Entailment Inference for Indonesian and Javanese Language

Jalaluddin Al-Mursyidy Fadhlurrahman[1], Ayu Purwarianti[2]
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
23521059@mahasiswa.itb.ac.id[1], ayu@staff.stei.itb.ac.id[2]

Alham Fikri Aji
Natural Language Processing Department
Mohamed bin Zayed University of Artificial Intelligence
Abu Dhabi, United Arab Emirate

***Abstract*—Natural Language Inference (NLI) is a task that determines whether a model can recognize a hypothetical sentence from a given premise sentence. An NLI model is commonly trained in one language and cannot be directly used in another. This research aims to build a model to solve low-resource cross-lingual NLI problems.**

***Keywords*—Natural Language Inference, Recognition Textual Entailment, Natural Language Processing, Low-resource NLI**

I. Introduction

Natural Language Inference (NLI) is a task that determines whether a model can infer a hypothetical sentence from the given premise sentence [34]. Researchers have used NLI as a benchmark for Natural Language Understanding (NLU). This is indicated by the existence of large-scale English language datasets such as Multi-Genre NLI (MultiNLI) which cover various genres of written and spoken text and support cross-genre evaluation [59], Stanford NLI (SNLI) [11], as well as SciTail which originates from science question answering [29].

Currently there are three Indonesian language NLI datasets, namely WReTE, INARTE, and IndoNLI. Mahendra & Setya [36] built WReTE using revision history from the Indonesian Wikipedia. On 2018, Abdiansah et al. [1] developed INARTE based on question-answer pairs from Web data. Finally on 2021, Mahendra et al. [35] built IndoNLI using a human-elicitation approach and has 18,000 pairs of sentences annotated by the public and experts.

An NLI model is commonly trained in one language and cannot be directly used in another [16]. In reality, a multinational system needs to handle input from multiple languages. This rules out the option of annotating an entire language. One way to create a multilingual system is through cross-lingual language understanding. A model is trained in one language and will be evaluated in another.

Common problems in NLI models cover various aspects such as premise-hypothesis sentence pairs that contain antonyms, reasoning involving numbers, overlapping words, the presence of strong negation words, premise sentences that are longer than the hypothesis sentence, and spelling errors, lexical overlap, inference without lexical and real-world knowledge, and syntactic structures [25], [37], [46], [49].

The main problem in cross-language NLI is that texts in different languages have different semantic, syntactic, and grammatical structures [4]. Model understanding of sentences no longer revolves around translation. Research conducted by Mohammad et al. [44] and Smith et al. [54] showed that the decrease in the performance of cross-language models occurs due to differences in word choice, causing the semantic relationship between sentences to be lost. Another problem in cross-language NLI is that if the language used is a low resource language, not only the word representation loses semantic relationships, but also coarse-grained relationships and lexicon induction will be lost [60]. Therefore, there is a need for further research regarding entailment inference which can maintain semantic relationships between sentences, especially in Indonesian as a high resource language and Javanese as a low resource language.

II. Related Work

A. Natural Language Inference

Reasoning and concluding are fundamental tasks in Natural Language Understanding [34]. Natural Language Inference (NLI) specifically aims to determine a hypothesis sentence h can be concluded from the premise sentence p. Conneau et al. [17] provide examples of hypothesis sentence pairs and premise sentences along with their labels, as can be seen in Table I. The example sentence pairs shown in Table 1 are a form of recognizing textual entailment.

TABLE I: Examples of premise-hypothesis sentence pairs and its labels [17].

Premise	**Hypothesis**	**Label**
You can't stay here	You can leave	Entailment
Conceptually cream skimming has two basic dimensions - product and geography	Product and geography are what make cream skimming work	Neutral
At the end of Rue des Francs-Bourgeois is what many consider to be the city 's most handsome residential square, the Place des Vosges, with its stone and red brick facades	Place des Vosges is constructed entirely of gray marble	Contradict

B. Recognizing Text Entailment

Recognizing Textual Entailment (RTE) is an NLI challenge aiming to evaluate the ability of machines to obtain background knowledge and the reasoning abilities needed by humans to determine whether a text can be inferred from other texts [18]. The most common challenge is giving a premise and a hypothesis to a system, and the system will predict whether the hypothesis can be concluded from the premise. In 2008 and 2009, another challenge emerged as a three-way decision task where the system will identify contradictory sentence relationships between a text and hypothetical sentences [7], [23]. The next challenge in 2010 and 2011 was a hypothetical sentence that could be deduced from several sentences in a corpus [5], [6]. In answering the NLI challenge, a model needs knowledge sources to be able to provide arguments.

C. Linguistic Knowledge Source

NLI is a principal part of natural language understanding (NLU). Therefore, linguistic knowledge sources are crucial for computers to identify the syntactic and semantic structures of languages. This structure also often needs to be supplemented by general knowledge to gain a more thorough understanding

Knowledge sources that provide annotations for semantic, syntactic, and discourse structures have been developed for Machine Learning modeling. Several knowledge bases are also available for semantic processing. In addition, researchers have developed various embedding vectors of word meanings. This section discusses sources of linguistic knowledge such as lexical, semantic frames, and pre-trained semantic vectors for linguistic knowledge.

D. Lexical Source

Lexical resources for the most commonly used verbs, adverbs, nouns, and adjectives are available at WordNet [41]. WordNet is capable of organizing words in terms of concepts, such as lists of synonyms and semantic relationships between words. Another lexical source is VerbOcean [14], which can capture the more subtle relationships between smaller sets of common verbs. In addition to annotating large corpora or lexical hierarchies, semantic relationships in verbs can also be captured by frames.

E. Frame Semantics

Frame, according to Minsky [42], is a cognitive data structure for information about prototype events and situations. Research conducted by Fillmore et al. [22] produced FrameNet which provides a database of semantic frames that describes situations, information, relations, and sentences annotated by frame elements. FrameNet also provides a lexical database of nouns, adjectives, and verbs that are paired with frames. This FrameNet then became the basis for research on semantic role labeling [25], [50]. Another approach to linguistic knowledge resources is with pre-trained semantic vectors.

F. Pre-trained Semantic Vector

The initial approach evolved from using n-gram counts to statistical language modeling by Brown et al. [12] instead of using these calculations to produce a spare semantic vector [32]. Subsequent research by Mikolov, Chen, et al. [39] produced word2vec, which trained a neural network on classifying word occurrences generated from large text. The training result weights then used as a dense vector representation of words, which can then be used as a source to represent words in the neural network input for other tasks. In addition to word2vec, the result of research by Pennington et al. [47], Glove, is an example of a well-known pre-training vector increasing efficiency and performance by training the model only on words with non-zero occurrences. Another subsequent research by Bojanowski et al. [9] made use of sub-word information to better deal with rare words that were not widely available in training data. This is achieved by splitting the training text into characters when training a semantic vector, such as with FastText.

There are advantages to using pre-trained semantic vectors. Similar words appear in the text simultaneously, and are trained based on their occurrence. This vectorization causes the vectors of similar words to appear close together in the semantic vector space, so that further word meanings can be captured [39].

The main drawback of these vectors is that the vector representation of a word is always the same regardless of the context. The use of a large parallel corpus causes gender, racial, and cultural stereotypes [10]. The other weakness is that word embedding can only be generated from words that already exist in the training corpus [39]. If a word does not exist in the training corpus, then the word does not have an embedding vector. This can cause the model to have domain constraints. Moreover, word embedding is difficult to interpret the meaning of each embedding. This can be a problem if the model makes an error, or the embedding is used in a domain that emphasizes interpretation [31].

G. Sentence Embedding

Another approach to conducting NLI is to use learning with sentence representation [2], [8], [58]. One way to get sentence embedding is to do a weighted average of word representations, which is often called a Continuous Bag-of-Words (CBOW) [16]. Another way to get sentence embedding is to use the SkipThought model, which is a study by Kiros et al. [30], and the continuation of the skip-gram model from research by Mikolov, Chen, et al. [38]. This SkipThought model can capture semantic and syntactic dependencies on sentence representation [30]. Research by Bandyopadhyay et al. [4] and Howard & Ruder [27] show that the use of pre-trained language models can work well when using the hidden state as a contextual vector, or when using the entire model for fine-tuning.

In a study conducted by Espana-Bonet et al. [20] and Schwenk & Douze [53], the method of presenting sentences between languages was carried out by training a sequence-to-sequence Machine Translation system in several languages to study the space for sentence embedding across languages. Moritz Hermann & Blunsom [45] in their research proposed a vector model to study representations at the document level using unigrams and bigrams. Zhou et al. [61] in their research study the representation of bilingual documents by minimizing the Euclidean distance between a document and its translation.

The sentence embedding method has a weakness, namely the limited context. The words in a sentence could form a sentence embedding vector but do not describe the context of the entire document [52]. In addition, sentence embedding also has difficulty understanding long sentences. This is caused by missing information when applied to long sentences, especially when the main information is situated at the beginning or end of a sentence [13]. Similar to word embedding, according to Conneau et al. [15], sentence embedding is also usually trained in a general corpus, which does not capture the nuances of language in a particular domain. This causes the model to not work well in specific domains, such as medical or legal. Cer et al. [13] in their research found that negation words are difficult to capture with sentence embedding because the meaning of a sentence can change drastically depending on whether a word is negated or not. All of these drawbacks led researchers to use contextual models as a technique for conducting NLI.

H. Contextual Model

Recent studies on NLI focus on increasing the accuracy of the model by using contextual information. Contextual models use pre-trained language models, such as BERT, to generate representations of words and sentences that take into account the surrounding context. Mahendra et al. [35] found in their research that the model's performance against cross-language NLI challenges, the performance produced by BERT was below XLM-Roberta (XLMR). This happens because the size of the pre-training data on XLMR has a larger portion of Indonesian compared to BERT.

Hu et al. [28], Tandon et al. [57], Qi et al. [51], and Soudani et al. [56] in their study using XLMR as a baseline model gave similar results to research by Mahendra et al. [35], where XLMR provides the highest accuracy results. This is because XLMR performs cross-language learning tasks using a 42 GB parallel corpus.

However, this contextual model has not been able to solve the classic problems in cross-language NLI. In the case of low-resource NLI, word representations not only lose their semantic relationships but also coarse-grained relationships and lexicon induction [60]. This problem makes the case of low-resource NLI an interesting topic for research.

References

[1] Abdiansah, Azhari, & Sari, A. K. (2018). INARTE: An Indonesian Dataset for Recognition Textual Entailment. 4th International Conference on Science and Technology (ICST), 1–5.

[2] Arora, S., Liang, Y., & Ma, T. (2017). A Simple But Tough-To-Beat Baseline For Sentence Embeddings. ICLR, 1–16.

[3] Artetxe, M., Labaka, G., & Agirre, E. (2017). Learning bilingual word embeddings with (almost) no bilingual data. ACL 2017 - 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers), 1, 451–462. https://doi.org/10.18653/v1/P17-1042

[4] Bandyopadhyay, D., De, A., Gain, B., Saikh, T., & Ekbal, A. (2022). A Deep Transfer Learning Method for Cross-Lingual Natural Language Inference. Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), 20–25. https://github.com/google-research/

[5] Bentivogli, L., Clark, P., Dagan, I., & Giampiccolo, D. (2010). The Sixth PASCAL Recognizing Textual Entailment Challenge. Proceedings of the Third Text Analysis Conference (TAC 2010).

[6] Bentivogli, L., Clark, P., Dagan, I., & Giampiccolo, D. (2011). The Seventh PASCAL Recognizing Textual Entailment Challenge. Proceedings of the Fourth Text Analysis Conference (TAC 2011).

[7] Bentivogli, L., Dagan, I., Dang, H. T., Giampiccolo, D., & Magnini, B. (2009). The Fifth PASCAL Recognizing Textual Entailment Challenge. Proceedings of the Second Text Analysis Conference (TAC 2009), 15.

[8] Blacoe, W., & Lapata, M. (2012). A Comparison of Vector-based Representations for Semantic Composition. Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, 546–556.

[9] Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching Word Vectors with Subword Information. Transactions of the Association for Computational Linguistics, 5, 135–146. http://www.isthe.com/chongo/tech/comp/fnv

[10] Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., & Kalai, A. (2016). Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. Advances in Neural Information Processing Systems, 4349–4357. http://arxiv.org/abs/1607.06520

[11] Bowman, S. R., Angeli, G., Potts, C., & Manning, C. D. (2015). A large annotated corpus for learning natural language inference. http://arxiv.org/abs/1508.05326

[12] Brown, P. F., deSouza, P. V, Mercer, R. L., Della Pietra, V. J., & Lai, J. C. (1992). Class-Based n-gram Models of Natural Language. Computational Linguistics, 18(4), 467–480.

[13] Cer, D., Yang, Y., Kong, S., Hua, N., Limtiaco, N., St John, R., Constant, N., Guajardo-Céspedes, M., Yuan, S., Tar, C., Sung, Y.-H., Strope, B., & Kurzweil Google Mountain View, R. A. (2018). Universal Sentence Encoder for English. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 169–174.

[14] Chklovski, T., & Pantel, P. (2004). VerbOcean: Mining the Web for Fine-Grained Semantic Verb Relations. Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP 2004), 33–40.

[15] Conneau, A., Kiela, D., Schwenk, H., Barrault, L., & Bordes, A. (2017). Supervised Learning of Universal Sentence Representations from Natural Language Inference Data. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 670–680. http://arxiv.org/abs/1705.02364

[16] Conneau, A., Rinott, R., Lample, G., Schwenk, H., Stoyanov, V., Williams, A., & Bowman, S. R. (2018a). XNLI: Evaluating Cross-lingual Sentence Representations. Association for Computational Linguistics.

[17] Conneau, A., Rinott, R., Lample, G., Schwenk, H., Stoyanov, V., Williams, A., & Bowman, S. R. (2018b). XNLI: Evaluating Cross-lingual Sentence Representations. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2475–2485.

[18] Dagan, I., Glickman, O., & Magnini, B. (2006). The PASCAL Recognising Textual Entailment Challenge. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 3944 LNAI, 177–190.

[19] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. http://arxiv.org/abs/1810.04805

[20] Espana-Bonet, C., Varga, A. C., Barron-Cedeno, A., & Van Genabith, J. (2017). An Empirical Analysis of NMT-Derived Interlingual Embeddings and Their Use in Parallel Sentence Identification. IEEE Journal on Selected Topics in Signal Processing, 11(8), 1340–1350. https://doi.org/10.1109/JSTSP.2017.2764273

[21] Faruqui, M., & Dyer, C. (2014). Improving Vector Space Word Representations Using Multilingual Correlation. Proceedings of the 14th Con-

ference of the European Chapter of the Association for Computational Linguistics, 462–471.
[22] Fillmore, C. J., Baker, C. F., & Sato, H. (2002). The FrameNet Database and Software Tools. Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02), 1157–1160. http://framenet.icsi.berkeley.edu/
[23] Giampiccolo, D., Dang, H. T., Magnini, B., Dagan, I., & Dolan, B. (2008). The Fourth PASCAL Recognizing Textual Entailment Challenge. Proceedings of the First Text Analysis Conference (TAC 2008), 9.
[24] Gildea, D., & Jurafsky, D. (2000). Automatic Labeling of Semantic Roles. Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics, 512–520.
[25] Glockner, M., Shwartz, V., & Goldberg, Y. (2018). Breaking NLI Systems with Sentences that Require Simple Lexical Inferences. http://arxiv.org/abs/1805.02266
[26] Gouws, S., Bengio, Y., & Corrado, G. (2015). BilBOWA: Fast Bilingual Distributed Representations without Word Alignments. Proceedings of the 32 Nd International Conference on Machine Learning, 37, 748–756. http://arxiv.org/abs/1410.2455
[27] Howard, J., & Ruder, S. (2018). Universal Language Model Fine-tuning for Text Classification. http://arxiv.org/abs/1801.06146
[28] Hu, H., Richardson, K., Xu, L., Li, L., Kuebler, S., & Moss, L. S. (2020). OCNLI: Original Chinese Natural Language Inference. Findings of the Association for Computational Linguistics: EMNLP 2020, 3512–3526. http://arxiv.org/abs/2010.05444
[29] Khot, T., Sabharwal, A., & Clark, P. (2018, April 27). SCITAIL: A Textual Entailment Dataset from Science Question Answering. Proceedings of the AAAI Conference on Artificial Intelligence.
[30] Kiros, R., Zhu, Y., Salakhutdinov, R., Zemel, R. S., Torralba, A., Urtasun, R., & Fidler, S. (2015). Skip-Thought Vectors. Advances in Neural Information Processing Systems, 3294–3302. http://arxiv.org/abs/1506.06726
[31] Li, J., Chen, X., Hovy, E., & Jurafsky, D. (2018). Visualizing and Understanding Neural Models in NLP. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2, 91–96. http://arxiv.org/abs/1506.01066
[32] Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. Behavior Research Methods, Instruments, & Computers, 28(2), 203–208.
[33] Luong, M.-T., Pham, H., & Manning, C. D. (2015). Bilingual Word Representations with Monolingual Quality in Mind. Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing, 151–159.
[34] Maccartney, B., & Manning, C. D. (2008). Modeling Semantic Containment and Exclusion in Natural Language Inference. Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008), 521–528.
[35] Mahendra, R., Aji, A. F., Louvan, S., Rahman, F., & Vania, C. (2021). IndoNLI: A Natural Language Inference Dataset for Indonesian. https://doi.org/10.18653/v1/2021.emnlp-main.821
[36] Mahendra, R., & Setya, K. N. (2018). Semi-supervised Textual Entailment on Indonesian Wikipedia Data Indonesian Morphological Analysis View project Consumer Health Question Answering System for Indonesian Language View project. https://doi.org/10.13140/RG.2.2.18820.27521
[37] Mccoy, R. T., Pavlick, E., & Linzen, T. (2019). Right for the Wrong Reasons: Diagnosing Syntactic Heuristics in Natural Language Inference. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 3428–3448.
[38] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013a). Distributed Representations of Words and Phrases and their Compositionality.
[39] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013b, January 16). Efficient Estimation of Word Representations in Vector Space. Proceedings of the 1st International Conference on Learning Representations (ICLR 2013). http://arxiv.org/abs/1301.3781
[40] Mikolov, T., Le, Q. V., & Sutskever, I. (2013). Exploiting Similarities among Languages for Machine Translation. http://arxiv.org/abs/1309.4168
[41] Miller, G. A. (1995). WordNet: A Lexical Database for English George A. Miller. Communications of the ACM, 38(11), 39–41.
[42] Minsky, M. (1975). Framework for Knowledge Representation. McGraw-Hill.
[43] Mogadala, A., & Rettinger, A. (2016). Bilingual Word Embeddings from Parallel and Non-parallel Corpora for Cross-Language Text Classification. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 692–702.
[44] Mohammad, S. M., Salameh, M., & Kiritchenko, S. (2016). How Translation Alters Sentiment. Journal of Artificial Intelligence Research, 55, 95–130.
[45] Moritz Hermann, K., & Blunsom, P. (2014). Multilingual Models for Compositional Distributed Semantics. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 1, 58–68.
[46] Naik, A., Ravichander, A., Sadeh, N., Rose, C., & Neubig, G. (2018, June 2). Stress Test Evaluation for Natural Language Inference. http://arxiv.org/abs/1806.00692
[47] Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014), 1532–1543.
[48] Peters, M. E., Neumann, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2227–2237.
[49] Poliak, A., Haldar, A., Rudinger, R., Hu, J. E., Pavlick, E., White, A. S., & Van Durme, B. (2018). Collecting Diverse Natural Language Inference Problems for Sentence Representation Evaluation. http://arxiv.org/abs/1804.08207
[50] Punyakanok, V., Roth, D., Yih, W.-T., & Zimak, D. (2004). Semantic Role Labeling via Integer Linear Programming Inference. Proceedings of the 20th International Conference on Computational Linguistics, 1346–1352.
[51] Qi, K., Wan, H., Du, J., & Chen, H. (2022). Enhancing Cross-lingual Natural Language Inference by Prompt-learning from Cross-lingual Templates (Vol. 1). Long Papers.
[52] Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 3982–3992. http://arxiv.org/abs/1908.10084
[53] Schwenk, H., & Douze, M. (2017). Learning Joint Multilingual Sentence Representations with Neural Machine Translation. Proceedings of the 2nd Workshop on Representation Learning for NLP, 157–167.
[54] Smith, L. K., Giorgi, S., Solanki, R., Eichstaedt, J. C., Schwartz, H. A., Abdul-Mageed, M., Buffone, A., & Ungar, L. H. (2016). Does "well-being" translate on Twitter? Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processings, 2042–2047.
[55] Søgaard, A., Zeljko Agi, ˇ, Martínez Alonso, H., Plank, B., Bohnet, B., & Johannsen, A. (2015). Inverted indexing for cross-lingual NLP. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, 1, 1713–1722.
[56] Soudani, H., Mojab, M. H., & Beigy, H. (2022). Persian Natural Language Inference: A Meta-learning approach. Proceedings of the 29th International Conference on Computational Linguistics, 4306–4319.
[57] Tandon, C., Bongale, P. J., Palivela, H., R, S. R., & R, N. C. (2021). UT-ATD: Universal Transformer for Anomalous Trajectory Detection by Embedding Trajectory Information. Journal of Visual Language and Computing, 2021(1), 27–32. https://doi.org/10.18293/jvlc2021-n1-011
[58] Wieting, J., Bansal, M., Gimpel, K., & Livescu, K. (2016). Towards Universal Paraphrastic Sentence Embeddings. 1–14. http://arxiv.org/abs/1511.08198
[59] Williams, A., Nangia, N., & Bowman, S. R. (2017). A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1112–1122. http://arxiv.org/abs/1704.05426
[60] Zhang, Y., Gaddy, D., Barzilay, R., & Jaakkola, T. (2016). Ten Pairs to Tag-Multilingual POS Tagging via Coarse Mapping between Embeddings. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1307–1317.
[61] Zhou, X., Wan, X., & Xiao, J. (2016). Cross-Lingual Sentiment Classification with Bilingual Document Representation Learning. Proceedings

of the 54th Annual Meeting of the Association for Computational Linguistics, 1403–1412.