus (Arabic/French languages) for the Algerian culture field.
We aim to improve bilingual alignment based on contextual information of a term and to create a significant term bank i.e. a bilingual Arabic-French dictionary. We propose to create word embeddings for both Arabic and French languages using ELMO model focusing on contextual features of terms. Then, we map those embeddings using a Seq2seq model.
We use multilingual-BERT and All-MiniLM-L6 as baseline models to compare terminology alignment results. Experimentations showed quite satisfying alignment results.

1 Introduction

For many years now, humans have wanted to enhance the machine's learning and understanding capacity to reach our potential of thinking, awareness, and power of judgment. making us wonder, is it close enough for a machine to be able to recognize and realize as we do? In artificial intelligence and NLP tasks, new models are frequently created to automate and facilitate life in different areas.
However, some fields have a long road to go, such as cross-lingual alignment and contextual translation. Terminology alignment is a very tough task to handle in NLP since one term can have several meanings according to its position and use. Aligned terms are often incorrect or misplaced especially while working with non-similar language families. For example, a sentence or a term might be translated into 3 or more different expressions and still not have the correct corresponding meaning. We can define bilingual terminology alignment as the process of mapping two terms or sentences in two different languages. Alignment provides significant benefits in many NLP tasks when properly applied like machine translation, clustering, building bilingual dictionaries, multilingual data resourcing. . . etc. The primary purpose of this work is to build a bilingual term bank for Arabic and French languages. Bilingual Alignment can be applied either to sentences or terms, in this article, we focus on bilingual terms only. According to (Och and Ney, 2003), we have a source language sentence containing the terms:

$$f = f1, f2, ...fj.$$

and a target language sentence:

$$e = e1, e2, \ldots ei.$$

An alignment A is defined as a subset of the Cartesian product of the word positions(Mikolov et al., 2013).

$$A \subseteq (j, i) : j = 1, ..., J; i = 1, ..., I$$

As shown in this example (See Figure 1):

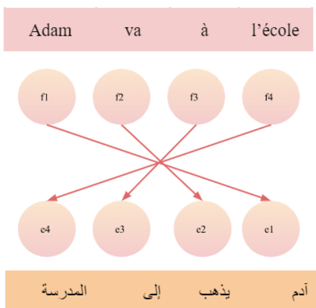


Figure 1: example of aligned terms in two languages

Proceedings of the First ConTenNTS Workshop and the 16th BUCC workshop, pages 1–8
Varna, Sep 7, 2023
https://doi.org/10.26615/978-954-452-090-8_001

The remainder of this paper is organized as follows. In section 2, we present related works on bilingual terminology alignment. Section 3 presents our methodology and system architecture. section 4 is dedicated to experiments and results of our method. We conclude the paper with a general conclusion and future perspectives.

2 Related Work

Although bilingual terminology alignment (referred to as BTA in the rest of the paper) task is challenging and tough, considerable efforts have been invested into this research field starting in the early 90s by IBM Watson research center (Brown et al., 1990) who introduced statistical alignment models, called IBM models using parallel corpora. Basically, there are 5 basic statistical models (IBM models, 2023) IBM1,2,3,4,5. Another one was added later combining IBM4 and HMM model(Hidden Markov model) based on assumptions such as:

- The target sentence length j is independent of source length i.
- For each target word, all alignments (including alignment to NULL) are equally likely and do not depend on the particular word or its position in the sentence.
- Once the alignments have been determined, the target word depends only on the source word to which it is aligned.
- The translation depends only on the source and target word pair, and not on any previous source or target words.
- The reordering depends only on the position of the target word, the position of the source word, and the lengths of the two sentences.

Many existing methods use IBM models, (Lee et al., 2010) applied IBM1 model using an unsupervised EM-based hybrid model[1] to extract bilingual terminology from comparable corpora through document alignment constraints. Using Giza++, (Moore, 2005) aligned their parallel corpus using the IBM4 model. As in (Macken et al., 2013) the famous TExSIS tool for terminology extraction is based on the IBM4 model for alignment. A combination of IBM1, IBM4 and HMM models is introduced in (Zhao and Xing, 2007) to perform alignment on parallel sentence pairs.

Besides IBM models, alternative statistical models focus on carried statistical properties of a given term or sentence, they vary from length-based, frequency-based, and lexical-based models. In (Salameh et al., 2011), the authors build a system to align English-Arabic sentences using a parallel corpus and focus on applying the best preprocessing steps to enhance their results. (Ittycheriah and Roukos, 2005) describes a maximum entropy-based method for Arabic-English term alignment. However, the recent state-of-the-art is basically governed by machine learning and deep learning models.

Generally, machine learning models treat alignment as a classification problem. In (Repar et al., 2018), the authors use an SVM model as a classifier for the task, adding some improvements to the model that was applied to the English-Slovinian language pair and applied to the Eurovoc thesaurus as the main dataset. (Kontonatsios et al., 2014) built a comparable corpus collected from Wikipedia as a 4k biomedical English term. The authors used a Logistic regression classifier for learning a string similarity measure of term translations. More recently, Deep Learning models achieved high scores and outstanding performance in understanding and translating words and phrases. A very interesting work by (Adjali et al., 2022) adopts the Compositional with Word Embedding Projection (CMWEP) approach of (Liu et al., 2018) to create dictionaries using a comparable corpus. They create WE's using FasText and learn the mapping using a linear transformation approach (Artetxe et al., 2016). (Dev et al., 2021) develop a family of techniques to align WEs, using several mechanisms such as glove, Word2Vec, and fastText, with Wikipedia as an initial dataset. In (Cao et al., 2020), the authors use multilingual BERT to align Bulgarian and Greek using a small parallel corpus extracted from Wikipedia. Another interesting work is (Garg et al., 2019) where the authors train a transformer and build an encoder-decoder model to build a framework for different language translations and where results outperform both Giza++ and IBM models results.

3 Proposed Approach

In this section, we describe our methods and models for BTA using context-based embeddings for both Arabic and French languages. We begin with

[1] Expectation Maximization model.

a system design that briefly shows the main used models and techniques for creating bilingual term pairs. starting with creating contextual WE to find the best equivalent of a given term from the source S to the target T language.

3.1 System Design

Our system depicted in Figure 2, involves the following steps:
Step1: create word vectors (the vocabulary) using the ELMO model for both source and target languages.
Step2: use a small dictionary to feed the models.
Step3: learn the alignments using a Seq2seq model.
Step4: align the list of terms from source to target languages.

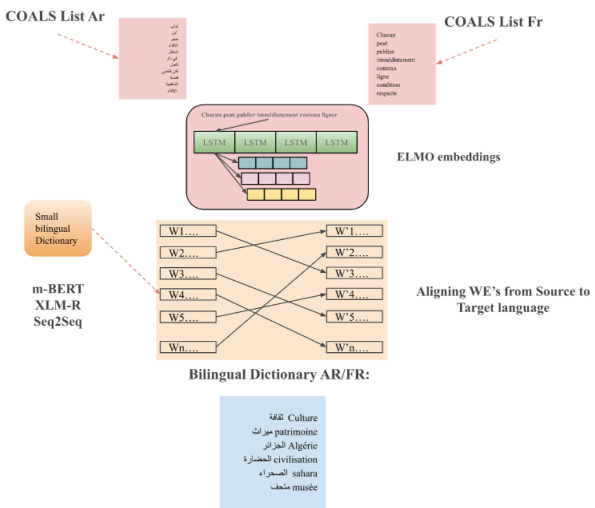


Figure 2: A global overview of the general system's architecture

3.2 Contextual Word Embeddings

Word embeddings (WE) are high-dimensional vector representations of words, based on the words' contexts. WE provide relevant, meaningful information for NLP tasks. The approaches for learning embeddings evolved from static free-word-order to contextualized and deeply contextualized. Word2Vec and Glove are context-independent, word-based representations that do not take word order into account in their training; for each word, we have just one vector as an output. This vector gathers all the meanings of the word. Elmo and BERT are contextual representations that take word order into account and can generate different vectors for a word, capturing all senses based on that word's position in the sentence. Our main goal is to capture the semantic features of a term, in order to compare term vectors across different languages. Therefore we chose Elmo to create our word embeddings WE.

3.3 ELMO

Contextual WE have been developed for better language modeling and to overcome the limitations of traditional methods. Elmo (Embedding for Language Models) (Gardner et al., 2018) has been developed by the Allen Institute NLP group. It is a bidirectional LSTM character-based model that learns word representations using character convolutions and can handle different vocabulary meanings. The main idea is to check all the sentences before creating the word vector, ELMO focuses only on the semantic features of terms, which makes ELMO highly relevant for the BTA task. Furthermore, the concatenation of right-to-left and left-to-right using LSTM should, in theory, generate more accurate word representations and therefore a better term alignment. In our work, we choose to use the Multilingual Elmo embeddings[2], which was pre-trained on 20 million words data randomly sampled from the raw text released by the shared task wiki dump + common crawl, (github, 2020) for 44 languages till this day.

3.4 Baseline Models

In order to evaluate our ELMO model, we have chosen to implement as baselines, recent models that have been successfully used in machine translation: Multilingual-Bert, All-MiniLM-L6, and Seq2seq combined with fasttext embeddings.

3.4.1 BERT-Base-Multilingual-Cased

Multilingual BERT (referred to as mBERT in the rest of the article) is an extension of the original BERT (Bidirectional Encoder Representations from Transformers) model. In other words, it is a multilingual version of BERT. BERT (Devlin et al., 2018) is the most powerful tool for language understanding in human history, and it is everywhere: e-mails, web pages, browsers…etc. It is an attention-based model that uses a transformer with positional encoding to represent word positions using a masked language modeling (MLM) objective.

[2]https://github.com/HIT-SCIR/ELMoForManyLangs

The transformer comprises an encoder to read the sentence and a decoder to predict the next lines. This means that BERT captures the context on both the left and right sides of the sentence to make a prediction. The main architecture comprises 12 layers(transformer blocks), 12 attention heads, and 110 million parameters (See Figure 3). The Google

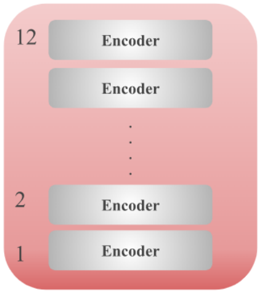


Figure 3: BERT's model general architecture

research team introduced mBERT (Devlin et al., 2018) very soon after the original BERT. It was initially pre-trained for 104 languages and it showed a great performance in several NLP tasks.

3.4.2 Seq2Seq Model With Fastext

Sequence to Sequence is a well-known machine translation model that was introduced by Google, it takes a sequence of items as inputs (terms, phrases, numbers…etc) and outputs another sequence of predicted items as well.(analyticsvidhya, 2023) Seq2Seq models use a powerful encoder-decoder neural mechanism, which is often based on Recurrent neural networks RNN (See Figure 4). Encoders read the input sequence and summarize the information in context vectors. We discard the outputs of the encoder by only preserving these vectors. Where context vectors aim to encapsulate the information for all input elements in order to help the decoder make accurate predictions(analyticsvidhya, 2023).

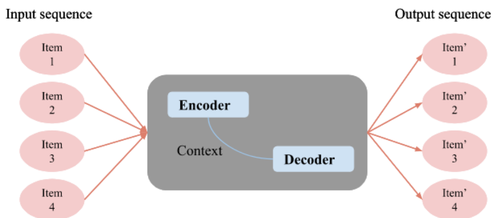


Figure 4: Seq2seq model's architecture

3.4.3 All-MiniLM-L6

All-MiniLM-L6 is a sentence transformer model that maps sentences and paragraphs to a 384-dimensional dense vector space and can be used for tasks like clustering or semantic search (huggingface, 2023). The model was pretrained on a 1B sentence pairs dataset using a contrastive learning objective: given a sentence from the pair, the model should predict which out of a set of randomly sampled other sentences, was actually paired with it in the dataset. This model is intended to be used as a sentence and short paragraph encoder. Given an input text, it outputs a vector that captures the semantic information. The sentence vector may be used for information retrieval, clustering, or sentence similarity tasks (huggingface, 2023)

4 Experiments & Results

In this section, we examine the performance of the baseline models for French and Arabic languages based on two tests. First, we start by using WEs in Seq2seq model with fasttext embeddings to compare WEs without contextual information with ELMO's embeddings for the mapping results. In the second experiment, we compare baseline models' results for the BTA task. Lastly, we evaluate the model's performance using evaluation metrics: Precision, Recall, and F1-score.

4.1 Dataset Resources

The main dataset of this work is provided from (Imene and Hassina, 2022) where a set of terms in Arabic and French languages were collected from Wikipedia pages in the "Algerian culture" domain pages and all related pages. Extracted pages went through a monolingual terminology extraction process using **COALS model** (Correlated Occurrence Analogue to Lexical Semantics)(Rohde et al., 2006). As we can see in Table 1, we use about 28k of Arabic tokens and 30k of French.

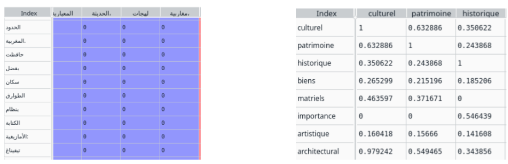
Terms language	Terms number
Arabic language	27 500 terms
French language	30 000 terms

Table 1: Dataset details.

4.2 Some Notes About The Dataset:

- The dataset contains 57 500K terms.

- We consider both simple and Muti-word terms for the process.
- Most of the Extracted terms are in-domain terms for the specific field of "Algerian culture" (See Figure 5).
- We feed some in-domain Multi-word terms into the dictionary to be recognized by the models.
- No preprocessing is applied, the vocabulary is already preprocessed in the terminology extraction step.



Index	المعياري	الحديثة.	لهجات	مغاربية.
الحدود		0	0	0
المغربية.		0	0	0
حافظت		0	0	0
بفضل		0	0	0
سكان		0	0	0
الطوارق		0	0	0
بنظام		0	0	0
الكتابة		0	0	0
الأمازيغية:		0	0	0
تيفيناغ		0	0	0

Index	culturel	patrimoine	historique
culturel	1	0.632886	0.350622
patrimoine	0.632886	1	0.243868
historique	0.350622	0.243868	1
biens	0.265299	0.215196	0.185206
matriels	0.463597	0.371671	0
importance	0	0	0.546439
artistique	0.160418	0.15666	0.141608
architectural	0.979242	0.549465	0.343856

Figure 5: The general form of the dataset in French and Arabic languages

4.3 Seed Dictionary

We use our dataset to create a small dictionary. it contains about 200 terms matched with their exact equivalent from source to target language. We manually review the dictionary pairs to confirm all mapped terms. It contains both single-word and multi-word terms. We also try to add a sufficient number of Multi-word in-domain terms, and acronyms to better feed the alignment models. for example:

ONU → هيئة الأمم المتحدة

unicef → منظمة اليونيسيف

unesco → منظمة اليونسكو

This small bilingual dictionary is used as an additional resource to feed the models with some in-domain terms.

4.4 Evaluation Metrics

According to (Sabet et al., 2020), given a set of predicted alignment edges A and a set of sure, possible gold standard edges S, P (where S is a subset of P).

We use the following evaluation measures:

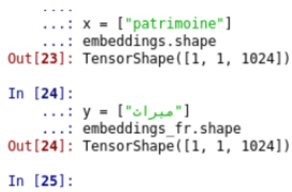
$$Recall = |A \cap S||S|$$

$$Precision : |A \cap P||A|$$

$$F1 - Score = (2PrecRec)/(Prec + Rec)$$

4.5 Contextual Space Vectors

Using the Elmo model, we create WE for source and target languages. Based on contextual features provided by the Elmo model, for instance, the term "patrimoine" and its translation conceivably share the same vector's structure as shown in Figure 6 below:



```
   ...: x = ["patrimoine"]
   ...: embeddings.shape
Out[23]: TensorShape([1, 1, 1024])

In [24]:
   ...: y = ["ميراث"]
   ...: embeddings_fr.shape
Out[24]: TensorShape([1, 1, 1024])

In [25]:
```

Figure 6: An example of two terms sharing the same WEs

- After finishing all previous steps we load the WE to apply our alignment method next.
- For the following tests we consider French as the source language and Arabic as the target language.
- We use Fasttext aligned monolingual vectors[3] to test with. The Facebook team provides these vectors in 89 languages and 78 aligned matrices including French and Arabic. Those matrices are aligned based on a linear transformation (matrix) using the SVD function.(Smith et al., 2017)

For the first test, we apply term alignment using the Seq2seq model with Elmo WE and Fasttext WE to compare them. We start by creating WEs using Elmo for our list of terms, then we use Fasttext vectors as well (We download the available multilingual space vectors for both Arabic and French).

Word vectors	Fasttext	ELMO
Alignment Precision on 100 terms of data	49.9%	62.3%

Table 2: Alignment results using Elmo & Fast-text

[3]https://github.com/facebookresearch/fastText

4.6 Alignment Process

In the upcoming experiments, we use a Desktop Computer with an Intel Core I5 7400 CPU with a 3.00 GHz frequency and 16 GB RAM. We also train our models on a workstation that contains 4 GPU RTX2080ti. We implement the proposed models using Python. Pytorch, tensorflow, and transformers libraries are used in the following experiments.

4.6.1 Multilingual-Bert

Multilingual Bert is a pre-trained model on 104 Wikipedia for 104 languages. Trained with 12 transformer layers, with 12 heads and 768 hidden dimensions each with a total number of 110M parameters. It scores high precision for translation tasks that reached 82% in English and 71% in Arabic. We load the model and apply it directly to our term's list, results are shown in Table 4.

4.6.2 Seq2Seq Model With Fasttext

Our second baseline model has been used for machine translation since 2014. We upload our Fasttext WEs, then we pass directly to create the RNN encoder-decoder networks using the Pytorch library. We train the model on 30 epochs to predict our list of Arabic terms.

4.6.3 All-MiniLM-L6

From the various available multilingual models that are based on sentence transformers, we chose All-MiniLM-L6, to align our vector spaces which is known for its fast results and good quality in semantic similarity search. We use the "Sentence-transformers" library to align not sentences but parts of them, which are in our case simple terms from source to target languages.

m-BERT	Seq2Seq	All-Mini-ML-6
le patrimoine = التراث	le patrimoine = ميراث	le patrimoine = التراث
Algérien = الجزائري	Algérien =جزائري	Algérien = الجزائري
la culture = الثقافة	la culture = الثقافة	la culture = الثقافة
la civilisation = الحضارة	la civilisation = الحضارة	la civilisation = الحضارة

Table 3: Alignment results from French to Arabic.

- Table 3 shows alignment results for the following terms respectfully: "patrimony", "Algerian" "Culture", and "civilization" (in French and Arabic languages).

4.7 Baselines Comparison

We compare previous baseline models to each other. We apply our test on the 100 first terms of both lists. to compare results between the models.

Alignment	Precision	Recall	F1-Score
M-BERT	84%	72%	77.5%
Seq2seq/Fasttext	50%	34%	40%
Seq2seq/ELMO	62%	46%	52.8%
All-MiniLM-L6	82%	70%	75.5%

Table 4: Evaluation results from French into Arabic.

4.8 Discussion

In this work, we tackle terminology alignment based on contextualized embeddings for a French/Arabic list of terms. We use three baseline models to apply the alignment. From the first experiment, we hypothesized that contextual embeddings would give better results in terminology alignment, which has shown to be true since Elmo's embeddings capture all meanings of terms and present it as a multiple vector choice to be aligned and Table 2 along with Figure 6 clearly confirms our hypothesis. In the second experiment, we align Arabic and French extracted terms using the proposed baseline models. Although Sentence transformer models are made to work mainly with phrases and paragraphs, results of mBERT and All-MiniLM-L6 are very close and alike, many translations are the same in both models and as shown in Table 3, we can see in the example of "civilization" term "la civilisation = "الحضارة" in mBERT while in All-MiniLM-L6 it means "حضارة".

The reason that those models perform better and give efficient results is related to the fact that the transformer's self-attention mechanism identifies the context which gives meaning to each position in the input sequence, allowing more parallelization than RNN models and reducing the training time. As for the Seq2seq model, we know that it is dedicated properly for long sequences i.e. paragraphs and sentences, however, the recurrent layer processes the input data in sequential order. These RNNs do not capture term position or order in the sentence which leads to a low term translation quality. Even So, our dataset is a comparable list of terms while the seq2seq model works better with parallel data. Overall, the manual comparison analysis we made for 100 first-aligned terms (See

Table 3) shows that transformer-based models are clearly the best choice for contextual terminology alignment. Therefore, mBERT and All-Mini-ML-6 score the highest precision(See Table 4).

5 Conclusion

In this paper, the new trending models in terminology alignment and machine translation are presented to improve the quality of alignment for several languages, especially Arabic. We chose to focus on the contextual angle of terminology alignment, to improve alignment quality. We use the ELMO model to create contextual Word vectors in order to capture terms' diversity of meaning, then use the Seq2seq model to align those vectors. We believe that the use of contextual word vectors might have a real impact on the alignment quality. We use mBERT, Seq2Seq(fast-text), and All-MiniLM-L6 Models to compare with our proposed method. Although mBERT outperforms all the models in our experiments, the results are very satisfying for the other models as well. Therefore, we think that the model we use in bilingual mapping should depend on the data size, data quality, and model parameters. In terms of future works, we are longing to create new aligned term banks, and dictionaries for other languages. We also hope to apply new models with new features to enhance the alignment quality.

References

Omar Adjali, Emmanuel Morin, and Pierre Zweigenbaum. 2022. Building comparable corpora for assessing multi-word term alignment. In *LREC 2022-Language Resources and Evaluation Conference*, pages 3103–3112.

analyticsvidhya. 2023. a simple introduction to sequence-to-sequence models. https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2,.

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2016. Learning principled bilingual mappings of word embeddings while preserving monolingual invariance. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, pages 2289–2294.

Peter F Brown, John Cocke, Stephen A Della Pietra, Vincent J Della Pietra, Frederick Jelinek, John Lafferty, Robert L Mercer, and Paul S Roossin. 1990. A statistical approach to machine translation. *Computational linguistics*, 16(2):79–85.

Steven Cao, Nikita Kitaev, and Dan Klein. 2020. Multilingual alignment of contextual word representations. *arXiv preprint arXiv:2002.03518*.

Sunipa Dev, Safia Hassan, and Jeff M Phillips. 2021. Closed form word embedding alignment. *Knowledge and Information Systems*, 63(3):565–588.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Matt Gardner, Joel Grus, Mark Neumann, Oyvind Tafjord, Pradeep Dasigi, Nelson Liu, Matthew Peters, Michael Schmitz, and Luke Zettlemoyer. 2018. Allennlp: A deep semantic natural language processing platform. *arXiv preprint arXiv:1803.07640*.

Sarthak Garg, Stephan Peitz, Udhyakumar Nallasamy, and Matthias Paulik. 2019. Jointly learning to align and translate with transformer models. *arXiv preprint arXiv:1909.02074*.

github. 2020. Elmoformanylangs. Https://github.com/HIT-SCIR/ELMoForManyLangs.

huggingface. 2023. sentence transformers all-minilm-l6-v2. https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2,.

IBM models. 2023. Types of models. https://www.ibm.com/docs/en/spss-modeler/18.1.0?topic=mining-types-models,.

Setha Imene and Aliane Hassina. 2022. An unsupervised semantic model for arabic/french terminology extraction. In *Proceedings of International Conference on Emerging Technologies and Intelligent Systems: ICETIS 2021 Volume 2*, pages 49–59. Springer.

Abraham Ittycheriah and Salim Roukos. 2005. A maximum entropy word aligner for arabic-english machine translation. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 89–96.

Georgios Kontonatsios, Ioannis Korkontzelos, Jun'ichi Tsujii, and Sophia Ananiadou. 2014. Combining string and context similarity for bilingual term alignment from comparable corpora. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Lianhau Lee, Aiti Aw, Min Zhang, and Haizhou Li. 2010. Em-based hybrid model for bilingual terminology extraction from comparable corpora. In *Coling 2010: Posters*, pages 639–646.

Jingshu Liu, Emmanuel Morin, and Sebastián Peña Saldarriaga. 2018. Towards a unified framework for bilingual terminology extraction of single-word and multi-word terms. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2855–2866.

Lieve Macken, Els Lefever, and Veronique Hoste. 2013. Texsis: Bilingual terminology extraction from parallel corpora using chunk-based alignment. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 19(1):1–30.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Robert C Moore. 2005. A discriminative framework for bilingual word alignment. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 81–88.

Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational linguistics*, 29(1):19–51.

Andraz Repar, Matej Martinc, and Senja Pollak. 2018. Machine learning approach to bilingual terminology alignment: Reimplementation and adaptation. In *4REAL 2018 Workshop on Replicability and Reproducibility of Research Results in Science and Technology of Language*, pages 1–8.

Douglas LT Rohde, Laura M Gonnerman, and David C Plaut. 2006. An improved model of semantic similarity based on lexical co-occurrence. *Communications of the ACM*, 8(627-633):116.

Masoud Jalili Sabet, Philipp Dufter, François Yvon, and Hinrich Schütze. 2020. Simalign: High quality word alignments without parallel training data using static and contextualized embeddings. *arXiv preprint arXiv:2004.08728*.

Mohammad Salameh, Rached Zantout, and Nashat Mansour. 2011. Improving the accuracy of english-arabic statistical sentence alignment. *Int. Arab J. Inf. Technol.*, 8(2):171–177.

Samuel L Smith, David HP Turban, Steven Hamblin, and Nils Y Hammerla. 2017. Offline bilingual word vectors, orthogonal transformations and the inverted softmax. *arXiv preprint arXiv:1702.03859*.

Bing Zhao and Eric Xing. 2007. Hm-bitam: Bilingual topic exploration, word alignment, and translation. *Advances in Neural Information Processing Systems*, 20.