TartuNLP at EvaLatin 2024: Emotion Polarity Detection

Aleksei Dorkin, Kairit Sirts
University of Tartu
Tartu, Estonia
{aleksei.dorkin, kairit.sirts}@ut.ee

Abstract

This paper presents the TartuNLP team submission to EvaLatin 2024 shared task of the emotion polarity detection for historical Latin texts. Our system relies on two distinct approaches to annotating training data for supervised learning: 1) creating heuristics-based labels by adopting the polarity lexicon provided by the organizers and 2) generating labels with GPT4. We employed parameter efficient fine-tuning using the adapters framework and experimented with both monolingual and cross-lingual knowledge transfer for training language and task adapters. Our submission with the LLM-generated labels achieved the overall first place in the emotion polarity detection task. Our results show that LLM-based annotations show promising results on texts in Latin.

Keywords: emotion polarity classification, adapter training, knowledge transfer, latin

1. Introduction

This short report describes the system developed the TartuNLP team for the Emotion Polarity Detection task of the EvaLatin 2024 Evaluation Campaign (Sprugnoli et al., 2024). The goal of the task was to label Latin texts from three historical authors with four emotion polarity labels as positive, negative, neutral or mixed. For this task, no training data was provided, but only a polarity lexicon and a small evaluation set with 44 annotated sentences.

Our approach entails two steps. First, we annotated data for supervised model training a) via heuristic rules using the provided polarity lexicon and b) using GPT-4 (see Section 2). Secondly, we adopted knowledge transfer with parameter-efficient training via adapters (Houlsby et al., 2019) followed by task-specific fine-tuning on the data annotated in the first step (see Section 3). The knowledge transfer was applied both cross-lingually via pretraining on an English sentiment analysis task, and monolingually by training on an unannotated Latin text corpus.

We made two submissions to the shared task: one with heuristically annotated training data and another with the GPT-4 annotated labels. Both submissions obtained competitive results, with the submission with GPT-4 labels obtaining the first place overall. The code for the system is available on GitHub.[1]

2. Data Annotation

For the Emotion Polarity Detection task, no training data was provided. However, the organizers provided two useful resources: a polarity lexicon and a small gold annotated sample. We employed two distinct approaches to annotate the training data based on these resources: a heuristics-based and an LLM-based. The annotated data from both approaches is available on HuggingFace Hub.[2] The label distribution for the annotated data is presented in Table 1.

Label	Heuristics	LLM-based
positive	6535	1334
negative	2243	1028
mixed	5884	221
neutral	735	4698
Total	15396	7281

Table 1: Statistics of the annotated training data.

2.1. Heuristics-based annotation

In this approach, we employed the provided polarity lexicon similarly to the lexicon-based classifier by Sprugnoli et al. (2023). First, data from all available Universal Dependencies (Zeman et al., 2023) sources (Version 2.13, the most recent one at the time of writing) in Latin was collected :

1) Index Thomisticus Treebank (ITTB);
2) Late Latin Charter Treebank (LLCT);
3) UDante;
4) Perseus;
5) PROIEL treebank.

Then, the sentences containing no nouns or adjectives in the lexicon were removed. The filtered sentences were assigned labels based on the following rules:

[1] https://github.com/slowwavesleep/ancient-lang-adapters/tree/lt4hala

[2] https://huggingface.co/datasets/adorkin/evalatin2024

LT4HALA 2024@LREC-COLING-2024, pages 223–228
25 May, 2024.

1) If all words in the sentence are neutral according to the polarity lexicon, the sentence was labeled as neutral;
2) If the mean polarity of the words in the sentence is in the range from -0.1 to 0.1, then the sentence was labeled as mixed;
3) If the mean polarity is larger than 0.1, then the sentence was labeled as positive;
4) If the mean polarity is less than 0.1, then the sentence was labeled as negative.

Our expectation from this approach was that training a model on lexicon-annotated data would result in a model with better generalization capabilities than simply applying the lexicon classifier. The total amount of sentences annotated this way was 15396.

2.2. LLM-based annotation

In this approach, we made use of the OpenAI's GPT-4 model via the API (`gpt-4-turbo-preview`[3]). The sentences were again sampled from the Universal Dependencies sources. The model was given the description of the problem and one example per label from the gold annotations file. The model was tasked with assigning the given sentence a label and providing an explanation as to why it assigned that particular label.

With this approach, we expected that GPT-4 could simulate the annotation process done by an expert in Latin. According to the first author's somewhat limited understanding of Latin and based on a small sample of annotations and explanations done by the model, the output seems reasonable. We set out to spend about 15 euros per data annotation, which after removing sentences with invalid labels resulted in 7281 annotated sentences.

3. Description of the system

The system in our submission is based on the BERT architecture (Devlin et al., 2019). More specifically, we employed the multilingual version of RoBERTa (Zhuang et al., 2021)—XLM-RoBERTa (Conneau et al., 2020), which was trained on the data that included Latin.

We treated Emotion Polarity Detection as a multi-class classification problem and fine-tuned the model accordingly. However, instead of full fine-tuning, we trained a stack of adapters: a language adapter and a task adapter. Training adapters involves adding a small number of trainable parameters to the model while freezing the rest of the parameters (Houlsby et al., 2019). In addition to making the training considerably faster, adapters mitigate overfitting and catastrophic forgetting, which are common problems when dealing with small amounts of training data. We implemented our system by using the transformers[4] and the adapters[5] libraries.

We expected the model to benefit from both mono-lingual and cross-lingual knowledge transfer; therefore, the training process comprised several stages. First, we fine-tuned a Latin language adapter on a publicly available Latin Corpus[6] collected from the Latin Library[7]. In the next phase of training, we trained a task-specific classification adapter on the English IMDB movie reviews dataset[8]. The dataset contains only two labels: positive and negative. We created an adapter with a classification head with four classes, two of which remained unused during this stage. Finally, we stacked the task adapter previously trained on English on top of the language adapter, and continued training the task adapter on the annotated data in Latin.

The language adapter was trained for ten epochs with a learning rate 1e-4. For further usage, we took the last checkpoint. The task adapter was trained on data in English for five epochs with a learning rate of 5e-4, and we also took the last checkpoint. Finally, for the submissions, we trained a model on both sets of annotated data for 50 epochs with a 5e-4 learning rate. We used the provided gold annotation example as the validation set for training and measured the F-score on it after each epoch. For submission, we selected the best checkpoint based on the validation F-score.

4. Results

We made two submissions to the Emotion Polarity Detection task; the first one (TartuNLP_1) fine-tuned on the dataset with the heuristic labels, and the second one (TartuNLP_2) fine-tuned on the dataset with the LLM-generated labels. Both submissions obtained competitive results, with the model trained on the LLM-annotated labels (TartuNLP_2) taking the overall first place and the model trained on the heuristics-annotated data (TartuNLP_1) taking the second place on micro average F1-score and the third place on macro average F1-score (see Table 2).

While the scores obtained by the two models are quite close, there is frequent disagreement in their predictions: out of 294 test examples, the models

[3] https://platform.openai.com/docs/models/gpt-4-and-gpt-4-turbo

[4] https://github.com/huggingface/transformers

[5] https://github.com/adapter-hub/adapters

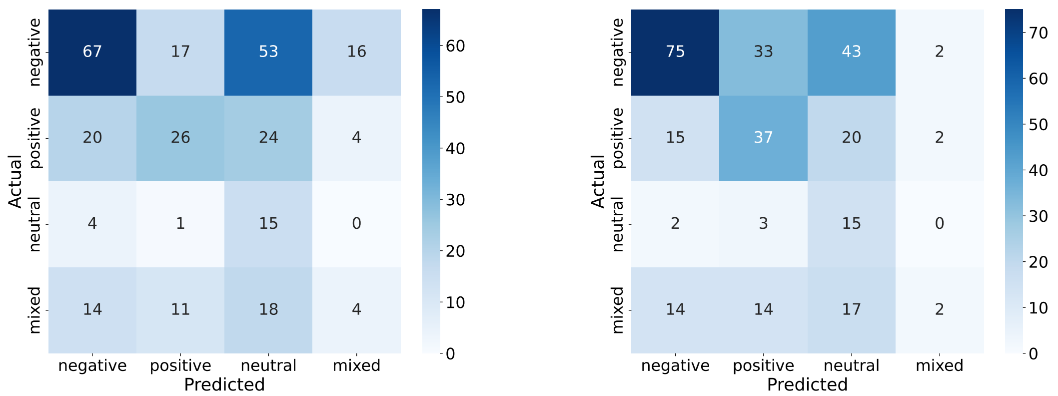
[6] https://github.com/mathisve/LatinTextDataset

[7] https://www.thelatinlibrary.com/

[8] https://huggingface.co/datasets/imdb

Model	Micro Average F1	Macro Average F1
TartuNLP_2	**0.34**	**0.29**
TartuNLP_1	0.32	0.27
NostraDomina_1	0.22	0.28
NostraDomina_2	0.22	0.22

Table 2: The overall results of all teams.



(a) TartuNLP_1 with lexicon-based heuristic labels.

(b) TartuNLP_2 with GPT4-generated labels.

Figure 1: Confusion matrices for both submissions.

disagreed in 140 examples. In case of disagreement, the heuristics- and LLM-based models made correct predictions in 40 and 57 examples respectively. Meanwhile, in case of agreement, the models correctly predicted the labels of 72 examples out of 154.

The confusion matrices for both models (see Figure 1) are similar. The models had the most trouble with the mixed class, while the negative class was the easiest to predict; this is in line with findings by Sprugnoli et al. (2023), who reported the lowest inter-annotator agreement for the mixed class, while the negative class had the highest agreement, assuming that the test data of the shared task was annotated in a similar manner.

We performed a small ablation study on the labeled test data released by the organizers after evaluating the shared task results to measure the effect of the knowledge transfer methods used:

1) Monolingual knowledge transfer from the wider Latin corpus in training the language adapter;
2) Cross-lingual knowledge transfer from the English IMDB sentiment dataset in training the task adapter.

The results of the study, shown in Table 3, were somewhat unexpected. First of all, we observe that the base model with no knowledge transfer is already as good or better than the submitted models adopting both types of knowledge transfer. Secondly, the monolingual knowledge transfer by training the language adapter improves the micro-averaged F1-score with both types of labels. Finally, the model with the LLM-generated labels benefits more from the monolingual language adapter training resulting in a model that noticeably outperforms our initial submission.

5. Discussion

The model with LLM-generated labels obtained better results than the model with lexicon-based heuristic labels, although the final results of both submitted systems are relatively close. However, the ablation study testing the effectiveness of both monolingual and cross-lingual knowledge transfer demonstrated that the model trained on the LLM-annotated data can show even better results when omitting the cross-lingual transfer from English. This is despite the fact that the number of LLM-annotated examples was nearly twice as small, suggesting that the LLM annotations are of higher quality than the labels based on lexicon-informed heuristics.

Despite our model trained on the LLM-annotated data taking the overall first place, the absolute values are somewhat low and sometimes below the baseline. There might be several reasons related to the choice of the data source and the annotation scheme and procedures. First, many of the exam-

Ablation	Micro Avg F1	Macro Avg F1	Val F1
Heuristic labels without knowledge transfer	0.33	0.26	0.48
Heuristic labels + Monolingual language transfer	0.34	0.25	0.48
Heuristic labels + Cross-lingual task transfer	0.30	0.23	0.55
Heuristic labels + Both (TartuNLP_1)	0.32	0.27	0.47
LLM labels without knowledge transfer	0.37	0.30	0.55
LLM labels + Monolingual language transfer	**0.38**	**0.30**	**0.61**
LLM labels + Cross-lingual task transfer	0.37	0.29	0.53
LLM labels + Both (TartuNLP_2)	0.34	0.29	0.48

Table 3: The results of the ablation study.

ples appear to be expository or narrative in nature. It is difficult to assign a particular emotive polarity to the texts of that kind. Furthermore, Sprugnoli et al. (2023) mention that the annotators were instructed to assign labels on the sentence level. However, they were also presented with the wider context of the sentence. This leads us to believe that some labels are actually contextual, especially when the annotated sentence contains only a single word (for example, the sentence "Mentior?" is labeled as mixed). Secondly, the manual analysis of the examples shows that it is quite difficult to distinguish between mixed and neutral texts. This appears to be true for the trained models, as well.

One possibility of improvement is to reframe the task as a multi-label classification problem instead. The model would be expected to predict the probabilities for the negative and positive labels independently. If the probability of both labels is low, the assigned label can be "neutral"; if both probabilities are high, the label can be "mixed"; otherwise, the label corresponding to the highest probability would be assigned.

6. Conclusion

This paper described our solution to the Emotion Polarity Detection task of the EvalLatin Evaluation Campaign. Our submission obtained with a model trained on a dataset with LLM-generated labels achieved the overall first place, showing that LLM-based annotations can be useful for processing texts in Latin.

7. Bibliographical References

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin De Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. 2019. Parameter-efficient transfer learning for nlp. In *International Conference on Machine Learning*, pages 2790–2799. PMLR.

Rachele Sprugnoli, Federica Iurescia, and Marco Passarotti. 2024. Overview of the EvaLatin 2024 evaluation campaign. In *Proceedings of the Third Workshop on Language Technologies for Historical and Ancient Languages â€“ LT4HALA 2024*, Torino, Italy. European Language Resources Association.

Rachele Sprugnoli, Francesco Mambrini, Marco Passarotti, and Giovanni Moretti. 2023. The sentiment of latin poetry. annotation and automatic analysis of the odes of horace. *IJCoL. Italian Journal of Computational Linguistics*, 9(9-1).

Daniel Zeman, Joakim Nivre, Mitchell Abrams, Elia Ackermann, Noëmi Aepli, Hamid Aghaei, Željko Agić, Amir Ahmadi, Lars Ahrenberg, Chika Kennedy Ajede, Salih Furkan Akkurt, Gabrielė Aleksandravičiūtė, Ika Alfina, Avner Algom, Khalid Alnajjar, Chiara Alzetta, Erik Andersen, Lene Antonsen, Tatsuya Aoyama, Katya Aplonova, Angelina Aquino, Carolina Aragon, Glyd Aranes, Maria Jesus Aranzabe, Bilge Nas Arıcan, Ĥórunn Arnardóttir, Gashaw

Arutie, Jessica Naraiswari Arwidarasti, Masayuki Asahara, Katla Ásgeirsdóttir, Deniz Baran Aslan, Cengiz Asmazoğlu, Luma Ateyah, Furkan Atmaca, Mohammed Attia, Aitziber Atutxa, Liesbeth Augustinus, Mariana Avelãs, Elena Badmaeva, Keerthana Balasubramani, Miguel Ballesteros, Esha Banerjee, Sebastian Bank, Verginica Barbu Mititelu, Starkaður Barkarson, Rodolfo Basile, Victoria Basmov, Colin Batchelor, John Bauer, Seyyit Talha Bedir, Shabnam Behzad, Kepa Bengoetxea, İbrahim Benli, Yifat Ben Moshe, Gözde Berk, Riyaz Ahmad Bhat, Erica Biagetti, Eckhard Bick, Agnė Bielinskienė, Kristín Bjarnadóttir, Rogier Blokland, Victoria Bobicev, Loïc Boizou, Emanuel Borges Völker, Carl Börstell, Cristina Bosco, Gosse Bouma, Sam Bowman, Adriane Boyd, Anouck Braggaar, António Branco, Kristina Brokaitė, Aljoscha Burchardt, Marisa Campos, Marie Candito, Bernard Caron, Gauthier Caron, Catarina Carvalheiro, Rita Carvalho, Lauren Cassidy, Maria Clara Castro, Sérgio Castro, Tatiana Cavalcanti, Gülşen Cebiroğlu Eryiğit, Flavio Massimiliano Cecchini, Giuseppe G. A. Celano, Slavomír Čéplö, Neslihan Cesur, Savas Cetin, Özlem Çetinoğlu, Fabricio Chalub, Liyanage Chamila, Shweta Chauhan, Ethan Chi, Taishi Chika, Yongseok Cho, Jinho Choi, Jayeol Chun, Juyeon Chung, Alessandra T. Cignarella, Silvie Cinková, Aurélie Collomb, Çağrı Çöltekin, Miriam Connor, Daniela Corbetta, Francisco Costa, Marine Courtin, Mihaela Cristescu, Ingerid Løyning Dale, Philemon Daniel, Elizabeth Davidson, Leonel Figueiredo de Alencar, Mathieu Dehouck, Martina de Laurentiis, Marie-Catherine de Marneffe, Valeria de Paiva, Mehmet Oguz Derin, Elvis de Souza, Arantza Diaz de Ilarraza, Carly Dickerson, Arawinda Dinakaramani, Elisa Di Nuovo, Bamba Dione, Peter Dirix, Kaja Dobrovoljc, Adrian Doyle, Timothy Dozat, Kira Droganova, Puneet Dwivedi, Christian Ebert, Hanne Eckhoff, Masaki Eguchi, Sandra Eiche, Marhaba Eli, Ali Elkahky, Binyam Ephrem, Olga Erina, Tomaž Erjavec, Farah Essaidi, Aline Etienne, Wograine Evelyn, Sidney Facundes, Richárd Farkas, Federica Favero, Jannatul Ferdaousi, Marília Fernanda, Hector Fernandez Alcalde, Amal Fethi, Jennifer Foster, Cláudia Freitas, Kazunori Fujita, Katarína Gajdošová, Daniel Galbraith, Federica Gamba, Marcos Garcia, Moa Gärdenfors, Fabrício Ferraz Gerardi, Kim Gerdes, Luke Gessler, Filip Ginter, Gustavo Godoy, Iakes Goenaga, Koldo Gojenola, Memduh Gökırmak, Yoav Goldberg, Xavier Gómez Guinovart, Berta González Saavedra, Bernadeta Griciūtė, Matias Grioni, Loïc Grobol, Normunds Grūzītis, Bruno Guillaume, Céline Guillot-Barbance, Tunga Güngör, Nizar Habash, Hinrik Hafsteinsson, Jan Hajič, Jan Hajič jr., Mika Hämäläinen, Linh Hà Mỹ, Na-Rae Han, Muhammad Yudistira Hanifmuti, Takahiro Harada, Sam Hardwick, Kim Harris, Dag Haug, Johannes Heinecke, Oliver Hellwig, Felix Hennig, Barbora Hladká, Jaroslava Hlaváčová, Florinel Hociung, Petter Hohle, Marivel Huerta Mendez, Jena Hwang, Takumi Ikeda, Anton Karl Ingason, Radu Ion, Elena Irimia, Ọlájídé Ishola, Artan Islamaj, Kaoru Ito, Siratun Jannat, Tomáš Jelínek, Apoorva Jha, Katharine Jiang, Anders Johannsen, Hildur Jónsdóttir, Fredrik Jørgensen, Markus Juutinen, Hüner Kaşıkara, Nadezhda Kabaeva, Sylvain Kahane, Hiroshi Kanayama, Jenna Kanerva, Neslihan Kara, Ritván Karahóğa, Andre Kåsen, Tolga Kayadelen, Sarveswaran Kengatharaiyer, Václava Kettnerová, Jesse Kirchner, Elena Klementieva, Elena Klyachko, Arne Köhn, Abdullatif Köksal, Kamil Kopacewicz, Timo Korkiakangas, Mehmet Köse, Alexey Koshevoy, Natalia Kotsyba, Jolanta Kovalevskaitė, Simon Krek, Parameswari Krishnamurthy, Sandra Kübler, Adrian Kuqi, Oğuzhan Kuyrukçu, Aslı Kuzgun, Sookyoung Kwak, Kris Kyle, Veronika Laippala, Lorenzo Lambertino, Tatiana Lando, Septina Dian Larasati, Alexei Lavrentiev, John Lee, Phuong Lê Hồng, Alessandro Lenci, Saran Lertpradit, Herman Leung, Maria Levina, Lauren Levine, Cheuk Ying Li, Josie Li, Keying Li, Yixuan Li, Yuan Li, KyungTae Lim, Bruna Lima Padovani, Yi-Ju Jessica Lin, Krister Lindén, Yang Janet Liu, Nikola Ljubešić, Olga Loginova, Stefano Lusito, Andry Luthfi, Mikko Luukko, Olga Lyashevskaya, Teresa Lynn, Vivien Macketanz, Menel Mahamdi, Jean Maillard, Ilya Makarchuk, Aibek Makazhanov, Michael Mandl, Christopher Manning, Ruli Manurung, Büşra Marşan, Cătălina Mărănduc, David Mareček, Katrin Marheinecke, Stella Markantonatou, Héctor Martínez Alonso, Lorena Martín Rodríguez, André Martins, Cláudia Martins, Jan Mašek, Hiroshi Matsuda, Yuji Matsumoto, Alessandro Mazzei, Ryan McDonald, Sarah McGuinness, Gustavo Mendonça, Tatiana Merzhevich, Niko Miekka, Aaron Miller, Karina Mischenkova, Anna Missilä, Cătălin Mititelu, Maria Mitrofan, Yusuke Miyao, AmirHossein Mojiri Foroushani, Judit Molnár, Amirsaeid Moloodi, Simonetta Montemagni, Amir More, Laura Moreno Romero, Giovanni Moretti, Shinsuke Mori, Tomohiko Morioka, Shigeki Moro, Bjartur Mortensen, Bohdan Moskalevskyi, Kadri Muischnek, Robert Munro, Yugo Murawaki, Kaili Müürisep, Pinkey Nainwani, Mariam Nakhlé, Juan Ignacio Navarro Horñiacek, Anna Nedoluzhko, Gunta Nešpore-Bērzkalne, Manuela Nevaci, Luong Nguyễn Thị, Huyền Nguyễn Thị Minh, Yoshihiro Nikaido, Vitaly Nikolaev, Rattima Nitisaroj, Alireza Nourian, Hanna Nurmi, Stina Ojala, Atul Kr. Ojha, Hulda Óladóttir,

Adédayọ Olúòkun, Mai Omura, Emeka Onwuegbuzia, Noam Ordan, Petya Osenova, Robert Östling, Lilja Øvrelid, Şaziye Betül Özateş, Merve Özçelik, Arzucan Özgür, Balkız Öztürk Başaran, Teresa Paccosi, Alessio Palmero Aprosio, Anastasia Panova, Hyunji Hayley Park, Niko Partanen, Elena Pascual, Marco Passarotti, Agnieszka Patejuk, Guilherme Paulino-Passos, Giulia Pedonese, Angelika Peljak-Łapińska, Siyao Peng, Siyao Logan Peng, Rita Pereira, Sílvia Pereira, Cenel-Augusto Perez, Natalia Perkova, Guy Perrier, Slav Petrov, Daria Petrova, Andrea Peverelli, Jason Phelan, Jussi Piitulainen, Yuval Pinter, Clara Pinto, Tommi A Pirinen, Emily Pitler, Magdalena Plamada, Barbara Plank, Thierry Poibeau, Larisa Ponomareva, Martin Popel, Lauma Pretkalniņa, Sophie Prévost, Prokopis Prokopidis, Adam Przepiórkowski, Robert Pugh, Tiina Puolakainen, Sampo Pyysalo, Peng Qi, Andreia Querido, Andriela Rääbis, Alexandre Rademaker, Mizanur Rahoman, Taraka Rama, Loganathan Ramasamy, Joana Ramos, Fam Rashel, Mohammad Sadegh Rasooli, Vinit Ravishankar, Livy Real, Petru Rebeja, Siva Reddy, Mathilde Regnault, Georg Rehm, Arij Riabi, Ivan Riabov, Michael Rießler, Erika Rimkutė, Larissa Rinaldi, Laura Rituma, Putri Rizqiyah, Luisa Rocha, Eiríkur Rögnvaldsson, Ivan Roksandic, Mykhailo Romanenko, Rudolf Rosa, Valentin Roșca, Davide Rovati, Ben Rozonoyer, Olga Rudina, Jack Rueter, Kristján Rúnarsson, Shoval Sadde, Pegah Safari, Aleksi Sahala, Shadi Saleh, Alessio Salomoni, Tanja Samardžić, Stephanie Samson, Manuela Sanguinetti, Ezgi Sanıyar, Dage Särg, Marta Sartor, Mitsuya Sasaki, Baiba Saulīte, Yanin Sawanakunanon, Shefali Saxena, Kevin Scannell, Salvatore Scarlata, Nathan Schneider, Sebastian Schuster, Lane Schwartz, Djamé Seddah, Wolfgang Seeker, Mojgan Seraji, Syeda Shahzadi, Mo Shen, Atsuko Shimada, Hiroyuki Shirasu, Yana Shishkina, Muh Shohibussirri, Maria Shvedova, Janine Siewert, Einar Freyr Sigurðsson, João Silva, Aline Silveira, Natalia Silveira, Sara Silveira, Maria Simi, Radu Simionescu, Katalin Simkó, Mária Šimková, Haukur Barri Símonarson, Kiril Simov, Dmitri Sitchinava, Ted Sither, Maria Skachedubova, Aaron Smith, Isabela Soares-Bastos, Per Erik Solberg, Barbara Sonnenhauser, Shafi Sourov, Rachele Sprugnoli, Vivian Stamou, Steinĥór Steingrímsson, Antonio Stella, Abishek Stephen, Milan Straka, Emmett Strickland, Jana Strnadová, Alane Suhr, Yogi Lesmana Sulestio, Umut Sulubacak, Shingo Suzuki, Daniel Swanson, Zsolt Szántó, Chihiro Taguchi, Dima Taji, Fabio Tamburini, Mary Ann C. Tan, Takaaki Tanaka, Dipta Tanaya, Mirko Tavoni, Samson Tella, Isabelle Tellier, Marinella Testori, Guillaume Thomas, Sara Tonelli, Liisi Torga, Marsida Toska, Trond Trosterud, Anna Trukhina, Reut Tsarfaty, Utku Türk, Francis Tyers, Sveinbjörn Ĥórðarson, Vilhjálmur Ĥorsteinsson, Sumire Uematsu, Roman Untilov, Zdeňka Urešová, Larraitz Uria, Hans Uszkoreit, Andrius Utka, Elena Vagnoni, Sowmya Vajjala, Socrates Vak, Rob van der Goot, Martine Vanhove, Daniel van Niekerk, Gertjan van Noord, Viktor Varga, Uliana Vedenina, Giulia Venturi, Veronika Vincze, Natalia Vlasova, Aya Wakasa, Joel C. Wallenberg, Lars Wallin, Abigail Walsh, Jonathan North Washington, Maximilan Wendt, Paul Widmer, Shira Wigderson, Sri Hartati Wijono, Seyi Williams, Mats Wirén, Christian Wittern, Tsegay Woldemariam, Taksum Wong, Alina Wróblewska, Mary Yako, Kayo Yamashita, Naoki Yamazaki, Chunxiao Yan, Koichi Yasuoka, Marat M. Yavrumyan, Arife Betül Yenice, Olcay Taner Yıldız, Zhuoran Yu, Arlisa Yuliawati, Zdeněk Žabokrtský, Shorouq Zahra, Amir Zeldes, He Zhou, Hanzhi Zhu, Yilun Zhu, Anna Zhuravleva, and Rayan Ziane. 2023. Universal dependencies 2.12. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

Liu Zhuang, Lin Wayne, Shi Ya, and Zhao Jun. 2021. A Robustly Optimized BERT Pre-training Approach with Post-training. In *Proceedings of the 20th Chinese National Conference on Computational Linguistics*, pages 1218–1227, Huhhot, China. Chinese Information Processing Society of China.