A Sequence-to-Sequence Neural Network for Joint Aspect Term Extraction and Aspect Term Sentiment Classification Tasks

Hasna Chouikhi[1][a], Fethi Jarray[1,2][b] and Mohammed Alsuhaibani[3][c]
[1]*LIMTIC Laboratory, UTM University, Tunis, Tunisia*
[2]*Higher Institute of Computer Science of Medenine, Gabes University, Medenine, Tunisia*
[3]*Department of Computer Science, College of Computer, Qassim University, Buraydah, Saudi Arabia*

Keywords: Aspect Term Extraction, Aspect Term Sentiment Classification, Encoder-Decoder Model, seq2seq Model, Arabic BERT-Based Models.

Abstract: Aspect-based Sentiment Analysis (ABSA) consists in extracting the terms or entities described in a text (attributes of a product or service) and the user perception of each aspect. Most earlier approaches are traditionally programmed sequentially, extracting the terms and then predicting their polarity. In this paper, we propose a joint sequence-to-sequence model that simultaneously extracts the terms and determines their polarities. The seq2seq architecture comprises an encoder, which can be an Arabic BERT model, and a decoder, which also can be an Arabic BERT, GPT, or BiGRU model. The encoder aims to preprocess the input sequence and encode it into a fixed-length vector called a context vector. The decoder reads that context vector from the encoder and generates the aspect term sentiment pair output sequence. We conducted experiments on two accessible Arabic datasets: Human Annotated Arabic Dataset (HAAD) of Book Reviews and The ABSA Arabic Hotels Reviews (ABSA Arabic Hotels). We achieve an accuracy score of 77% and 96% for HAAD and ABSA Arabic Hotels datasets respectively using BERT2BERT pairing. The results clearly highlight the superiority of the joint seq2seq model over pipeline approaches and the outperformance of BERT2BERT architecture over the pairing of BERT and BiGRU, and the pairing of BERT and GPT.

1 INTRODUCTION

Sentiment analysis (SA), also known as opinion mining in (Bing, 2012), can be applied at three levels: document, sentence, and aspect. The most fine-grained level is Aspect Based Sentiment Analysis (ABSA) which aims at analyzing and understanding people's opinions at the aspect level. An aspect is an attribute or a characteristic of an entity, such as quality, or price.

There are four fundamental subtasks for ABSA: (i) Aspect Term Extraction (ATE), (ii) Aspect Term Sentiment (ATS), (iii) Aspect Category Detection (ACD) and (iv) Aspect Category Sentiment Analysis (ACSA). ATE can be cast as a sequence labeling problem where one has to assign a label for each word. The format BIO is the most used for sequence labeling (beginning, inside, and outside). Aspect-term sentiment (ATS) task may be thought of as a categorization issue of the sentiment towards a specific aspect as positive, negative, or neutral in a sequence, and can be framed as a sequence classification problem (Pontiki et al., 2016). Aspect Category Detection (ACD) is framed as a classification problem in which a term is classified into one and only one of several given categories. Aspect Category Sentiment Analysis (ACSA) is cast as a sequence classification problem where the sentiment polarities toward each aspect category are predicted. Table 1 summarises these four subtasks with an example. In this contribution, we are interested in solving the first two subproblems, i.e., ATE and ATS. For this purpose, we propose a sequence-to-sequence architecture that jointly solves both subproblems.

Recently, the sequence-to-sequence (Seq2Seq for short) learning framework has been effectively applied to various NLP tasks (Cho et al., 2014; Sutskever et al., 2014; Bahdanau et al., 2014; Nallapati et al., 2016). A sequence-to-sequence model consists of two parts, the encoder, and the decoder.

[a] https://orcid.org/0000-0002-3733-2063
[b] https://orcid.org/0000-0002-5110-1173
[c] https://orcid.org/0000-0001-6567-6413

Chouikhi, H., Jarray, F. and Alsuhaibani, M.
A Sequence-to-Sequence Neural Network for Joint Aspect Term Extraction and Aspect Term Sentiment Classification Tasks.
DOI: 10.5220/0011620500003393
In *Proceedings of the 15th International Conference on Agents and Artificial Intelligence (ICAART 2023) - Volume 3*, pages 117-123
ISBN: 978-989-758-623-1; ISSN: 2184-433X

Table 1: Example of fundamental subtasks for ABSA.

Subtasks	Review Although many of the novel's passages seem to have a Hollywood influence. However, it is Arabic in nature.
ATE	Novel's passages
ATS	Negative
ACD	Feelings
ACSA	Positive

Practically, both parts are two different neural network models combined into one large network. The main task of an encoder network is to process and understand the input sequence and produce a fixed-length dimensional representation of it. This latent representation is then fed to the decoder part, which generates its own sequence that represents the output. Perhaps the most pertinent example of the Seq2seq model is part-of-speech tagging, which consists of labeling each word in a sentence by its appropriate part of speech such as noun, verb, adjective, etc.

The main contributions of this paper can be summarized as follows:

- Design a joint Seq2Seq architecture based on BERT embedding for Aspect Term Extraction (ATE) and Aspect Term Sentiment.
- prepare the datasets HAAD (Al-Smadi et al., 2015) and ABSA Arabic Hotels (Mohammad et al., 2016) so that the classification task could be performed appropriately on the joint model. Although it is a language-specific task, the extension of the datasets can be generalized to other languages.

The remainder of this paper is organized as follows. In Section 2, a review of previous research on Arabic ABSA and Seq2Seq models is presented. The proposed models are provided in Section 3. The details of the experiments and the evaluation results are presented in Section 4. Section 5 concludes the paper and provides future directions for this work.

2 RELATED WORKS

Aspect Based Sentiment Analysis (ABSA) is one type of sentiment analysis that specifically focuses on identifying and evaluating the individual aspects and the associated sentiment of a text. For each ABSA subtask, AL-Smadi et al. (Al-Smadi et al., 2015) covered many standard models. They received an F1 score of 23.4% for aspect term extraction and 15.2% for aspect category, respectively, and 29.7% and 42.6% of accuracy for ATS and ACSA (Bensoltane and Zaki, 2022).

Compared with other NLP tasks, deep learning approaches are currently in an early stage of advancement for ABSA (Oueslati et al., 2020) and especially for Arabic, which is more complicated than English ABSA(Al-Dabet et al., 2020). (Al-Smadi et al., 2019) carried out Long-Short-Term Memory (LSTM) based network in order to enhance the findings on HARD dataset ((Mohammad et al., 2016), (Pontiki et al., 2016)) in slots 2 and 3. The best result for slot 2 was accomplished with BiLSTMCRF (FastText), a 39% improvement over standard outcomes (F1 score = 69.9% versus 30.9%). Slot 3 gave results similar to the best model of SemEval 2016 Task 5 (Pontiki et al., 2016) (Accu = 82.6%). In (Al-Dabet et al., 2020), the ATS model was improved by inserting two additional layers: a CNN layer for character-level extraction and a CNN layer for character-level extraction and connection. It achieved an F1 score of 72.8% using a CBOW model prepared on a Wikipedia dataset. They showed that the character-level vectors extracted by CNN have a positive effect on the exhibition of convolution tweets.

SOTA performance in many downstream applications, including sentiment analysis, has lately been attained by large-scale pre-trained models like OpenAI GPT (Radford et al., 2019), XLNET (Yang et al., 2019), and BERTBERT (Devlin et al., 2018). These models may be adjusted for further NLP tasks because they have already been pre-trained with a massive amount of data. In (Fadel et al., 2022), the authors concatenated the embedding of BERT and Flair to better represent the words for the Arabic language. For Arabic ATE, two models were developed by concatenating AraBERT (Antoun et al., 2020), and Flair embedding and following them with an extended layer, BiLSTM-CRF or BiGRU-CRF (noted BF-BiLSTM-CRF and BF-BiGRU-CRF). The experimental results achieved an F1 score of 79.7% HARD dataset.

In this paper, we propose a Seq2Seq model for joint ATE and ATS. As far as we know, this is the first attempt that a Seq2Seq learning method built upon the BERT has been used to tackle ATE tasks using an Arabic dataset.

3 PROPOSED APPROACH Seq2Seq ABSA ARCHITECTURE

In this section, we explain the Seq2Seq architecture, for Aspect Extraction and Aspect Term Sentiment (see Figure 1).

Table 2: Characteristics of Arabic BERT versions.

Models	Variant	Size	Word	Tokens	Vocab size	Steps
ArabicBERT (Safaya et al., 2020)	MSA	95GB	8.2B	WordPiece	32K	4M
AraBERTv0.2 (Fouad et al., 2019)	MSA	77GB	8.6B	WordPiece	60K	3M
MARBERT (Abdul-Mageed et al., 2020)	MSA/DA	128GB	15.6B	WordPiece	100K	17M
ARBERT (Abdul-Mageed et al., 2020)	MSA	61GB	6.5B	WordPiece	100K	8M
CAMeLBERT-MSA (Inoue et al., 2021)	MSA	107GB	12.6B	WordPiece	30k	1M
GPT-2 (Inoue et al., 2021)	MSA	-	1.5B	-	50K	124M

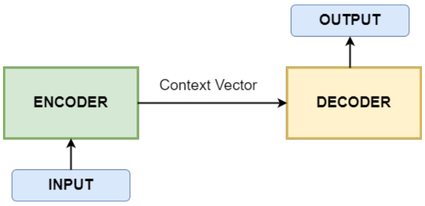

Figure 1: Seq2Seq architecture.

3.1 Datasets Annotation

The BIO format (short for Beginning, Inside, Outside) is the common tagging format for tagging tokens in ATE and named entity recognition. In this contribution, we introduce a more fine-grained labels system (Extended-BIO) by crossing the classes of both subtasks ATE and ATS to obtain a new classification. The new set of labels is {B-Positive, B-Negative, B-Conflict, B-Neutral, I-Positive, I-Negative, I-Conflict, I-Neutral, O} where for example B-Positive stands for the beginning of a term with positive sentiment. The joint model is subject to many constraints, e.g. the sequence ”O I-Positive” is not allowed. These constraints are implicitly considered throughout the learning process.

3.2 Encoder-Decoder Architecture

A Seq2Seq model is a generic framework composed of two parts: encoder and decoder, where the input sequence is first processed by the encoder and then fed to the decoder to generate the output sequence (Brownlee, 2021).

- **Encoder part:** The Encoder part consists of a BERT model that is used to learn the representation of the input sequence. BERT (Bidirectional Encoder Representations from Transformers) **BERT** (Devlin et al., 2018)BERT is a language model pre-trained on unsupervised plain text corpora, making it easier for smaller, more defined tasks such as (Chouikhi et al., 2021; Chouikhi. et al., 2021; Saidi et al., 2021; Saidi and Jarray, 2022; Saidi et al., 2022). For the Arabic language, there are many versions of BERT with different parameters and trained on different corpora such as ArabicBERT(Safaya et al., 2020), ARABERT (Fouad et al., 2019) and MARBERT (Abdul-Mageed et al., 2020). ArabicBERT (Safaya et al., 2020) used the standard configuration of BERT, including maximum sequence length of 512 tokens, 12 attention heads, 768 hidden dimensions, and 12 transformer blocks. CAMeLBERT-MSA (Inoue et al., 2021) was made as an assortment of pre-trained BERT models on Arabic texts with various sizes and variations (Modern Standard Arabic (MSA), Dialectal Arabic (DA), Classic Arabic (CA), and a blend of the three). ARABERT (Fouad et al., 2019) utilized the BERT base setup with 12 encoder blocks, 768 hidden dimensions, 12 attention heads, and 512 maximum sequence lengths. MARBERT (Abdul-Mageed et al., 2020) is an enormous scope pre-trained masked language model focused in on both Dialectal Arabic (DA) and MSA. Table 2 shows the characteristics of the Arabic BERT versions.
- **Decoder part:** The decoder part consists of either a BERT model or a Bidirectional Gated recurrent units (BiGRU) or a GPT. A Gated recurrent units (GRU) cell is a variant of the Recurrent Neural Network that learns dependencies in sequence data. A BiGRU is made up of a forward GRU $\overrightarrow{h_t}$ and a backward GRU $\overleftarrow{h_t}$ and therefore has the ability to capture long-term contextual dependency from both the past and the future. In mathematical terms: $\overrightarrow{h_t} = \overrightarrow{GRU}(x_t, \overrightarrow{h_{t-1}})$ and $\overleftarrow{h_t} = \overleftarrow{GRU}(x_t, \overrightarrow{h_{t-1}})$. A Generative Pre-Trained Transformer (GPT) is a SOTA deep learning language model that has been trained to generate human-like output in various NLP tasks(Radford et al., 2018).

Globally, we follow the framework proposed by (Rothe et al., 2020) for pairing encoder and decoder models. More precisely, we design three families of the seq2seq model: BERT2BERT (see Figure 2)

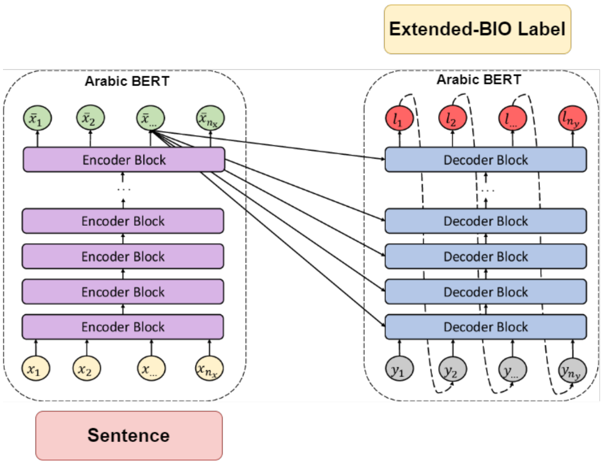

Figure 2: BERT2BERT family models (Naous et al., 2021). A BERT version as an encoder paired with another BERT version as a decoder.

Table 3: Classes distribution of HAAD and ABSA Arabic Hotels datasets.

Dataset	Sub-dataset	#Positive	#Negative	#Neutral	#Conflict	Total	Variant
HAAD	Train	1252	855	126	26	2259	MSA
	Test	124	432	22	1	579	
	Total	1376	1287	148	27	2838	
ABSA Arabic Hotels	Train	6197	3629	683	0	10509	MSA
	Test	1508	927	169	0	2604	
	Total	7705	4556	852	0	13113	

and BERT2GPT. Table 4 shows the different proposed pairings of encoders and decoders architecture.

In BERT2BERT, a BERT-initialized encoder is paired with a BERT-initialized decoder. Similarly, in BERT2GPT, a BERT-initialized encoder is paired with a GPT-initialized decoder. The final layer of the decoder in the aforementioned models is followed by a softmax that outputs the probability of each label.

In BERT2BIGRU, a BERT-initialized encoder is paired with a bidirectional Gated Recurrent Unit (BiGRU). The forward $\overrightarrow{h_t}$ and backward $\overleftarrow{h_t}$ GRU are concatenated and pass to fully connected layers (FC). The FC is followed by a softmax output layer consisting of 9 different nodes which are required for the 9 different classes (Extended-BIO).

4 EXPERIMENTS AND RESULTS

4.1 Parameter Settings

Word embeddings and hidden states (for both encoder and decoder) were configured to have a dimension of 128 and 256, respectively, in all of our experiments.

For stochastic optimization, the Adam optimizer (Kingma and Ba, 2014) is employed using the hyper-parameter values β1 = 0.9, β2 = 0.999, and ε = 10^8. The batch is 16, and the learning rate is fixed at 0.0001. We utilized the accuracy metric to assess performance.

4.2 Datasets

In this study, we conducted experiments using two existing datasets: the HAAD and ABSA Arabic Hotels datasets. Following the standard way, we divided each dataset into two subsets, with 80% of the data being used for training and 20% for testing. Most ABSA approaches can only be applied to a small number of academic datasets because there is a shortage of user reviews that are properly labeled according to aspects.

- **HAAD dataset:** (Al-Smadi et al., 2015) is considered as the most readily accessible dataset. There are 1513 Arabic book reviews. There are a total of 2838 aspect words in HAAD, and Table 3 summarizes their distribution across the four Aspect Term Sentiment classes (Positive, Negative, Conflict, and Neutral) in both the training and testing datasets.
- **ABSA Arabic Hotels Dataset:** was introduced

Table 4: Experimental results for encoder and decoder pairing in Seq2Seq architecture over HAAD and ABSA datasets. Models are classified into three families: BERT2BERT, BERT2BIGRU and BERT2GPT.

Pairing	Model	Encoder	Decoder	HAAD	ABSA Arabic Hotels
BERT2BERT	**M1**	ArabicBERT	ArabicBERT	**77%**	92%
	M2	AraBERTv0.2	AraBERTv0.2	75%	91%
	M3	MARBERT	MARBERT	72%	88%
	M4	ARBERT	ARBERT	70%	85%
	M5	CAMeLBERT-MSA	CAMeLBERT-MSA	73%	85%
	M6	ArabicBERT	AraBERTv0.2	76%	**96%**
	M7	ArabicBERT	MARBERT	75%	90%
	M8	ArabicBERT	ARBERT	72%	87%
	M9	ArabicBERT	CAMeLBERT-MSA	73%	90%
	M10	AraBERTv0.2	MARBERT	76%	89%
	M11	AraBERTv0.2	ARBERT	71%	84%
	M12	AraBERTv0.2	CAMeLBERT-MSA	70%	82%
	M13	MARBERT	ARBERT	69%	68%
	M14	MARBERT	CAMeLBERT-MSA	67%	65%
	M15	ARBERT	CAMeLBERT-MSA	65%	64%
BERT2BiGRU	**M16**	ArabicBERT	BiGRU	73%	88%
	M17	AraBERTv0.2	BiGRU	74%	89%
	M18	MARBERT	BiGRU	72%	85%
	M19	ARBERT	BiGRU	66%	70%
	M20	CAMeLBERT-MSA	BiGRU	70%	69%
BERT2GPT	**M21**	ArabicBERT	GPT	48%	56%
	M22	AraBERTv0.2	GPT	50%	55%
	M23	MARBERT	GPT	47%	52%
	M24	ARBERT	GPT	42%	50%
	M25	CAMeLBERT-MSA	GPT	44%	51%

in SemEval-2016 on the side of ABSA's multilingual task, including tasks in 8 dialects and 7 areas (Mohammad et al., 2016; Pontiki et al., 2016; Al-Smadi et al., 2019). There are 19,226 preparation tuples and, 4802 testing tuples in the dataset. The dataset includes many reviews, and each review contains multiple sentences. Each sentence includes three parts: the aspect category, the extracted opinion term, and the aspect polarity.

4.3 Results

Table 4 presents the results obtained using the Arabic Seq2Seq method. The bold values are the best results in their respective columns. With HAAD dataset, The pairing of ArabicBERT as the encoder and the decoder gives us good performance with HAAD dataset. For ABSA Arabic Hotels dataset, the optimal result is given by ArabicBERT as an encoder and AraBERT as a decoder.

Moreover, MARBERT seems to be more adequate than ARBERT as a decoder and ArabicBERT as an encoder (90% in front of 87% with ABSA Arabic Hotels dataset) Concerning BERT2BiGRU, we note that for both datasets, there is a competition between ArabicBERT and AraBERT as encoder and BiGRU as decoder (the difference is about 1%). For HAAD dataset, we noticed that AraBERT gives the best performance. For ABSA Arabic Hotels dataset, AraBERT2BiGRU gives a quite good result. It should be mentioned that BERT2BERT outperforms BERT2GPT because warm-starting (Rothe et al., 2020) a BERT decoder with a BERT checkpoint is more efficient than warm-starting a GPT2 decoder with a BERT checkpoint. In fact, any pair of BERT versions share more parameters than a pair of BERT and GPT do. Finally, we note that BERT2BERT outperforms BERT2BiGRU because the BiGRU layer is randomly initialized as there is no weight sharing between BERT and BiGRU.

5 CONCLUSION

In this paper, we propose a joint Seq2Seq model devoted to the Arabic ATE and ATS classification tasks. The experiments were carried out using the HAAD and ABSA Arabic Hotels datasets. We also annotated the datasets by manually adding polarity to each aspect. The experimental results show that pairing a BERT encoder with a BERT decoder achieves the best performance.

As a future continuation of this work, it would be interesting to jointly solve the aspect term extraction and the other subtasks of the Aspect Based Sentiment Analysis.

REFERENCES

Abdul-Mageed, M., Elmadany, A., and Nagoudi, E. M. B. (2020). Arbert & marbert: deep bidirectional transformers for arabic. *arXiv preprint arXiv:2101.01785*.

Al-Dabet, S., Tedmori, S., and Al-Smadi, M. (2020). Extracting opinion targets using attention-based neural model. *SN Computer Science*, 1(5):1–10.

Al-Smadi, M., Qawasmeh, O., Talafha, B., and Quwaider, M. (2015). Human annotated arabic dataset of book reviews for aspect based sentiment analysis. In *2015 3rd International Conference on Future Internet of Things and Cloud*, pages 726–730. IEEE.

Al-Smadi, M., Talafha, B., Al-Ayyoub, M., and Jararweh, Y. (2019). Using long short-term memory deep neural networks for aspect-based sentiment analysis of arabic reviews. *International Journal of Machine Learning and Cybernetics*, 10(8):2163–2175.

Antoun, W., Baly, F., and Hajj, H. (2020). Arabert: Transformer-based model for arabic language understanding. *arXiv preprint arXiv:2003.00104*.

Bahdanau, D., Cho, K., and Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Bensoltane, R. and Zaki, T. (2022). Towards arabic aspect-based sentiment analysis: a transfer learning-based approach. *Social Network Analysis and Mining*, 12(1):1–16.

Bing, L. (2012). Sentiment analysis and opinion mining (synthesis lectures on human language technologies). *University of Illinois: Chicago, IL, USA*.

Brownlee, J. (2021). Encoder-decoder recurrent neural network models for neural machine translation. *Machine Learning Mastery*.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Chouikhi, H., Chniter, H., and Jarray, F. (2021). Arabic sentiment analysis using bert model. In *International Conference on Computational Collective Intelligence*, pages 621–632. Springer.

Chouikhi., H., Chniter., H., and Jarray., F. (2021). Stacking bert based models for arabic sentiment analysis. In *Proceedings of the 13th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management - KEOD,*, pages 144–150. INSTICC, SciTePress.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Fadel, A. S., Saleh, M. E., and Abulnaja, O. A. (2022). Arabic aspect extraction based on stacked contextualized embedding with deep learning. *IEEE Access*, 10:30526–30535.

Fouad, M. M., Mahany, A., and Katib, I. (2019). Masdar: a novel sequence-to-sequence deep learning model for arabic stemming. In *Proceedings of SAI Intelligent Systems Conference*, pages 363–373. Springer.

Inoue, G., Alhafni, B., Baimukan, N., Bouamor, H., and Habash, N. (2021). The interplay of variant, size, and task type in arabic pre-trained language models. *arXiv preprint arXiv:2103.06678*.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Mohammad, A.-S., Qwasmeh, O., Talafha, B., Al-Ayyoub, M., Jararweh, Y., and Benkhelifa, E. (2016). An enhanced framework for aspect-based sentiment analysis of hotels' reviews: Arabic reviews case study. In *2016 11th International conference for internet technology and secured transactions (ICITST)*, pages 98–103. IEEE.

Nallapati, R., Zhou, B., Gulcehre, C., Xiang, B., et al. (2016). Abstractive text summarization using sequence-to-sequence rnns and beyond. *arXiv preprint arXiv:1602.06023*.

Naous, T., Antoun, W., Mahmoud, R. A., and Hajj, H. (2021). Empathetic bert2bert conversational model: Learning arabic language generation with little data. *arXiv preprint arXiv:2103.04353*.

Oueslati, O., Cambria, E., HajHmida, M. B., and Ounelli, H. (2020). A review of sentiment analysis research in arabic language. *Future Generation Computer Systems*, 112:408–430.

Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., et al. (2016). Semeval-2016 task 5: Aspect based sentiment analysis. In *International workshop on semantic evaluation*, pages 19–30.

Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., et al. (2018). Improving language understanding by generative pre-training.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.

Rothe, S., Narayan, S., and Severyn, A. (2020). Leveraging pre-trained checkpoints for sequence generation tasks. *Transactions of the Association for Computational Linguistics*, 8:264–280.

Safaya, A., Abdullatif, M., and Yuret, D. (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.

Saidi, R. and Jarray, F. (2022). Combining bert representation and pos tagger for arabic word sense disambiguation. In *International Conference on Intelligent Systems Design and Applications*, pages 676–685. Springer.

Saidi, R., Jarray, F., and Alsuhaibani, M. (2022). Comparative analysis of recurrent neural network architectures for arabic word sense disambiguation. In *Proceedings of the 18th International Conference on Web Information Systems and Technologies, WEBIST 2022, Valletta, Malta, October 25-27, 2022*, pages 272–277. SCITEPRESS.

Saidi, R., Jarray, F., and Mansour, M. (2021). A bert based approach for arabic pos tagging. In *International Work-Conference on Artificial Neural Networks*, pages 311–321. Springer.

Sutskever, I., Vinyals, O., and Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, 27.

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., and Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, 32.