

UoB-NLP at SemEval-2025 Task 11: Leveraging Adapters for Multilingual and Cross-Lingual Emotion Detection

Frances Laureano De Leon and Yixiao Wang and Yue Feng and Mark G. Lee

University of Birmingham

Birmingham B15 2TT

Abstract

Emotion detection in natural language processing is a challenging task due to the complexity of human emotions and linguistic diversity. While significant progress has been made in high-resource languages, emotion detection in low-resource languages remains underexplored. In this work, we address multilingual and cross-lingual emotion detection by leveraging adapter-based fine-tuning with multilingual pre-trained language models. Adapters introduce a small number of trainable parameters while keeping the pre-trained model weights fixed, offering a parameter-efficient approach to adaptation. We experiment with different adapter tuning strategies, including task-only adapters, target-language-ready task adapters, and language-family-based adapters. Our results show that target-language-ready task adapters achieve the best overall performance, particularly for low-resource African languages with our team ranking 7th for Tigrinya, and 8th for Kinyarwanda in Track A. In Track C, our system ranked 3rd for Amharic, and 4th for Oromo, Tigrinya, Kinyarwanda, Hausa, and Igbo. Our approach outperforms large language models in 11 languages and matches their performance in four others, despite our models having significantly fewer parameters. Furthermore, we find that adapter-based models retain cross-linguistic transfer capabilities while requiring fewer computational resources compared to full fine-tuning for each language.

1 Introduction

Emotion detection in Natural Language Processing (NLP) remains a challenging task due to the complexity and nuance of human emotions. Language is often used in subtle and intricate ways to express emotion (Wiebe et al., 2005; Mohammad and Kiritchenko, 2018) making emotion detection difficult not only for humans but also for state-of-the-art machine learning models. Accurately identifying emotions in text has broad applications

across various fields, including commerce, public health, disaster response, and policymaking (Mohammad et al., 2018), making continued research in this area essential. Emotion detection can take different forms, such as identifying the emotions of the speaker, determining what emotion a piece of text conveys, or detecting emotions evoked in a reader (Mohammad, 2023). While significant progress has been made in high-resource languages like English and Spanish, emotion detection in low-resource languages remains underexplored.

This work focuses on multilingual and cross-lingual emotion detection, in which the goal is to predict the perceived emotions of a speaker based on a given sentence or short text snippet. This task presents several challenges. First, its multilingual and cross-lingual nature complicates emotion detection due to linguistic diversity. This diversity arises not only from the wide range of language families included in the datasets we use, but also from the varied sources used to compile the datasets (Muhammad et al., 2025a). Secondly, emotions are subjective and influenced by cultural norms, making it difficult to standardise emotion labels across languages (Van Woensel and Nevil, 2019). Thirdly, multiple emotions can coexist within a single text, requiring models to handle multi-label classification effectively.

To address these challenges, we leverage adapter-based fine-tuning and multilingual language models. Adapters are parameter-efficient and modular components that introduce a small number of trainable parameters while keeping the pre-trained model weights fixed (Houlsby et al., 2019). We investigate different adapter tuning strategies, incorporating both multilingual and cross-lingual approaches to improve model performance. Our results indicate that target-language-ready task adapters outperform standard fine-tuning methods. Additionally, using adapters with multilingual Pre-trained Language Models (PLMs) yields better per-

formance than Large Language Models (LLMs) for many of the languages (Muhammad et al., 2025a). However, we also recognise challenges in applying the adapter approach to low-resource languages, where PLMs have limited pretraining data, and note its constraints in higher-resource languages.

Our contributions include:

- Evaluating different adapter-based approaches for multilingual and cross-lingual emotion detection.
- Investigating the impact of language-family-based adapter tuning on model performance.

Our code and trained adapters will be publicly available on GitHub to support further research in this area ¹

2 Related Literature

As text-based communication continues to grow, extracting emotions from text has become a key area of research (Maruf et al., 2024). While sentiment analysis has been widely studied, it primarily categorises text into broad classes such as positive, negative, and neutral sentiment (Muhammad et al., 2025a). In contrast, emotion detection aims to identify finer-grained emotions such as anger, fear, joy, and sadness, aligning with discrete emotion models like Ekman’s six basic emotions (Ekman, 1992) and Plutchik’s Wheel of Emotions (Plutchik, 1980). Most emotion detection research has been conducted in high-resource languages such as English, Spanish, and Arabic (Maruf et al., 2024; Muhammad et al., 2018), with some studies exploring Hindi, Bangla, and code-mixed languages like Hindi-English, Punjabi-English, and Dravidian-English (Maruf et al., 2024). However, challenges remain, particularly for low-resource languages. These include limited dataset availability, the ambiguity of emotional boundaries (necessitating multi-label classification), and the difficulty of detecting emotions from text alone (Deng and Ren, 2023).

To address these challenges, we leverage adapters—lightweight modules that allow for efficient fine-tuning of language models without modifying their original parameters (Houlsby et al., 2019). Adapters are inserted within the layers of a frozen model and trained to learn task-specific representations, making them effective for low-resource and multilingual settings (Houlsby et al., 2019; Pfeiffer et al., 2020b). Previous work has

shown that multilingual transformer models such as mBERT and XLM-RoBERTa perform poorly in low-resource languages, but adapters help mitigate these limitations (Pfeiffer et al., 2020b). The standard adapter paradigm involves training Language Adapters (LAs) on unlabelled data with a Masked Language Modelling (MLM) objective, while Task Adapters (TAs) are trained on labelled task-specific data in a source language (Pfeiffer et al., 2020b; Parovic et al., 2023). For cross-lingual transfer, the task adapter is trained on top of a frozen source language adapter and later used for inference by swapping in a target language adapter. This approach enhances transferability but does not fully bridge the performance gaps in low-resource languages. Recent studies have introduced target-language-ready task adapters, which train the task adapter by cycling through multiple language adapters, improving generalisation across languages (Parovic et al., 2023). Another approach, phylogeny-inspired adapter training, incorporates linguistic relationships by structuring adapters according to language families, improving cross-lingual transfer (Faisal and Anastasopoulos, 2022). Phylogeny-based adapter tuning has also been applied to multilingual sentiment analysis in African languages (Alam et al., 2023). Building on these approaches, we explore whether combining target-language-ready task adapters with the idea of training the adapters per language family, enhances emotion detection across diverse languages. We do this because languages within the same family often share structural and lexical similarities, potentially making cross-lingual transfer more effective compared to languages from different families (Faisal and Anastasopoulos, 2022). Based on this, we train task adapters using language-family groupings to leverage these shared features, aiming to enhance knowledge transfer and improve performance within related languages. We evaluate our models on the BRIGHTER (Muhammad et al., 2025a) and EthioEmo (Belay et al., 2025) datasets, which spans a broad set of languages, including many that are underrepresented in NLP research (Muhammad et al., 2025a).

3 Experimental Setup

We conduct experiments using the SemEval-2025 Task 11 (Muhammad et al., 2025b) emotion detection datasets (Muhammad et al., 2025a; Belay et al., 2025). The datasets have predefined train-

¹<https://github.com/francesita/Adapters-EmoDetection-SemEval2025>

ing, development, and test splits. The development sets are used for hyperparameter tuning and model selection, while the test sets are used for the final evaluation. The development sets are not used for training final models or adapters. SemEval Task 11 consists of three tracks, of which we participate in two: Track A, which focuses on multi-label emotion detection, and Track C, which involves cross-lingual emotion detection. In both tracks, the objective is to determine whether each of six emotions—joy, sadness, fear, anger, surprise, and disgust—is present in the text. (For English and Afrikaans there are only five emotions.) All experiments are conducted using xlm-roberta-base and afro-xlmr-base.

The organisers provide a baseline which we use as a reference. Additionally, we establish our own baseline by fine-tuning xlm-roberta-base and afro-xlmr-base individually for each language. These baselines serve as comparison points, and their results are presented in Table 1. We perform four main experiments. The first experiment follows the original adapter setup, where a task adapter is trained stacked on an LA, where the parameters are frozen. The source language selected for this setup is Spanish. We choose Spanish, rather than English, because it is both one of the most high-resource languages in the dataset, and it contains all 6 emotion labels. Additionally, we utilise the SemEval-2018 Task 1 dataset in Spanish to augment the emotion data for this experiment, ensuring the training of a robust TA. During inference, the TA trained on Spanish emotion data is stacked on a LA corresponding to the target language. The second experiment investigates a task-only setup, where task adapters are trained for emotion detection without LAs. The third experiment involves training target-language-ready task adapters (TLR), following the approach of Parovic et al. (2023), in which a task adapter is trained using all available LAs corresponding to all the languages in the dataset. The fourth experiment focuses on language-family target-language-ready task adapters (lang-fam TLR), where TAs are trained using the previously described TLR cycling method, but only cycling through the LAs of languages within the same family. Figure 1 provides an overview of the language families considered in this work. Figure 2 illustrates the adapter setup within the transformer, which is applied in the TLR, and lang-fam TLR adapter experiments. We

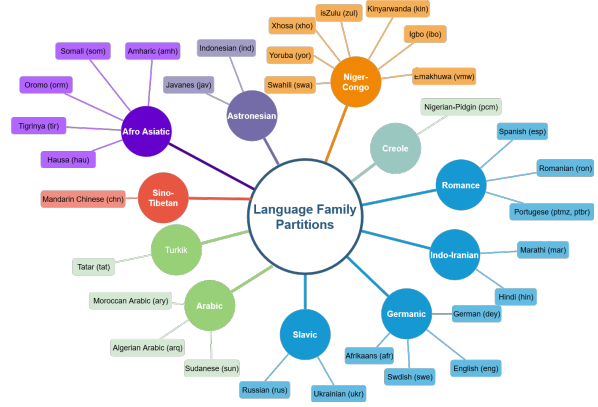


Figure 1: Languages used in this study and their partitioning into language families for adapter training.

use the HuggingFace Transformers² and AdapterHub³ (Pfeiffer et al., 2020a) libraries for model training and evaluation.

4 Methodology

Our approach consists of two key stages: LA training and TA training. We use pre-existing LAs for English, Spanish, German, Arabic, Chinese, Swahili, Hindi, and Russian, all available through the AdapterHub library (Pfeiffer et al., 2020a). For most other languages, we train new LAs, except for Romanian and Emakhuwa, due to resource limitations. In total, our experiments involve 27 LAs.

LAs are trained using the OSCAR dataset (Ortiz Su’arez et al., 2020, 2019) for languages such as Portuguese and Amharic, while Wikipedia (Foundation) is used for low-resource African languages. Training follows prior work, with each LA trained for 100 epochs or 100,000 steps. The batch size is set to 8, the learning rate to 1e-5, and the maximum sequence length to 256. For low-resource languages, training is conducted for 30,000 steps. We use a bottleneck size of 16 of the adapters. Task adapters are trained by stacking them on top of the LAs, while keeping the parameters of both the base model and the LA frozen during training. A bottleneck size of 16 is also used, consistent with prior work. For TLR TAs, training is conducted using all LAs relevant to the dataset and experimental setup.

5 Results

The best-performing model for emotion detection in our experiments was the TLR-TA, using XLM-RoBERTa-base for non-African languages

²<https://huggingface.co/>

³<https://adapterhub.ml/>

Language	Competition results	LA_t and TA	Xlm_r_target_ready_TA	Afro_xlm_r_target_ready_TA	Family_target_ready_TA	Xlm_r_TA	Afro_xlmr_TA	Xlm_r_baseline	Afro_xlmr_baseline
afr	0.4495	0.2639	0.3874	0.3629	0.4495	0.3594	0.3877	0.2962	0.2962
amh	0.67	0.3757	0.6181	0.67	0.5609	0.6272	0.543	0.522	0.5369
arq	0.4691	0.211	0.4691	0.4406	0.482	0.4475	0.4145	0.1721	0.1721
ary	0.5148	0.306	0.4328	0.4244	0.5148	0.4186	0.4017	0.2958	0.2457
chn	0.5692	0.209	0.5916	0.5692	0.5158	0.5894	0.4794	0.5126	0.4081
deu	0.548	0.4019	0.6041	0.548	0.5937	0.5947	0.4587	0.4538	0.4538
eng	0.6451	0.3953	0.6451	0.6236	0.6581	0.6526	0.6236	0.6327	0.6468
esp	0.7291	0.7324	0.7291	0.6985	0.7465	0.7283	0.6441	0.7303	0.7016
hau	0.6628	0.2297	0.5864	0.6628	0.6271	0.5916	0.5557	0.5169	0.5169
hin	0.8423	0.4049	0.8423	0.8	0.8411	0.8353	0.7404	0.8233	0.8378
ibo	0.4718	0.1826	0.4585	0.4842	0.4718	0.4854	0.4331	0.3688	0.3398
kin	0.5159	0.125	0.3178	0.5159	0.4657	0.3185	0.4561	0.3288	0.3288
mar	0.7996	0.0989	0.8384	0.7996	0.8416	0.8291	0.5886	0.7924	0.7988
orm	0.5261	0.1986	0.438	0.5261	0.4912	0.4514	0.4268	0.3931	0.2755
pcm	0.5152	0.2582	0.5274	0.513	0.5152	0.5357	0.4902	0.112	0.112
ptbr	0.4776	0.4467	0.4788	0.4398	0.4776	0.4797	0.3919	0.2185	0.3091
ptmz	0.4004	0.242	0.3944	0.3247	0.4004	0.3849	0.2616	0.0778	0.1845
ron	0.6811	0.4576	0.6811	0.6576	0.6592	0.6671	0.6349	0.1897	0.1886
rus	n/a	0.518	0.8246	0.8057	0.8282	0.8256	0.7591	0.8133	0.8137
som	0.445	0.0717	0.3355	0.445	0.3873	0.3695	0.2315	0.2744	0.2944
sun	0.3646	0.231	0.3646	0.2965	0.3359	0.3636	0.289	0.1872	0.1872
swa	0.2624	0.2377	0.2624	0.2197	0.1624	0.2187	0.1497	0.1582	0.1582
swe	0.5215	0.2342	0.5215	0.4829	0.5086	0.5165	0.4774	0.4119	0.4039
tat	0.6371	0.2929	0.6371	0.5257	0.2743	0.64	0.4078	0.3776	0.3959
tir	0.5029	0.1129	0.4149	0.5029	0.4154	0.4446	0.216	0.3357	0.3357
ukr	0.5841	0.253	0.5841	0.4983	0.5846	0.5776	0.4531	0.4307	0.4276
vmw	0.0564	0.0652	0.0564	0.0336	0.031	0.0766	0.011	0	0
yor	0.1931	0.0343	0.1446	0.1931	0.1153	0.1406	0.1599	0.0987	0.0989

Table 1: Track A results for all languages included in our experiments. Model names ending in target_TA refer to target-language ready adapters. LA_t refers to LA for some target language.

Language	Competition results	New Column	Xlm_r_target_ready_TA	Afro_xlm_r_target_ready_TA	Family_target_ready_TA	Xlm_r_TA	Afro_xlmr_TA
afr	0.3629	0.2639	0.3874	0.3629	0.4495	0.3594	0.3877
amh	0.6272	0.3757	0.6181	0.67	0.5609	0.6272	0.543
arq	0.4406	0.211	0.4691	0.4406	0.482	0.4475	0.4145
ary	0.4244	0.306	0.4328	0.4244	0.5148	0.4186	0.4017
chn	0.5692	0.209	0.5916	0.5692	0.5158	0.5894	0.4794
deu	0.5543	0.4019	0.6041	0.548	0.5937	0.5947	0.4587
eng	0.6451	0.3953	0.6451	0.6236	0.6581	0.6526	0.6236
esp	0.7291	-	0.7291	0.6985	0.7465	0.7283	0.6441
hau	0.6271	0.2297	0.5864	0.6628	0.6271	0.5916	0.5557
hin	0.8	0.4049	0.8423	0.8	0.8411	0.8353	0.7404
ibo	0.4842	0.1826	0.4585	0.4842	0.4718	0.4854	0.4331
ind	0.3329	0.3887	0.405	0.3329	0.4354	0.4027	0.3082
jav	n/a	0.2825	0.311	0.2607	0.3519	0.2766	0.2201
kin	0.4657	0.125	0.3178	0.5159	0.4657	0.3185	0.4561
mar	0.7996	0.0989	0.8384	0.7996	0.8416	0.8291	0.5886
orm	0.4912	0.1986	0.438	0.5261	0.4912	0.4514	0.4268
pcm	0.513	0.2582	0.5274	0.513	0.5152	0.5357	0.4902
ptbr	0.4398	0.4467	0.4788	0.4398	0.4776	0.4797	0.3919
ptmz	0.3247	0.242	0.3944	0.3247	0.4004	0.3849	0.2616
ron	0.6811	0.4576	0.6811	0.6576	0.6580	0.6671	0.6349
rus	0.8057	0.518	0.8246	0.8057	0.8282	0.8256	0.7591
som	0.3873	0.0717	0.3355	0.445	0.3873	0.3695	0.2315
sun	0.2965	0.231	0.3646	0.2965	0.3359	0.3636	0.289
swa	0.2197	0.2377	0.2624	0.2197	0.1624	0.2187	0.1497
swe	0.4829	0.2342	0.5215	0.4829	0.5086	0.5165	0.4774
tat	0.6371	0.2929	0.6371	0.5257	0.2743	0.64	0.4078
tir	0.4446	0.1129	0.4149	0.5029	0.4154	0.4446	0.216
ukr	0.4983	0.253	0.5841	0.4983	0.5846	0.5776	0.4531
vmw	0.0336	0.0652	0.0564	0.0336	0.031	0.0766	0.011
xho	n/a	0.1254	0.0254	0.0825	0.0568	0.0397	0.048
yor	0.1931	0.0343	0.1446	0.1931	0.1153	0.1406	0.1599
zul	n/a	0.0882	0.037	0.1159	0.082	0.0441	0.08

Table 2: Track C results for all languages included in our experiments. Model names ending in target_TA indicate target-language-ready adapters. We exclude the esp score because we trained the TA using the Spanish language LA.

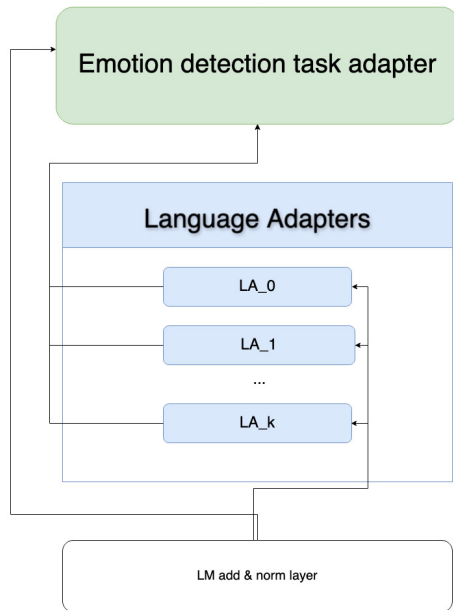


Figure 2: Illustration of a multilingual task adapter (TA) with a target language-ready (TLR) module at a single PLM layer. The diagram shows language adapters (LAs) for K target languages alongside the emotion detection TA. This method is from Parovic et al. (2023).

and Afro-XLMR-base for African languages (see Tables 1 and 2).

However, performance varied across languages. In Track A, our system ranked 7th for Kinyarwanda, 8th for Tigrinya, 11th for Hausa, 12th for Amharic, and Oromo. In Track C, it ranked 3rd for Amharic, and 4th for Oromo, Tigrinya, Kinyarwanda, Hausa, and Igbo. These results suggest that our approach was particularly effective for low-resource African languages.

While our system showed moderate overall performance, it outperformed LLMs in 11 languages under a few-shot setting. In four additional languages, it achieved performance on par with LLMs, despite being significantly smaller and more computationally efficient (Muhammad et al., 2025a). LLMs are trained on large-scale web data and are considerably larger in size, but they continue to struggle in low-resource settings. This is likely due to the limited presence of many languages in online data, which affects the models’ ability to generalise across diverse linguistic contexts. Our use of adapters enables efficient parameter sharing and supports cross-linguistic performance while requiring fewer computational resources compared to full fine-tuning for each language. These results

suggest that using adapters with smaller pre-trained language models remains a relevant and effective approach, particularly in low-resource scenarios.

The best model was selected based on development set performance during the competition. However, when evaluated on the test set, some models performed better than expected, while others underperformed. This discrepancy may be due to overfitting to the development set, differences in data distribution between the development and test sets, or varying levels of pretraining exposure to certain languages. Performance also varied depending on the type of adapter used. The task-only adapters performed best for Igbo, Pidgin, Tatar, and Emakhuwa, while the TLR task adapters achieved the highest scores for 14 languages. The Family-based task adapters performed best for two Arabic languages, two Romance languages, and Slavic languages. In our experiments, the traditional TA stacked on an LA for the target language performed poorly, except in the case of Xhosa. In all cases, using TLR adapters performed similarly to or better than fine-tuning a PLM per language while retaining multilingual transfer capabilities.

These results demonstrate that adapters provide a modular and efficient approach to multilingual NLP. A single base model can be adapted for different tasks and languages by inserting lightweight modules, reducing the need for extensive retraining. The findings align with previous research, reinforcing adapters as a flexible and scalable alternative to full-model fine-tuning.

6 Conclusion

Adapters offer a parameter-efficient method for training models in multiple languages for a given task. They are lightweight and allow language models to be applied to different tasks and languages without altering the original model weights. This approach is effective for cross-lingual transfer, outperforming the baseline in 28 languages and achieving strong results in multilingual emotion detection across 16 languages. Furthermore, using adapters with pre-trained language models continues to improve performance over LLMs in mid- and low-resource languages. Future research will examine the use of prompt tuning methods with LLMs to assess their potential for specialised tasks and low-resource languages.

7 Ethical considerations

Adapters enhance multilingual performance, but rely on pre-trained models that may introduce biases. These biases can affect specific languages, dialects, or groups, which requires further evaluation and mitigation. Although adapters are effective for medium- and low-resource languages, performance gaps persist compared to high-resource languages, requiring improvement without reinforcing digital inequalities. This study covers a limited set of languages, leaving many low-resource languages underrepresented.

Acknowledgments

The computations described in this paper were performed using the University of Birmingham's BEAR Cloud service, which provides flexible resource for intensive computational work to the University's research community. See <http://www.birmingham.ac.uk/bear> for more details.

References

- Md Mahfuz Ibn Alam, Ruoyu Xie, Fahim Faisal, and Antonios Anastasopoulos. 2023. [GMNLP at SemEval-2023 Task 12: Sentiment Analysis with Phylogeny-Based Adapters](#). In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 1172–1182, Toronto, Canada. Association for Computational Linguistics.
- Tadesse Destaw Belay, Israel Abebe Azime, Abinew Ali Ayele, Grigori Sidorov, Dietrich Klakow, Philip Slusallek, Olga Kolesnikova, and Seid Muhie Yimam. 2025. [Evaluating the Capabilities of Large Language Models for Multi-label Emotion Understanding](#). In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 3523–3540, Abu Dhabi, UAE. Association for Computational Linguistics.
- Jiawen Deng and Fuji Ren. 2023. [A Survey of Textual Emotion Recognition and Its Challenges](#). *IEEE Transactions on Affective Computing*, 14(1):49–67.
- Paul Ekman. 1992. [An argument for basic emotions](#). *Cognition and Emotion*, 6(3-4):169–200.
- Fahim Faisal and Antonios Anastasopoulos. 2022. [Phylogeny-Inspired Adaptation of Multilingual Models to New Languages](#). In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 434–452, Online only. Association for Computational Linguistics.
- Wikimedia Foundation. [Wikimedia Downloads](#).
- 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](#). *ArXiv*, abs/1902.00751.
- Abdullah Al Maruf, Fahima Khanam, Md. Mahmudul Haque, Zakaria Masud Jiyad, Muhammad Firoz Mridha, and Zeyar Aung. 2024. [Challenges and Opportunities of Text-Based Emotion Detection: A Survey](#). *IEEE Access*, 12:18416–18450.
- Saif Mohammad. 2023. [Best Practices in the Creation and Use of Emotion Lexicons](#). In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 1825–1836, Dubrovnik, Croatia. Association for Computational Linguistics.
- Saif Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018. [SemEval-2018 Task 1: Affect in Tweets](#). In *Proceedings of the 12th International Workshop on Semantic Evaluation*, pages 1–17, New Orleans, Louisiana. Association for Computational Linguistics.
- Saif Mohammad and Svetlana Kiritchenko. 2018. [Understanding Emotions: A Dataset of Tweets to Study Interactions between Affect Categories](#). In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).
- Shamsuddeen Hassan Muhammad, Nedjma Ousidhoum, Idris Abdulmumin, Jan Philip Wahle, Terry Ruas, Meriem Beloucif, Christine de Kock, Nirmal Surange, Daniela Teodorescu, Ibrahim Said Ahmad, David Ifeoluwa Adelani, Alham Fikri Aji, Felermينو D M A Ali, Ilseyar Alimova, Vladimir Araujo, Nikolay Babakov, Naomi Baes, Ana-Maria Bucur, Andiswa Bukula, Guanqun Cao, Rodrigo Tufino Cardenas, Rendi Chevi, Chiamaka Ijeoma Chukwuneke, Alexandra Ciobotaru, Daryna Dementieva, Murja Sani Gadanya, Robert Geislinger, Bela Gipp, Oumaima Hourrane, Oana Ignat, Falalu Ibrahim Lawan, Rooweither Mabuya, Rahmad Mahendra, Vukosi Marivate, Andrew Piper, Alexander Panchenko, Charles Henrique Porto Ferreira, Vitaly Protasov, Samuel Rutunda, Manish Shrivastava, Aura Cristina Udrea, Lilian Diana Awuor Wanzare, Sophie Wu, Florian Valentin Wunderlich, Hanif Muhammad Zhafran, Tianhui Zhang, Yi Zhou, and Saif M Mohammad. 2025a. [BRIGHTER: BRIdging the Gap in Human-Annotated Textual Emotion Recognition Datasets for 28 Languages](#).
- Shamsuddeen Hassan Muhammad, Nedjma Ousidhoum, Idris Abdulmumin, Seid Muhie Yimam, Jan Philip Wahle, Terry Ruas, Meriem Beloucif, Christine De Kock, Tadesse Destaw Belay, Ibrahim Said Ahmad, Nirmal Surange, Daniela Teodorescu, David Ifeoluwa Adelani, Alham Fikri Aji, Felermينو Ali, Vladimir Araujo, Abinew Ali Ayele, Oana Ignat,

- Alexander Panchenko, Yi Zhou, and Saif M Mohammad. 2025b. SemEval Task 11: Bridging the Gap in Text-Based Emotion Detection. In *Proceedings of the 19th International Workshop on Semantic Evaluation (SemEval-2025)*, Vienna, Austria. Association for Computational Linguistics.
- Pedro Javier Ortiz Su'arez, Laurent Romary, and Benoit Sagot. 2020. [A Monolingual Approach to Contextualized Word Embeddings for Mid-Resource Languages](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1703–1714, Online. Association for Computational Linguistics.
- Pedro Javier Ortiz Su'arez, Benoit Sagot, and Laurent Romary. 2019. [Asynchronous pipelines for processing huge corpora on medium to low resource infrastructures](#). Proceedings of the Workshop on Challenges in the Management of Large Corpora (CMLC-7) 2019, Cardiff, 22nd July 2019, page 9 – 16, Mannheim. Leibniz-Institut f"ur Deutsche Sprache.
- Marinela Parovic, Alan Ansell, Ivan Vulić, and Anna Korhonen. 2023. [Cross-Lingual Transfer with Target Language-Ready Task Adapters](#). In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 176–193, Toronto, Canada. Association for Computational Linguistics.
- Jonas Pfeiffer, Andreas Rücklé, Clifton Poth, Aishwarya Kamath, Ivan Vulić, Sebastian Ruder, Kyunghyun Cho, and Iryna Gurevych. 2020a. [AdapterHub: A Framework for Adapting Transformers](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 46–54, Online. Association for Computational Linguistics.
- Jonas Pfeiffer, Ivan Vulić, Iryna Gurevych, and Sebastian Ruder. 2020b. [MAD-X: An Adapter-Based Framework for Multi-Task Cross-Lingual Transfer](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7654–7673, Online. Association for Computational Linguistics.
- Robert Plutchik. 1980. [Chapter 1 - A GENERAL PSYCHOEVOLUTIONARY THEORY OF EMOTION](#). In Robert Plutchik and Henry Kellerman, editors, *Theories of Emotion*, pages 3–33. Academic Press.
- Lieve Van Woensel and Nissy Nevil. 2019. What if your emotions were tracked to spy on you? Technical report, European Parliamentary Research Service.
- Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. [Annotating Expressions of Opinions and Emotions in Language](#). *Language Resources and Evaluation*, 39:165–210.