SemanticCUETSync at SemEval-2024 Task 1: Finetuning Sentence Transformer to Find Semantic Textual Relatedness

Md. Sajjad Hossain, Ashraful Islam Paran, Symom Hossain Shohan, Jawad Hossain and **Mohammed Moshiul Hoque**
Department of Computer Science and Engineering
Chittagong University of Engineering and Technology
{u1904031,u1904029,u1904048,u1704039}@student.cuet.ac.bd
moshiul_240@cuet.ac.bd

Abstract

Semantic textual relatedness is crucial to Natural Language Processing (NLP). Methodologies often exhibit superior performance in high-resource languages such as English compared to low-resource ones like Marathi, Telugu, and Spanish. This study leverages various machine learning (ML) approaches, including Support Vector Regression (SVR) and Random Forest, deep learning (DL) techniques such as Siamese Neural Networks, and transformer-based models such as MiniLM-L6-v2, Marathi-sbert, Telugu-sentence-bert-nli, and Roberta-bne-sentiment-analysis-es, to assess semantic relatedness across English, Marathi, Telugu, and Spanish. The developed transformer-based methods notably outperformed other models in determining semantic textual relatedness across these languages, achieving a Spearman correlation coefficient of 0.822 (for English), 0.870 (for Marathi), 0.820 (for Telugu), and 0.677 (for Spanish). These results led to our work attaining rankings of 22$^{\text{th}}$ (for English), 11$^{\text{th}}$ (for Marathi), 11$^{\text{th}}$ (for Telegu) and 14$^{\text{th}}$ (for Spanish), respectively.

1 Introduction

Semantic textual relatedness measures the conceptual and contextual similarity of two sentences. It specifies how alike the two sentences are in terms of meaning. Determining semantic textual relatedness is crucial for various language-processing tasks, including contemporary technology, search engines, chatbots, virtual assistants, plagiarism detection, paraphrasing, question answering, text generation, and other related applications. It is possible to determine how comparable two natural language sentences are based on the quantity and quality of matched elements in each sentence. These matches offer essential insights into the relationship between and degree of semantic similarity between the two sentences and the likelihood of successful word matching in semantically equivalent text pairs.

Another critical aspect of semantic relatedness is understanding the context of sentences. Singnificant research has been done in this area, specifically in SemEval competition from 2012 to 2017 (Agirre et al., 2012, 2013, 2014, 2015, 2016; Cer et al., 2017). Most proposed methodologies perform better in high-resource languages like English and Spanish but could be better in other low-resource languages like Telegu, Marathi, and Bengali. In certain studies, translating from a low-resource language to English is conducted (Wu et al., 2017) prior to semantic analysis, contributing to suboptimal performance. Other factors include the failure of specific methods to capture the meaning of phrases and idioms within sentences effectively (Śpiewak et al., 2017), as well as a tendency for some approaches to focus excessively on individual word meanings (Śpiewak et al., 2017). This work addresses these shortcomings by proposing a transformer-based model, which gives us a good result in finding semantic textual relatedness. The main contributions of this work are:

- Investigated several ML, DL, and transformer-based models to find the semantic relatedness in various texts of four languages.
- Explored various pre-trained transformer-based methods with necessary tuning to identify semantic textual relatedness in English, Spanish, Telegu, and Marathi.

The code will be publicly available at https://github.com/ashrafulparan2/SemanticCUETSync-at-SemEval-2024-Task-1.

2 Related Work

Numerous studies have been accomplished on semantic relatedness in high-resource languages. Hasan et al., 2020 presented the process for calculating semantic similarity and proposed a feature-based metric for building semantic vectors. Their

Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024), pages 1222–1228
June 20-21, 2024

knowledge feature-based method found similarity measure of 0.82. Abdalla et al., 2021 proposed a dataset to explore questions on what makes sentences semantically related. Reimers and Gurevych, 2019 suggested using a sentence transformer model, which creates a 768-dimensional dense vector space from sentences and paragraphs. The relatedness of two sentences can be assessed using this model. Their approach achieved the best score of 0.8492 with SRoBERTa-STSb-base model. Three transformer-based clinical semantic textual similarity models intended to detect semantic relatedness in medical data were presented by Yang et al., 2020. Chen et al., 2022 proposed a semi-supervised sentence embedding technique called GenSE, which effectively uses large-scale unlabeled data. It achieved promising results on several STS datasets with an average correlation score of 0.8519. Meanwhile, Gatto et al., 2023 evaluated sentence similarity among texts using large language models (LLM). According to their research, ChatGPT and other models are proficient in recognizing textual similarity within particular areas. In addition to the transformer models, Verma and Muralikrishna, 2020 introduced a deep learning model, specifically a Recurrent Neural Network (RNN). This model utilized document embedding vectors to infer the meaning of small paragraphs comprising one, two, or three sentences.

Significant challenges arise when representing sentences with low-resource languages (LRLs), such as Telugu, Marathi, and Bengali. Furthermore, limited datasets make the process of textual similarity detection more challenging in LRLs. Joshi et al., 2023 focused on two low-resource Indian languages (Hindi and Marathi), and their proposed model was evaluated on real text classification datasets to show embeddings obtained from synthetic data, which will be an effective training strategy for low-resource languages. They achieved the highest score of 0.83 utilizing the MahaBERT and LaBSE models. A cross-lingual model for finding similarity between sentences was proposed by Deode et al., 2023a. Their system obtained an accuracy of 0.82 for finding semantic relatedness in low-resource languages like Hindi, Marathi, Kannada, Telugu, Malayalam, Tamil, Gujarati, Odia, Bengali, and Punjabi. Tang et al., 2018 proposed a multilingual framework for finding semantic textual similarity in low-resource languages utilizing rich annotation data from a high-resource language. Their shared sentence encoder approach archived score of 0.825 for Spanish language.

3 Dataset and Task Description

The dataset is developed by Ousidhoum et al. (Ousidhoum et al., 2024a) to evaluate the semantic textual similarity to perform the shard task at SemEval-2024. It includes data for Telugu, Marathi, English, and Spanish, which assess semantic relatedness between sentences. There are labeled and unlabeled data in the dataset. Labeled data is used for Track-A; each row has two sentences and a score that ranges from 0 to 1, representing the semantic relatedness of the sentences. Moreover, the dev, test, and train categories are applied to every language dataset. Table 1 shows the distribution of dev, test, and train sets for the English, Marathi, Telugu, and Spanish datasets, respectively.

Language	Dev set	Test set	Train set
English	250	2600	11000
Marathi	294	299	2400
Telegu	130	297	2340
Spanish	140	600	1562
Total	814	3796	17302

Table 1: Dataset statistics for Track-A

The task for Track-A (Ousidhoum et al., 2024b) calculates the semantic relationship and provides a score (degree of semantic relatedness) between 0 to 1. Figure 1 illustrates few examples of the task of sentence relatedness with the score.

4 System Overview

Various textual features will be extracted to employ ML and DL models. Several transformer models are exploited for the task, including MiniLM-L6-v2, Marathi-SBERT, Telugu-Sentence-BERT-NLI, and Roberta-BNE-Sentiment-Analysis-ES, to assess semantic relatedness. Figure 2 illustrates the schematic structure of the proposed approach.

Textual Feature Extraction: Feature extraction is necessary for ML and DL models to learn from text. We used TF-IDF (Takenobu, 1994) to extract the features to apply different ML algorithms. Word2Vec (Mikolov et al., 2013) and FastText (Grave et al., 2018) embeddings were used to extract features for DL models.

ML-based Approaches: This work employed traditional ML approaches, including SVR and RF. Following dataset preprocessing, we trained

Sentence #1	Sentence #2	Score
I very much enjoy Billy Talent. They are one of my favorite groups.	I just love Billy Talent, they are one of my favourites.	0.97
El cocinero estÃ¡ rociando queso sobre la pizza.	Un hombre estÃ¡ poniendo un poco de queso en una pizza.	0.57
పశ్చిమబంగలో రైల్వే ఇరుసు కర్మాగారాన్ని నెలకొల్పనున్నట్లు ఆయన వెల్లడించారు	ఈ దాడులు రాష్ట్రవ్యాప్తంగా కొనసాగుతాయని ఈ సందర్భంగా విజిలెన్స్ డీజీ స్పష్టం చేశారు.	0.37
या विश्वचषकानंतर जयसूर्याने मागे वळून पाहिले नाही.	त्यातूनच तिने इसिसकडे भारतातील मुस्लिमांना आकृष्ट करण्याचे काम सुरू केले.	0.04

Figure 1: Track-A task sample with Similarity score for English, Spanish, Telugu and Marathi

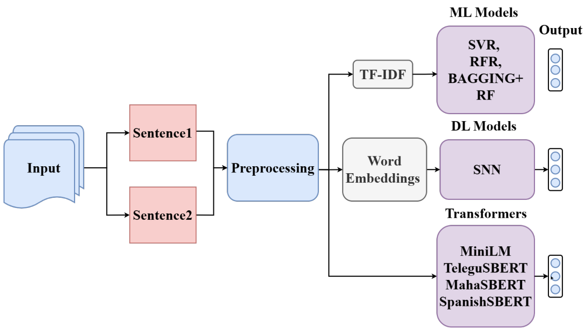

Figure 2: Schematic process of finding semantic textual relatedness

the model using the SVR for the English language. Similarly, we utilized the RF, configuring "n-estimators = 100" during training. A BaggingRegressor model is employed with the RF as the base estimator, with "n-estimators = 10" set for the base estimator.

DL-based Approaches: This work explored a Siamese Neural Network (SNN) architecture implemented in PyTorch to perform the task. The SNN model was applied exclusively to the English dataset. The text is transformed into vectors and subsequently passed through LSTM layers. The similarity between the two texts was determined using cosine embedding loss, and optimization was carried out using the 'Adam' optimizer. We set learning rate = 10, hidden-dim = 128, embedding-dim = 128 and max-length = 1000 and ran it for 10 epochs. Table 2 shows the several ML parameters and DL hyperparameters used for the models.

Classifier	Parameters	Value
SVR	kernel	rbf
	degree	3
	gamma	scale
	C	1.0
RF	n-estimator	100
Bagging+RF	n-estimator	10
SNN	epoch	10
	lr	9e-5
	hidden-dim	128
	embedding-dim	128
	max-words	10000
	max-length	1000

Table 2: Several ML parameters and DL hyperparameters of the employed models

Transformer-based Models: This work developed the task solutions in four languages: English, Marathi, Telugu, and Spanish. Thus, four different kinds of transformers were explored, including all-MiniLM-L6-v2 (Wang et al., 2020), telugu-sentence-bert-nli (Deode et al., 2023b), marathi-sentence-similarity-sbert (Joshi et al., 2023) and dccuchile-bert-base-spanish-wwm-uncased (Cañete et al., 2023). MiniLM-L6-v2 is a sentence transformer that maps sentences and paragraphs to a 384-dimensional dense vector space and can be used for tasks like clustering or semantic search. Similarly, telugu-sentence-bert-nli is a TeluguBERT model trained on a large dataset. MahaSBERT(marathi-sentence-similarity-sbert) is also fine-tuned on the STS dataset. Before applying these models, we have used other transformer models such as msmarco-distilbert-cos-v5, all-mpnet-base-v2, and all-MiniLM-L12-v2 for English. For evaluating Marathi, we also used pre-trained models like stsb-xlm-r-multilingual , marathi-roberta, and marathi-sentence-bert-nli. Similarly, we have fine-tuned transformer models like LaBSE and telugu-sentence-similarity-sbert for Tamil. Finally, for Spanish, we also used roberta-bne-sentiment-analysis-es and stsb-xlm-r-multilingual. We called all these models from `Huggingface`[1] sentence transformers library. All models were trained on the task datasets. MahaSBERT and TeluguBERT usually perform well for low-resource languages. Dataloader was used from a torch to prepare the data before passing it to the model. We

[1] https://huggingface.co/sentence-transformers

followed a similar approach to train the corresponding transformer model for all languages. Table 3 demonstrates the hyperparameters used to train transformer-based models.

Models	**LR**	**WD**	**WS**
all-Minilm-L6-v2	9e-5	9e-2	750
all-Minilm-L12-v2	9e-5	9e-2	750
marathi–sbert	9e-5	5e-2	500
telugu–bert-nli	9.5e-5	5e-5	750
bert-base-spanish	9e-5	9.5e-7	700

Table 3: Tuned hyperparameter for the transformer-based models, where LR, WD, and WS denotes learning rate, weight decay and warmup steps, respectively

5 Experiments

During the development, this study utilized Python 3 (3.10.12) and Python-based packages from `PyTorch`[2] framework to implement sentence transformers (MiniLM-L6, MarathiSBERT, TeluguSBERT, SpanishSBERT). To implement the models, 29GB of RAM, 16GB of VRAM, and 73.1GB of storage space were used. We utilized NVIDIA Tesla P100 GPU from `Kaggle`[3]. We used pandas (2.1.4) and numpy (1.24.3) to analyze and prepare the data. The ML models were developed with the scikit-learn (1.2.2) packages, and the DL models were trained with Keras (2.13.1) and TensorFlow (2.13.0). The PyTorch (2.0.0) packages, transformers (4.36.2), and sentence transformers (2,6,1) were used to implement transformer models.

The superiority of the models is determined based on the Spearman rank correlation coefficient (ρ) (Sennrich et al., 2015), which measures how well the system predicted rankings of test instances. This work also measures the Kendall correlation (τ) and Pearson correlation (R).

6 Results and Analysis

Table 4 exhibits the evaluation results of ML, DL, and transformer-based models for four languages: English, Marathi, Telegu, and Spanish.

The results demonstrate that the ML models perform poorly. DL models are slightly better, but they need to be better. Transformer-based models demonstrated exceptional performance across all languages. For the English language, Mpnet-v2, Distilbert, and MiniLM-L12 scored 0.821, 0.821, and 0.815, respectively. The top-performing model for English was MiniLM-L6, with a maximum score of 0.822. For the Marathi language, Stsb-xlm, Marathi-BERT-nli and MarathiRoberta received scores of 0.764, 0.866, and 0.810, respectively. The MarathiSBERT model performed best, with a Spearman correlation score of 0.870. For the Telugu language, Telugu-SBERT and LaBSE had scores of 0.761 and 0.804, respectively. The best model for this language is Telugu-BERT-nli, which has a Spearman correlation rank of 0.804. For Spanish, Roberta-bne and Stsb-xlm have scores of 0.659 and 0.677, respectively, but Spanish-BERT outperforms both by 0.677. Figure 3 illustrates the summary of the best-performed models in four languages of the task.

Language	**Models**	ρ	τ	R
English	SVR	0.161	0.105	0.181
	RF	0.177	0.115	0.153
	Bagging + RF	0.178	0.114	0.156
	SNN	0.418	0.284	0.473
	MiniLM-L12	0.815	0.614	0.821
	Mpnet-v2	0.821	0.620	0.832
	Distilbert	0.821	0.619	0.829
	MiniLM-L6	**0.822**	0.620	0.832
Marathi	Stsb-xlm	0.764	0.566	0.779
	Marathi-Roberta	0.810	0.619	0.810
	Marathi-BERT-nli	0.866	0.684	0.866
	Marathi-SBERT	**0.870**	0.688	0.875
Telegu	Telugu-SBERT	0.761	0.567	0.795
	LaBSE	0.804	0.608	0.814
	Telugu-BERT-nli	**0.820**	0.617	0.827
Spanish	Roberta-bne	0.659	0.479	0.713
	Stsb-xlm	0.655	0.473	0.707
	Spanish-BERT	**0.677**	0.503	0.719

Table 4: Performance of the employed models on the test set

Transformer-based outperformed other models

[2]https://pytorch.org/

[3]https://www.kaggle.com/

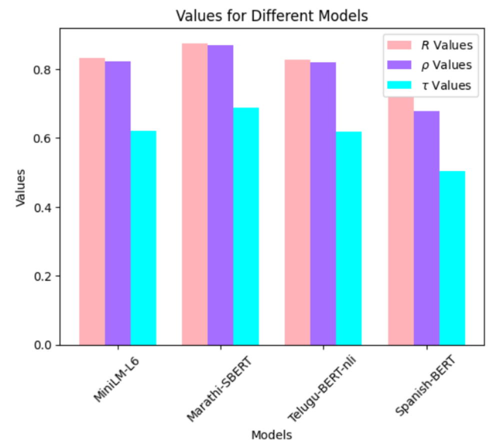

Figure 3: Performance summary of the best model in each task language

by a wide margin for four languages in the task. One explanation is its ability to capture information's bidirectional context. In addition, it produces contextual word representations that make polysemy understandable and enables capturing minute variations in meaning depending on context. Since these transformer-based model has been pre-trained on various linguistic tasks using a sizable dataset, fine-tuning these models improves results.

6.1 Error Analysis

Transformer models can recognize sentence patterns more apparently with better textual features. Thus, these models outperformed ML and DL techniques. Figure 4 depicts some example scores (actual and predicted) regarding two sentences in task languages.

Based on annotations using Best-Worst Scaling, actual scores were computed by deducting the number of times a phrase pair was selected as the least related from the fraction of times it was selected as the most related. The predicted and actual scores are incredibly close in samples 1, 2, and 4. However, the predicted score is more significant for sample 3. This may happen when a sentence is shorter than the longer sentence and contains similar terms. In this case, the model has a more challenging time figuring out the Spearman correlation between these two uneven-length sentences.

7 Conclusion

This study explores the efficacy of various machine learning (ML), deep learning (DL), and transformer-based models for analyzing semantic relatedness within texts across four languages: English, Spanish, Telugu, and Marathi. Experimental assessments reveal subpar performance of both ML and DL models across all languages. However, transformer-based models exhibit superior capabilities in discerning semantic relatedness within the given task. Specifically, the MiniLM-L6 model excels for English, MarathiSBERT for Marathi, TeluguSBERT for Telugu, and SpanishSBERT for Spanish, achieving peak ρ scores of 0.822, 0.870, 0.820, and 0.677, respectively. The study suggests that augmenting training data could enhance the performance of current models. Additionally, leveraging advanced techniques such as Large Language Models (LLM) and Generative Pre-trained Transformers (GPT) holds promise for further improvement.

Sentence #1	Sentence #2	AS	PS
Egypt's Brotherhood stands ground after killings	Egypt: Muslim Brotherhood Stands Behind Morsi	0.7	0.69
Los menonitas amish conascendencia suiza de Galicia se establecieronen 1815 cercade Dubno.	Los amonios menonitas de origen suizo de Galicia se establecieron cerca de Dubno en 1815.	0.8	0.82
మాస్ మహారాజ సినిమాలో మరో ప్రముఖ నటుడు.	మాస్ మహారాజా రవితేజ ఈ వేసవిలో ‘క్రాక్’ సినిమాతో సందడి చేయాలనుకుంటే తన స్పీడుకు కోవిడ్ 19 బ్రేకులేసింది.	0.38	0.52
सोनूसोबत ज्युनियर कलकत्ता नावाचा बुकीही तेव्हा रॅकेटमध्ये होता.	सोनू आणि अरबाज यांच्यात तेव्हा मोठ्या प्रमाणात पैशांचा व्यवहार झाला.	0.54	0.56

Figure 4: Sample prediction with Similarity scores: Actual (AS) and Predicted (PS)

References

Mohamed Abdalla, Krishnapriya Vishnubhotla, and Saif M Mohammad. 2021. What makes sentences semantically related: A textual relatedness dataset and empirical study.

Eneko Agirre, Carmen Banea, Claire Cardie, Daniel Cer, Mona Diab, Aitor Gonzalez-Agirre, Weiwei Guo, Inigo Lopez-Gazpio, Montse Maritxalar, Rada Mihalcea, et al. 2015. Semeval-2015 task 2: Semantic textual similarity, english, spanish and pilot on interpretability. In *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*, pages 252–263.

Eneko Agirre, Carmen Banea, Claire Cardie, Daniel Cer, Mona Diab, Aitor Gonzalez-Agirre, Weiwei Guo, Rada Mihalcea, German Rigau, and Janyce Wiebe. 2014. Semeval-2014 task 10: Multilingual semantic textual similarity. In *Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014)*, pages 81–91.

Eneko Agirre, Carmen Banea, Daniel Cer, Mona Diab, Aitor Gonzalez Agirre, Rada Mihalcea, German Rigau Claramunt, and Janyce Wiebe. 2016. Semeval-2016 task 1: Semantic textual similarity, monolingual and cross-lingual evaluation. In *SemEval-2016. 10th International Workshop on Semantic Evaluation; 2016 Jun 16-17; San Diego, CA. Stroudsburg (PA): ACL; 2016. p. 497-511.* ACL (Association for Computational Linguistics).

Eneko Agirre, Daniel Cer, Mona Diab, and Aitor Gonzalez-Agirre. 2012. Semeval-2012 task 6: A pilot on semantic textual similarity.* sem 2012: The first joint conference on lexical and computational semantics—. In *Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012), Montréal, QC, Canada*, pages 7–8.

Eneko Agirre, Daniel Cer, Mona Diab, Aitor Gonzalez-Agirre, and Weiwei Guo. 2013. * sem 2013 shared task: Semantic textual similarity. In *Second joint conference on lexical and computational semantics (* SEM), volume 1: proceedings of the Main conference and the shared task: semantic textual similarity*, pages 32–43.

José Cañete, Gabriel Chaperon, Rodrigo Fuentes, Jou-Hui Ho, Hojin Kang, and Jorge Pérez. 2023. Spanish pre-trained bert model and evaluation data. *arXiv preprint arXiv:2308.02976*.

Daniel Cer, Mona Diab, Eneko Agirre, Inigo Lopez-Gazpio, and Lucia Specia. 2017. Semeval-2017 task 1: Semantic textual similarity-multilingual and cross-lingual focused evaluation.

Yiming Chen, Yan Zhang, Bin Wang, Zuozhu Liu, and Haizhou Li. 2022. Generate, discriminate and contrast: A semi-supervised sentence representation learning framework.

Samruddhi Deode, Janhavi Gadre, Aditi Kajale, Ananya Joshi, and Raviraj Joshi. 2023a. L3cube-indicsbert: A simple approach for learning cross-lingual sentence representations using multilingual bert. *arXiv preprint arXiv:2304.11434*.

Samruddhi Deode, Janhavi Gadre, Aditi Kajale, Ananya Joshi, and Raviraj Joshi. 2023b. L3cube-indicsbert: A simple approach for learning cross-lingual sentence representations using multilingual bert. *arXiv preprint arXiv:2304.11434*.

Joseph Gatto, Omar Sharif, Parker Seegmiller, Philip Bohlman, and Sarah Masud Preum. 2023. Text encoders lack knowledge: Leveraging generative llms for domain-specific semantic textual similarity.

Edouard Grave, Piotr Bojanowski, Prakhar Gupta, Armand Joulin, and Tomas Mikolov. 2018. Learning word vectors for 157 languages. *arXiv preprint arXiv:1802.06893*.

Ali Muttaleb Hasan, Noorhuzaimi Mohd Noor, Taha H Rassem, and Ahmed Muttaleb Hasan. 2020. Knowledge-based semantic relatedness measure using semantic features. *International Journal*, 9(2).

Ananya Joshi, Aditi Kajale, Janhavi Gadre, Samruddhi Deode, and Raviraj Joshi. 2023. L3cube-mahasbert and hindsbert: Sentence bert models and benchmarking bert sentence representations for hindi and marathi. In *Science and Information Conference*, pages 1184–1199. Springer.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.

Nedjma Ousidhoum, Shamsuddeen Hassan Muhammad, Mohamed Abdalla, Idris Abdulmumin, Ibrahim Said Ahmad, Sanchit Ahuja, Alham Fikri Aji, Vladimir Araujo, Abinew Ali Ayele, Pavan Baswani, Meriem Beloucif, Chris Biemann, Sofia Bourhim, Christine De Kock, Genet Shanko Dekebo, Oumaima Hourrane, Gopichand Kanumolu, Lokesh Madasu, Samuel Rutunda, Manish Shrivastava, Thamar Solorio, Nirmal Surange, Hailegnaw Getaneh Tilaye, Krishnapriya Vishnubhotla, Genta Winata, Seid Muhie Yimam, and Saif M. Mohammad. 2024a. Semrel2024: A collection of semantic textual relatedness datasets for 14 languages.

Nedjma Ousidhoum, Shamsuddeen Hassan Muhammad, Mohamed Abdalla, Idris Abdulmumin, Ibrahim Said Ahmad, Sanchit Ahuja, Alham Fikri Aji, Vladimir Araujo, Meriem Beloucif, Christine De Kock, Oumaima Hourrane, Manish Shrivastava, Thamar Solorio, Nirmal Surange, Krishnapriya Vishnubhotla, Seid Muhie Yimam, and Saif M. Mohammad. 2024b. SemEval-2024 task 1: Semantic textual relatedness for african and asian languages. In *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*. Association for Computational Linguistics.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2015. Improving neural machine translation models with monolingual data. *arXiv preprint arXiv:1511.06709*.

Martyna Śpiewak, Piotr Sobecki, and Daniel Karaś. 2017. Opi-jsa at semeval-2017 task 1: Application of ensemble learning for computing semantic textual similarity. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 139–143.

Tokunaga Takenobu. 1994. Text categorization based on weighted inverse document frequency. *Information Processing Society of Japan, SIGNL*, 94(100):33–40.

Xin Tang, Shanbo Cheng, Loc Do, Zhiyu Min, Feng Ji, Heng Yu, Ji Zhang, and Haiqin Chen. 2018. Improving multilingual semantic textual similarity with shared sentence encoder for low-resource languages. *arXiv preprint arXiv:1810.08740*.

Dhruv Verma and SN Muralikrishna. 2020. Semantic similarity between short paragraphs using deep learning. In *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, pages 1–5. IEEE.

Wenhui Wang, Furu Wei, Li Dong, Hangbo Bao, Nan Yang, and Ming Zhou. 2020. Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers. *Advances in Neural Information Processing Systems*, 33:5776–5788.

Hao Wu, He-Yan Huang, Ping Jian, Yuhang Guo, and Chao Su. 2017. Bit at semeval-2017 task 1: Using semantic information space to evaluate semantic textual similarity. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 77–84.

Xi Yang, Xing He, Hansi Zhang, Yinghan Ma, Jiang Bian, Yonghui Wu, et al. 2020. Measurement of semantic textual similarity in clinical texts: comparison of transformer-based models. *JMIR medical informatics*, 8(11):e19735.