Foul at SemEval-2023 Task 12: MARBERT Language model and lexical filtering for sentiments analysis of tweets in Algerian Arabic

Faiza BELBACHIR
IPSA - Ecole d'ingénieurs Aéronautique et Spatiale Paris
63 Bd de Brandebourg Bis, 94200 Ivry-sur-Seine
PHDUPS@gmail.com

Abstract

This paper describes the system we designed for our participation in SemEval-2023 Task 12 Track 6 about Algerian dialect sentiment analysis. We propose a transformer language model approach combined with a lexicon mixing terms and emojis which is used in a post-processing filtering stage. The Algerian sentiment lexicons were extracted manually from tweets. We report on our experiments on the Algerian dialect, where we compare the performance of MARBERT to the one of ArabicBERT and CAMeLBERT on the training and development datasets of Task 12. We also analyze the contribution of our post-processing lexical filtering for sentiment analysis. Our system obtained an F1 score equal to 70%, ranking 9th among 30 participants.

1 Introduction

AfriSenti-SemEval shared task (Task 12) (Muhammad et al., 2023b) aims at sentiment analysis on African dialectal languages of different African countries. It covers three main tasks and twelve languages with a corpus made of messages collected from the micro-blogging service Twitter (Muhammad et al., 2023a). Our participation is for the first task (Task A): monolingual opinion detection, in the specific case of the Algerian dialect. In case a tweet conveys both positive and negative sentiment expression, the strongest one must be selected for the whole message.

Algerian dialect is considered a less-resourced language for sentiment analysis as few works have addressed the task in the past despite the relatively large number of speakers of the language[1] unlike other languages like English (Belbachir and Boughanem, 2018; Pak and Paroubek, 2010). Various works on opinion detection have addressed Arabic in the past (Mohammad et al., 2018), but most of them concern classical Arabic language the exception of the work of Touileb and Barnes (2021). However everyday life communication uses the dialect. Each Arabic country is characterized by its dialect different from the others by a large number of aspects: phonology, orthography, morphology, lexicon,...(Saâdane et al., 2018). It can be written in different ways as shown as follow Table 1:

أنا أحب فواكه
انا نبغي لي فغوي
انا نحب ق3 les fruits

Table 1: Various writings in the Algerian dialect for the sentence *I like fruits.*

The importance of social media in everyday communication and the fact that Algerian Arabic is used by a majority of the population warrants our interest in developing an opinion-mining algorithm able to handle the dialect. To do so we need to solve the problems specific to the dialect, augmented by the peculiarities of the form communication takes in a microblogging service like Twitter for instance:

1. the mixing of words coming from different languages: Arabic, Berber, Turkish, French, and Spanish;
2. the liberty that tweets take with spelling, syntax, and language in general;
3. the relatively important presence of emojis for iconic communication;
4. the lack of translation resources and annotated corpora and the impossibility to use an-

[1] Algerian Arabic is spoken by the majority of Algerian population, counting around 32M speakers of the dialect. Source: https://www.visualcapitalist.com/100-most-spoken-languages/

Proceedings of the The 17th International Workshop on Semantic Evaluation (SemEval-2023), pages 389–396
July 13-14, 2023 ©2023 Association for Computational Linguistics

notated material in other languages. Even the resources in standard Arabic are of little help for handling dialectal variants.

For our approach, we chose transformers (Devlin et al., 2019) because of their ability to leverage context for semantic analysis and also because they have already been investigated for Arabic dialect sentiment analysis (Fsih et al., 2022). But the lack of Algerian dialect-specific resources needed by this type of approach for training leads us to consider an approach mixing transformers and specific lexicons. Since the pioneering work of Pak (2012), many have provided proof of the utility of emoticons for tweets sentiment analysis in Arabic Refaee and Rieser (2014) and other languages: Felbo et al. (2017), Chen et al. (2018), Choudhary et al. (2018) and Weissman (2022). This is why we complemented our lexicon of Algerian-specific subjective terms (relatively rare but with a strong subjective value) with a second lexicon based on emoticons.

This paper is organized as follows: Section 2 presents methods existing for sentiment analysis and describes the corpus data. Section 3 describes our system and explains our different experiments. Sections 4 and 5 present our experiments and results. Section 6 discusses negative results.

2 Background

2.1 State-of-the-Art Method

Several approaches have been proposed to address sentiment analysis (positive, negative, and neutral) in Arabic.

Some works are based on a subjective lexicon constructed manually, this lexicon can also be weighted, and each term has a sentiment score (Abdul-Mageed and Diab, 2012; Mataoui et al., 2016). The problem with these approaches is the difficulty of finding the subjective lexicon available. Other works focus on building their lexicon (automatically), they rely on machine translation (Mohammad et al., 2016; Guellil and Azouaou, 2016) they use English sentiment resources like the lexicon[2] of Hu and Liu (2004) or like the semantic network SentiWordNet (Baccianella et al., 2010) and then translate the subjective words into Arabic. The problem with these approaches is the difficulty to obtain a correct translation. There are also works that represent a document as a features vector with different representations such as Tf-idf, bag-of-words, or Wod2Vec, and then use different classifiers to return the polarity of the document (El Mahdaouy et al., 2017; Mikolov et al., 2013). Recent deep learning algorithms such as convolutional neural network (Kalchbrenner et al., 2014; Safaya et al., 2020), long short-term memory (Cheng et al., 2016), bidirectional LSTM (Sujana et al., 2020) improved performances for opinion mining tasks because of their ability to better take into account the sentiment word context. The transformer-based BERT approach showed particularly good performances for a wide variety of natural language understanding tasks (Devlin et al., 2019).

All these approaches require large amounts of training data and up to now have been essentially applied to the classical Arabic language, whose linguistic characteristics are more stabilized than the ones of the dialectal variants like the Algerian dialect. Furthermore, special genres like social media (Twitter) suffer from a higher variability of all language aspects than other media. Two points that we had to address when building a solution of opinion mining of tweets in the Algerian dialect, despite the existence of a few works for other dialects (Alharbi et al., 2018) and contributions for Algerian Saâdane et al. (2018), Touileb and Barnes (2021).

2.2 The Corpus

The shared Task corpus comprises tweet text. Three sets of the corpus were accessible for participants: (1) training, (2) development and (3) testing sets. Some statistics of the corpora are provided in Table 2 and Table 3. Table 2 represents the number of tweets in the three corpus and the obtained polarity (positive (pos), negative (neg), and neutral(neu)). Using three types of Emoticons corpus, Table 3 shows the obtained positive (pos), negative (neg), and neutral (neu) polarity.

Corpus	pos	neg	neu	tot
Train	892	417	342	1651
Dev	223	105	86	414
Test	-	-	-	958

Table 2: Tweet Corpus statistics on polarity.

Corpus	pos	neg	neu	tot
Train	164	295	109	568
Dev	40	72	33	
Test	-	-	-	349

Table 3: Emoticons Corpus statistics on polarity.

[2] http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar

3 System Description

To detect positive, negative, or neutral sentiments, we developed an approach that uses the contextual language model called MARBERT (Abdul-Mageed et al., 2021). This model is Arabic-focuses Language Model (LM) exploiting large and massive diverse data.
Note that it has been adapted for Arabic dialect detection task (AlKhamissi et al., 2021) on short utterances but not for language understanding. To complement the performance of the language model which we found during our experiments to be under-performing on some Algerian dialect sentiment terms, we constructed two lexicons: one for subjective terms and another for emoticons, both regrouping highly opinionated lexical items identified by hand. The lexicon was used in a post-processing stage of the language model as described in Figure 1. In this section, we detail the three modules of our pipeline: a) filtering, b) language Model, and c) lexical post-processing.

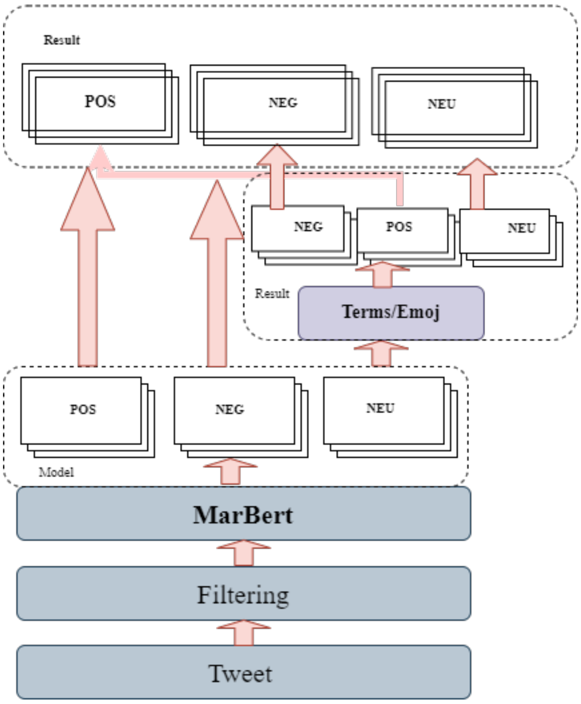

Figure 1: The overall framework of our system proposed for AfriSemEval-2023 Task 12.

3.1 Filtering

The pre-processing phase impact on the performance of sentiment analysis. The Arabic language is known by the property to have multiple forms of a given letter, for instance, ا، آ، إ، أ، ا are several forms for the letter ا *alif* and also for ؤ، و *waw* and *ya* ئ، ي. Classic Arabic imposes contextual rules for selections of these letters, while dialectal Arabic variants consider three equivalences classes, the characters being undistinguished inside each class. For our letter normalization, we select an arbitrary letter as a representative of each equivalence class. We Pre-process the data tweet and remove all the @user and RT. We choose to not remove stop-words cause we consider they provide important information for sentiment analysis. For example in: انا لن احب شوكولاتة (*I don't like chocolate*), the form لن could be considered as a stop word, but it changes the meaning of the sentence from positive to negative.

3.2 The MARBERT cased Model

We use MARBERT model transformer (Abdul-Mageed et al., 2021) to detect polarity tweets. We chose this model because it considers a dimension of the Arabic dialect language. The pre-training model's data includes 6B of Arabic tweets. The data set makes up to 1 GB of text which means 15.6 B tokens. The model use's the same network architecture as BERTbase with 160 M parameters, 17 steps, and 36 epochs.

3.3 Lexical post-processing

We build our lexicon of subjective terms, according to their respective frequency in the collection. Figure 2 and Figure 3 shows the top 50 positive and negative terms used in tweets. Our hypothesis is to determine the least frequent terms, therefore poorly represented by the model, but carrying a strong opinion. We manually filter the terms to obtain a few opinionated words. For example, words are considered as negatives وعلاه، والو، طير، فارغة، خرذة، تعلمي، سامطة، ما، طنح ، and words are considered as positives وعلاهمسكين، وعلاهالمخلوق، مزية، بتسامة. We proceed in the same way to build the emoticon lexicon. Figure 4 and Figure 5 show the top 20 positive and negative emoticons used in tweets.

Only in the case where MARBERT tags a document as "neutral", we then apply our lexicons on the document in a post-processing step and change the tag of the document if the lexical filtering yields results different from the one of MARBERT .

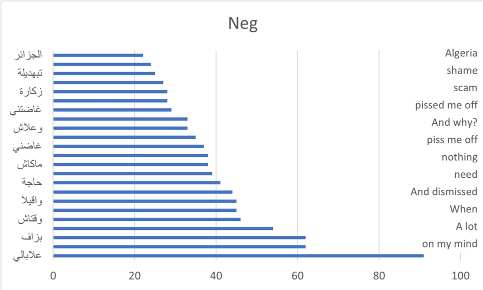

Figure 2: fr neg term

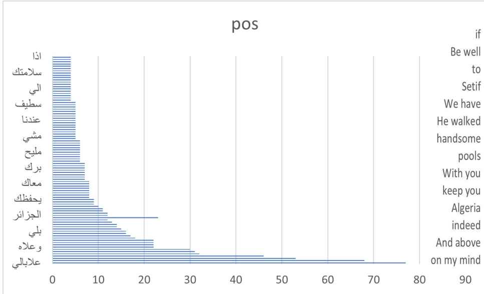

Figure 3: fr pos term

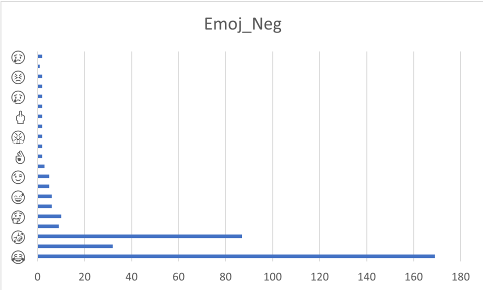

Figure 4: frequency neg emoj

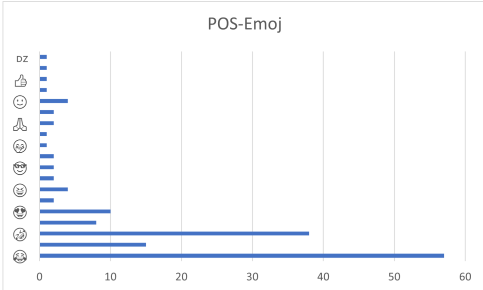

Figure 5: frequency pos emoj

4 Experiments

4.1 Training and Validation Data

Since we were limited to 100 submissions in total for the validation phase, we decided to create our internal validation set by splitting the original training and development set and conducting several experiments. Thus, our internal training set consists of 1858 annotated tweets, and the internal development set included 207 tweets. The labeled test collection was not provided. all our experiments were done by the train and dev collection; For the standard hyper-parameter setting, we based it on the work of (Alamro et al., 2021). We test only some variations shown in the following table 4.

Hyper-parameter	Range/Value
Epoch	5-10
Batch size	16
Weight Decay	1e-8
Learning Rate	1.215e-05-1.782551e-05

Table 4: : Main hyper-parameters tuned in our system.

4.2 The features influencing sentiment analysis

In our study, we test the impact of different language models on opinion detection. We complement our transformer model with a post-processing stage with lexical filtering on the part of the data that were labeled as neutral by the transformer. We ask ourselves three main questions:

- What is the best language model that represents the opinion using the Algerian dialect?
- Does our two manual subjective lexicons (Algerian terms and emoticons) have a significant impact on opinion detection for tweets?

We answer these questions in the next section.

5 Results on Official Sets

For our system, we obtain an F1 score equal to 0.70 on test data and ranking 9^{th} among 30 participants. Table 5 summarizes the results.

Precision	Recall	F1 score(weighted)
0.704	0.703	0.703

Table 5: Performance of MarBert model on test data.

We analyze in more detail the behavior of our model on the positives, negatives, and neutrals tweets classes. For this, we use the train and development collection which is annotated for all our experiments as explained in subsection 4.1.

Table 6 shows the results of the MarBet model. We can observe that our model determines the negative and positive documents better than the neutral ones. This is probably due to the distribution of the number of tweets in the collection.

Sentiment	Precision	Recall	F1-score
Negative	0.78	0.81	0.79
Positive	0.79	0.76	0.78
Neutral	0.53	0.50	0.51

Table 6: Performance of MarBert model on train and Dev data

5.1 What about Other Language Models ?

Regarding the first question, we compare during, the development phase, MARBERT (Abdul-Mageed et al., 2021) against Arabic-BERT(Safaya et al., 2020), a pre-trained language model for standard Arabic. After the test phase, we confirmed our choice of language model by comparing this time MARBERT against CAMeLBERT(Inoue et al., 2021), a language model for Arabic pre-trained on a dataset that contains a part of dialectal Arabic. For the last question, we look at the impact that the lexicon post-processing has on sentiment detection performance.

Interpreting the results tables 8, 7, we can notice that the two models yield more than 72% of the F1 score for the positive and negative classes but produce only around 45% of the F1 score for the neutral documents. Both models perform worse than the MARBERT .

Sentiment	Precision	Recall	F1-score
Negative	0.74	0.80	0.76
Positive	0.85	0.78	0.81
Neutral	0.42	0.38	0.39

Table 7: Performance of CamelBert model on train and Dev data

Sentiment	Precision	Recall	F1-score
Negative	0.71	0.81	0.75
Positive	0.84	0.63	0.72
Neutral	0.50	0.50	0.50

Table 8: Performance of ArabicBert model on train and Dev data

5.2 What about our lexicons ?

Regarding the second question, we introduce our lexicons related to Algerian terms and emoticons; and analyze their impact on sentiment analysis. The goal of this experimentation is to see the role of our lexicons in sentiment analysis.

Table 9 represents the comparison of MARBERT lexicons and MARBERT regarding precision, recall, and F1. We can see that there is an improvement of 0.4%, 0.5%, and 0.4% respectively in precision, recall, and F1 on sentiment detection with

Model	precision	recall	F1
MARBERT lexicons	0.735	0.739	0.736
MARBERT	0.731	0.734	0.732

Table 9: Performance of MarBert model with lexicons on train and Dev data

MARBERT lexicons. We can probably improve this score by further enriching our lexicons but we can say that taking the lexicon into account plays a role in sentiment analysis.

6 Negative Results

We use different methods to determine sentiment tweets which did not yield better results than the pre-training language models on Arabic. We used the SVM classifier with a vectorization based on TF-IDF. This reports 2 points drop on the F1 measure. We use the long-short-term memory model (LSTM) which yielded an accuracy of around 60%. We can conclude that the language models trained in Arabic more specifically in dialectal Arabic report better results.

7 Conclusion

In this article, we describe our approach based on a language model and sentiment lexicons that we build manually to detect positive, negative, and neutral documents. Our approach ranks among the top 10 for the Algerian dialect sentiment analysis task. We show that the MARBERT language model performs better than Arabic-BERTor CAMeLBERT. We prove that the introduction of lexical filtering with both Algerian dialectal terms and emoticons as a post-processing step to analysis with MARBERT increases the performance of our system.

We believe that this system can be improved by adding new terms or emoticons to lexicons used in complementing analysis with a BERT-based transformer and by using other combination strategies between the different models. For future work, we plan to improve our system by using other subjective lexicons of different languages.

Acknowledgments

I would like to thank Patrick Paroubek, research engineer at LISN lab of CNRS-U. Paris Saclay, for his help. Without his support and valuable contributions, this work would not have been possible.

References

Muhammad Abdul-Mageed and Mona Diab. 2012. Toward building a large-scale arabic sentiment lexicon. *GWC 2012: 6th International Global Wordnet Conference, Proceedings*, pages 18–22.

Muhammad Abdul-Mageed, AbdelRahim Elmadany, and El Moatez Billah Nagoudi. 2021. ARBERT & MARBERT: Deep bidirectional transformers for Arabic. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 7088–7105, Online. Association for Computational Linguistics.

Hind Alamro, Manal Alshehri, Basma Alharbi, Zuhair Khayyat, Manal M. Kalkatawi, Inji Ibrahim Jaber, and Xiangliang Zhang. 2021. Overview of the arabic sentiment analysis 2021 competition at kaust. This paper provides an overview of the Arabic Sentiment Analysis Challenge organized by King Abdullah University of Science and Technology (KAUST). The task in this challenge is to develop machine learning models to classify a given tweet into one of the three categories Positive, Negative, or Neutral. From our recently released ASAD dataset, we provide the competitors with 55K tweets for training, and 20K tweets for validation, based on which the performance of participating teams are ranked on a leaderboard, https://www.kaggle.com/c/arabic-sentiment-analysis-2021-kaust. The competition received in total 1247 submissions from 74 teams (99 team members). The final winners are determined by another private set of 20K tweets that have the same distribution as the training and validation set. In this paper, we present the main findings in the competition and summarize the methods and tools used by the top ranked teams. The full dataset of 100K labeled tweets is also released for public usage, at https://www.kaggle.com/c/arabic-sentiment-analysis-2021-kaust/data.

Randah Alharbi, Walid Magdy, Kareem Darwish, Ahmed AbdelAli, and Hamdy Mubarak. 2018. Part-of-speech tagging for Arabic Gulf dialect using Bi-LSTM. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Badr AlKhamissi, Mohamed Gabr, Muhammad El-Nokrashy, and Khaled Essam. 2021. Adapting MARBERT for improved Arabic dialect identification: Submission to the NADI 2021 shared task. In *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pages 260–264, Kyiv, Ukraine (Virtual). Association for Computational Linguistics.

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

Faiza Belbachir and Mohand Boughanem. 2018. Using language models to improve opinion detection. *Information Processing and Management*, 54:958–968.

Chia-Ping Chen, Tzu-Hsuan Tseng, and Tzu-Hsuan Yang. 2018. Sentiment analysis on social network: Using emoticon characteristics for Twitter polarity classification. In *International Journal of Computational Linguistics & Chinese Language Processing, Volume 23, Number 1, June 2018*, Taipei, Taiwan. Association for Computational Linguistics and Chinese Language Processing.

Jianpeng Cheng, Li Dong, and Mirella Lapata. 2016. Long short-term memory-networks for machine reading. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 551–561, Austin, Texas. Association for Computational Linguistics.

Nurendra Choudhary, Rajat Singh, Vijjini Anvesh Rao, and Manish Shrivastava. 2018. Twitter corpus of resource-scarce languages for sentiment analysis and multilingual emoji prediction. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1570–1577, Santa Fe, New Mexico, USA. 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.

Abdelkader El Mahdaouy, Eric Gaussier, and Saïd Ouatik El Alaoui. 2017. Arabic text classification based on word and document embeddings. In *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016*, pages 32–41, Cham. Springer International Publishing.

Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann. 2017. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1615–1625, Copenhagen, Denmark. Association for Computational Linguistics.

Emna Fsih, Sameh Kchaou, Rahma Boujelbane, and Lamia Hadrich-Belguith. 2022. Benchmarking transfer learning approaches for sentiment analysis of Arabic dialect. In *Proceedings of the The Seventh Arabic Natural Language Processing Workshop (WANLP)*, pages 431–435, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Imene Guellil and Faiçal Azouaou. 2016. Arabic dialect identification with an unsupervised learning (based on a lexicon). application case: ALGERIAN dialect. In *2016 IEEE Intl Conference on Computational Science and Engineering, CSE 2016, and IEEE Intl Conference on Embedded and Ubiquitous Computing, EUC 2016, and 15th Intl Symposium on Distributed Computing and Applications for Business Engineering, DCABES 2016, Paris, France, August 24-26, 2016*, pages 724–731. IEEE Computer Society.

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '04)*, pages 168–177, New York, NY, USA. Association for Computing Machinery.

Go Inoue, Bashar Alhafni, Nurpeiis Baimukan, Houda Bouamor, and Nizar Habash. 2021. The interplay of variant, size, and task type in Arabic pre-trained language models. In *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pages 92–104, Kyiv, Ukraine (Virtual). Association for Computational Linguistics.

Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. 2014. A convolutional neural network for modelling sentences. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 655–665, Baltimore, Maryland. Association for Computational Linguistics.

M'hamed Mataoui, Omar Zelmati, and Madiha Boumechache. 2016. A proposed lexicon-based sentiment analysis approach for the vernacular algerian arabic. *Res. Comput. Sci.*, 110:55–70.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

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, Mohammad Salameh, and Svetlana Kiritchenko. 2016. Sentiment lexicons for arabic social media. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France. European Language Resources Association (ELRA).

Shamsuddeen Hassan Muhammad, Idris Abdulmumin, Abinew Ali Ayele, Nedjma Ousidhoum, David Ifeoluwa Adelani, Seid Muhie Yimam, Ibrahim Sa'id Ahmad, Meriem Beloucif, Saif M. Mohammad, Sebastian Ruder, Oumaima Hourrane, Pavel Brazdil, Felermino Dário Mário António Ali, Davis David, Salomey Osei, Bello Shehu Bello, Falalu Ibrahim, Tajuddeen Gwadabe, Samuel Rutunda, Tadesse Belay, Wendimu Baye Messelle, Hailu Beshada Balcha, Sisay Adugna Chala, Hagos Tesfahun Gebremichael, Bernard Opoku, and Steven Arthur. 2023a. AfriSenti: A Twitter Sentiment Analysis Benchmark for African Languages.

Shamsuddeen Hassan Muhammad, Idris Abdulmumin, Seid Muhie Yimam, David Ifeoluwa Adelani, Ibrahim Sa'id Ahmad, Nedjma Ousidhoum, Abinew Ali Ayele, Saif M. Mohammad, Meriem Beloucif, and Sebastian Ruder. 2023b. SemEval-2023 Task 12: Sentiment Analysis for African Languages (AfriSenti-SemEval). In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*. Association for Computational Linguistics.

Alexander Pak. 2012. *Automatic, adaptive, and applicative sentiment analysis*. Theses, Université Paris Sud - Paris XI.

Alexander Pak and Patrick Paroubek. 2010. Twitter based system: Using twitter for disambiguating sentiment ambiguous adjectives. In *Proceedings of the 5th International Workshop on Semantic Evaluation, SemEval@ACL 2010, Uppsala University, Uppsala, Sweden, July 15-16, 2010*, pages 436–439. The Association for Computer Linguistics.

Eshrag Refaee and Verena Rieser. 2014. Evaluating distant supervision for subjectivity and sentiment analysis on arabic twitter feeds. In *Proceedings of the EMNLP 2014 Workshop on Arabic Natural Langauge Processing (ANLP)*, page 174–179. Association for Computational Linguistics.

Houda Saâdane, Hosni Seffih, Christian Fluhr, Khalid Choukri, and Nasredine Semmar. 2018. Automatic identification of maghreb dialects using a dictionary-based approach. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Ali Safaya, Moutasem Abdullatif, and Deniz Yuret. 2020. KUISAIL at SemEval-2020 task 12: BERT-CNN for offensive speech identification in social media. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 2054–2059, Barcelona (online). International Committee for Computational Linguistics.

Yudianto Sujana, Jiawen Li, and Hung-Yu Kao. 2020. Rumor detection on Twitter using multiloss hierarchical BiLSTM with an attenuation factor. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 18–26, Suzhou, China. Association for Computational Linguistics.

Samia Touileb and Jeremy Barnes. 2021. The interplay between language similarity and script on a novel multi-layer Algerian dialect corpus. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3700–3712, Online. Association for Computational Linguistics.

Benjamin Weissman. 2022. Emoji semantics/pragmatics: investigating commitment and lying. In *Proceedings of the Fifth International Workshop on Emoji Understanding and Applications in Social Media*, pages 21–28, Seattle, Washington, USA. Association for Computational Linguistics.