Metaphorical Expressions in Automatic Arabic Sentiment Analysis

Israa Alsiyat[1,2] **and Scott Piao**[1]
[1]School of Computing and Communications, Lancaster University, UK
[2]Northern Borders University, Saudi Arabia
{i.alsiyat,s.piao@lancaster.ac.uk; israa.alsiyat@nbu.edu.sa}

Abstract

Over the recent years, Arabic language resources and NLP tools have been under rapid development. One of the important tasks for Arabic natural language processing is the sentiment analysis. While a significant improvement has been achieved in this research area, the existing computational models and tools still suffer from the lack of capability of dealing with Arabic metaphorical expressions. Metaphors have an important role in Arabic language due to its unique history and culture. Metaphors provide a linguistic mechanism for expressing ideas and notions that can be different from their surface form. Therefore, in order to efficiently identify true sentiment of Arabic language data, a computational model needs to be able to "read between lines". In this paper, we examine the issue of metaphors in automatic Arabic sentiment analysis by carrying out an experiment, in which we observe the performance of a state-of-art Arabic sentiment tool on metaphors and analyse the result to gain a deeper insight into the issue. Our experiment evidently shows that metaphors have a significant impact on the performance of current Arabic sentiment tools, and hence it is an important task to develop Arabic language resources and computational models for Arabic metaphors.

Keywords: Sentiment Analysis, Metaphor, Arabic NLP, Evaluation

1. Introduction

With the increase of the social media and e-commerce activities in the Arabic-speaking world, Arabic online tools and services have been increasing rapidly. As a result, the Arabic-speaking users have access to an increasing number of Internet channels to express their thoughts and opinions towards commercial products or other topics. As the result, the amount of Arabic textual data for sentiment analysis (SA) available online has been growing substantially, and there is an urgent need for developing corpus resources, efficient algorithms and software tools for detecting and analysing sentiment information from the Arabic language data.

Over the past years, a variety of Arabic language resources, algorithms and tools have been suggested and developed, such as (Elarnaoty et al., 2012) and (Rushdi-Saleh et al., 2011), with various levels of success. However, automatically identifying sentiment information from the Arabic language data is highly challenging. One of the major reasons for the difficulty derives from rich Arabic metaphorical expressions. Metaphor is intuitively fundamental in our ordinary life of thinking, speaking and action (Lakoff and Johnson, 1980). The figurative expressions, which covers the most equivocal expressions, or metaphors, is a non-literal sub-language of an imaginative and embodies speech. A purpose of using such expressions is to hide true feelings in certain situations. Metaphor is an interesting cognitive linguistic as well as technical issue of the Natural Language Processing (NLP) for understanding and computing nuanced linguistic semantic mechanism.

While metaphor is a part of all human languages, Arabic language has a higher utilization of metaphorical expressions than most other languages. This is due to the unique history and culture of the Arabic world. In addition, the metaphorical expressions in general is considered to be an Arabic attractiveness (abdul-Mageed and Diab, 2012). As a result, this figure of speech is used frequently in Arabic language.

In NLP, metaphor is a challenging factor in efficiently identifying true sentiment polarity, because it introduces an extra level of ambiguity to the polarity of Arabic text. Therefore, automatic detection and analysis of sentiment of Arabic text involving metaphor expressions is highly challenging, and even more so when considering Arabic as a contextual and derivational language (Alkhatib and Shallan, 2016).

In this paper, we examine the nature of challenges presented by Arabic metaphors in automatically analysing sentiment of Arabic text by carrying out an experiment.

2. Related Work

In this section, we review some of the latest research on Arabic metaphor pertaining to the Arabic sentiment analysis. Most earlier works on Arabic metaphors tend to be conducted in the context of linguistic study, such as sentiment analysis in translation in the Arabic context, following the similar works done for the English language, such as (Gholami et al., 2016) and (Al-Harrasi, 2001). A typical such work was conducted by Faycel (2012) who, employing corpus-based approach, studied Arabic Tunisian metaphors about food in the Proverb domain, as Arabic proverbs are full of metaphors as they reflect Arabic history as well as culture. Faycel classified the proverbs from human physiological and psychological perspectives. As far as we are aware, there have been no published work on computationally detecting and analysing Arabic metaphors from Arabic discourse.

For the task of Arabic sentiment analysis, most of the previous works tend to build corpora for various entities with different levels of annotation, such as the works carried out by Refaee and Rieser (2014), Rushdi-Saleh et al. (2011) and Abdul-Mageed and Diab (2012). Rushdi-Saleh et al. (2011) constructed an Opinion Corpus for Arabic Language (OCA) of movie reviews collected from blogs and websites. The OCA is publicly available and it has been used for testing various machine learning algorithms. Abdul-Mageed and Diab (2012) built a multi-genre corpus of Modern Standard Arabic (MSA), which was annotated using crowdsourcing approach at the sentence level. In addition, they tagged linguistic difficulty

levels in their annotation based on evidence. Refaee and Rieser (2014) collected a corpus of over 8,860 tweets with human annotation. Some other works conducted for the Arabic sentiment depended on English text translated to Arabic, such as (Elarnaoty et al., 2012).

Metaphor is also a challenging issue for researchers working on the annotation of Arabic language resources for the sentiment analysis. Even human annotators found it hard to precisely label sentiments of sentences containing metaphors due to the ambiguity they introduce (Abdul-Mageed and Diab, 2012). However, ignoring the metaphors would affect the accuracy of sentiment analysis and credibility of the annotation.

Therefore, to develop an efficient sentiment analysis tool, the metaphors and their sentiment orientation need to be detected before the general sentiment analysis algorithm can be applied. Our work examines this issue by testing an existing state-of-art tool on an Arabic test data containing metaphors in an effort of finding a solution.

3. Metaphors in Arabic Language

There are different types of metaphors in Arabic language. Metaphor (isti'araa in Arabic استعاره) is an advanced simile where the *likeness* (target) or the *liking* (trait) is deleted in the context (Beust et al., 2003). Simile in the Arabic language must have both the likeness and linking, while a metaphor needs to have just one of them. As a part of the rhetoric concept in Arabic language, the metaphors have a major impact on the sentiment contrariety, because they drive the polarity of text. For instance, the expression " قلبه أبيض" (his heart is white) has a notion of kindness, which is positive (Raii, 2008).

The following example shows the difference between the *simile* and *metaphor*. For example, the sentence "the soldier is fighting as a lion" is obviously a simile, because both the *likeness* (soldier) and the *linking* (lion) exist. In order to change the sentence to metaphor we need to omit either the *likeness* or the *linking*, e.g. omit the *likeness* "the soldier roar". If we remove *lion* instead and add one of the lion's traits, it becomes an implicit metaphor. On the other hand, in the sentence "the lion shoots the enemy", we replace *solider* with *lion* and add one of the human behaviours, which is *shooting*, then it becomes an explicit metaphor.

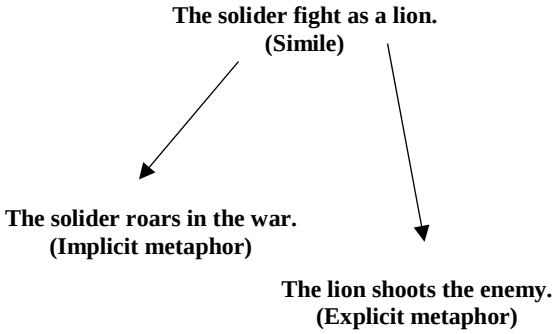

Figure 1: An example showing the difference between simile and metaphor.

In the example above, a metaphorical expression is used as informal language in social media (dialectal Arabic), while the formal Arabic language (Classical and MSA) is frequently used in rhetoric speech to express strong emotion. Note that the type of metaphor here is the representational metaphor.

The polarity orientation of a sentence could be metaphorically positive or negative, depending on the interpretation of the metaphors in the given contexts. We need to find a method or design an algorithm for the disambiguation. There have been some attempts to translate the metaphorical expressions from Arabic to English language. Yet, we have not found any work that computationally analyse metaphors with respect to the sentiment for Arabic language.

Metaphor is ubiquitous as an essential part of all languages, thus automatic detection and analysis of metaphor is a primary task for many NLP projects (Willks et al., 2013). However, so far the Arabic language has not received sufficient attention compared to other main languages concerning the metaphor detection and the sentiment analysis, partly due to the scarcity of reliable lexical and language resources (Alowisheq et al., 2016).

4. Experiment

In order to test existing Arabic sentiment analysis tools in terms of their capability of dealing with Arabic metaphors, we searched for such tools that are publically accessible. As the result, we found that there are very limited number of tools available for processing formal Arabic text. Most Arabic sentiment tools focus on the twitter data and assign sentiment polarity to the words in Tweets in order to create lexicons (word-level). For example, El-Masri et al. (2017) published a web-based tool to classify the sentiment of Arabic tweets using R language.

From the available tools, Mazajak[1] was selected for our experiment, which is an Arabic web-based sentiment analysis tool. The tool employs a deep learning model, which has achieved state-of-the-art performance on several Arabic idiom knowledge sets in the SemEval 2017 and ASTD (Farha and Magdy, 2019). Therefore, we assumed that our experiment result with this tool would be largely representative of the current performance of similar tools in general.

4.1 Test Data

We have collected a dataset of a collection of forty Arabic sentences, aiming to provide a testbed for the sentiment tool evaluation. Some of the sentences were collected from the HAAD dataset (Al-Smadi et al., 2015), and the others were collected from the LABR (Large-Scale Arabic Book Reviews) dataset (Aly and Atiya, 2013). LABR is the largest Arabic sentiment dataset, containing 63,000 reviews of Arabic. As a subset of the LABR, HAAD contains 2,389 book reviews in which the data is annotated with sentiment at aspect level,

There is a substantial amount of the sentiment dataset from which to choose the test data. However, we selected our test data carefully by selecting those sentences, which are

[1] http://mazajak.inf.ed.ac.uk:8000/

opinionated instead of purely informative such as news dataset (Maamouri et al., 2004). In detail, our test data contains 20 sentences that contain metaphors, coupled with 20 others without metaphors. Table 1 below lists the twenty sentences containing metaphors.

Sentence	SO
كتاب يجيبلك صرع (Book brings you epilepsy)	neu
دي مش بلد، دي خرابة (This is not a country, it is a corral)	neg
هيلحس دماغك (Going to lick your brain (make you confused))	neu
قاتلة هذه الرواية ! (This novel is a killer (means really good))	neg
كبسولات مركزة من النصائح. (An intense capsule of advices)	neu
الكتاب ده مش ممكن مالووووووووش حل (This book is impossible, has no solving)	neg
نزهة شعرية لذيذة مع الدكتور غازي القصيبي ، هذا الكتاب كفيل بتعديل المزاج (This is a delicious poetry trip with Dr. Gazi Alqusaibi, this book will set a good mood)	neu
من أخطر السير الذاتية على الإطلاق. (One of the dangerous autobiography (means really good)	neu
رواية رمزية مبهمة من الألف إلى الياء. لم يرق لي الترميز المعقّد والذي وصل حد الثمالة (An obscure symbolic novel from A to Z, I didn't like using complicated symbolic till drunk)	neu
رواية تتفجر منها ينابيع من عواطف الحزن والحنين والألم ، دائما يدهشك الرائع نجيب محفوظ بوصفه (A novel that explodes springs of emotions of sadness, nostalgia and pain, always surprised by the wonderful Naguib Mahfouz)	pos
ع الجرح(On the injury (on point))	neg
كتاب ملحد. (An atheist book (means the book contains text of not believing in god)); the book cannot be atheist (the writer or the person is)).	neu
القصة مريضة و تركت الفكرة الأولى التي تتبادر في الذهن عن اختلاف الثقافة بين بلدين و الغربة و ما بعدها إلى محور سيء ردئ (The story is sick, and explains the cultural differences between two countries and emigration as a negative thing)	neg
الرواية سوداء وربما تسأل نفسك لم لم يكن هناك ولا لحظة سعادة في هذه الرواية إطلاقاًوقد أجاب الراوي في النهاية بأنها من نسج خيال أحد التعساء (A black novel, maybe you ask yourself, why there is no happy moment of this novel at all?, the writer answered in the end that the novel is that novel from hapless imagination)	neg
أعتقد أني كسرت صيامي على بصل بعد 6 سنوات قضيتها بعيداً عن الروايات (I think I broke my fast on onion after 6 years of being away from novels)	neg
صاحب اليد الكبيرة السلطان حامد (The one who has the big hand, Sultan Hamid) means generosity	pos
رواية زبالة (a trash novel)	neg
الرواية دي من أفضل أنواع المخدرات (This novel is one of the best drugs), meaning a strong effect or influence	neu
خيال قاتل (Killer imagination), means being good	neg
دمااااغ (Brain) means being in a good mood)	pos

Table 1: Twenty Arabic sentences containing metaphors.

In the table, *SO* indicates "Sentiment Orientation" and *pos*, *neg* and *neu* indicate *positive*, *negative* and *neutral sentiments* respectively. In addition, the metaphors are highlighted.

Such a design of the test data is for examining the impact of metaphors on the tool's performance in comparison in order to avoid the possibility that the general underperformance of the tool on certain types of text to be interpreted as being caused by the metaphors. We assumed that such a choice of test data can ensure that we gain an objective view about how the metaphors affect the sentiment analysis algorithm and models.

4.2 Experiment of Sentiment Detection

As we mentioned earlier, our aim of this experiment is to assess how effectively the current tools can handle the Arabic metaphors in the context of automatic sentiment analysis and examine the impact of the metaphors. More specifically, we wanted to identify how the current Arabic sentiment tools perform on the sentences that do and do not contain metaphors respectively. Given the test data sentences were selected from the same dataset, we assume that, if the test tool shows a significant difference of performance on these two groups of sentences, it would demonstrate the impact of the metaphors. If such impact of metaphors can be identified, then we can design and develop better methodologies, resources and algorithms to effectively deal with the Arabic metaphors, enhancing and improving the existing tools.

For this purpose, we carried out the experiment following three steps below:

(1) Pass the Arabic sentences containing metaphors to the Mazajak tool via its web interface to automatically classify their sentiment orientation.
(2) Pass the sentences that do not contain metaphor to Mazajak for the same purpose.
(3) Manually analyse the sentiment classification results, particularly to compare the performance of the Mazajak tool on the sentences with and without metaphor contained.

As the result, in our experiment, the Mazajak tool performed significantly different on the two sets of sentences, as illustrated in Figure 2 and 3 below.

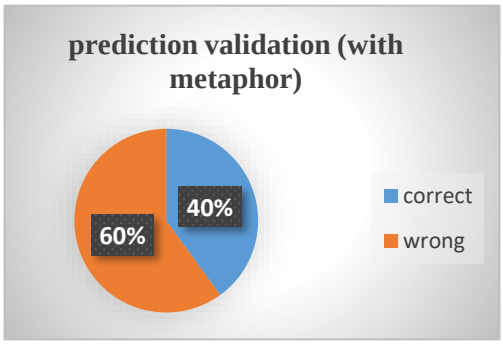

Figure 2: Experiment result without metaphor

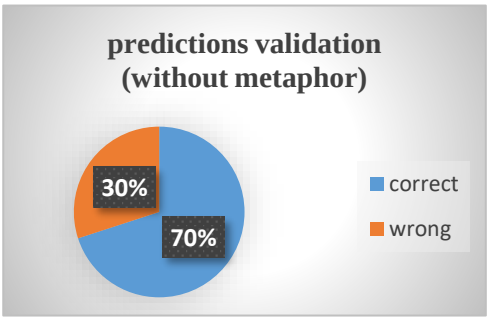

Figure 3: Experiment result with metaphors.

As shown in figure 2, the Mazajak tool achieved an impressive result on the sentences that do not contain any metaphors, correctly classifying the sentiment of 70% sentences with only 30% error rate. On the other hand, as shown in figure 3, when dealing with the sentences that contain metaphors, its performance degraded significantly by 10 percent, producing 60% and 40% accuracy and error rates respectively.

In fact, such a result is expected, because as we understand the Mazajak tool is not trained on any metaphor sentiment data. Our purpose of this experiment is not to simply evaluate existing tools on metaphors, but to find out what needs to be done to improve existing methods to handle Arabic metaphors when automatically analysing sentiment of text.

5. Discussion

Our experiment result clearly reveals the impact of the Arabic metaphors on the automatic sentiment classification. The performance of the Mazajak tool significantly drops when metaphors appear in the test sentences, compared with that on non-metaphorical sentences. The lower precision of 60% compared to that of 70% on non-metaphorical sentences is a clear indicator that metaphors have a significant influence on computing algorithms and models of Arabic sentiment analysis. Furthermore, because the test data were selected from the same larger dataset, sharing similar textual features, the difference of tool performance is not simply due to different features of test texts.

Although we acknowledge that the tools based on machine learning need to adapt the models for new styles of text, the significantly lower relative performance on sentences with metaphorical expressions provides a clear evidence that, in order to precisely classify sentiment orientation of Arabic language, the computational models must be adjusted to become capable of determining the hidden sentiment sense of the metaphorical expressions. Our experiment result also shows that the current Arabic sentiment tools suffer from lack of such capability.

In order to understand the types of errors, we manually checked the sentences which caused the errors. A such sentence contains the metaphorical expression "ع الجرح", which means "on the injury". The Mazajak tool mis-classified it as a negative sentiment, based on the literal meaning of the word "injury". It makes sense if this word is used with its normal contexts. However, in this particular sentence, this word has a hidden meaning, being part of a metaphorical expression. The sentence under discussion, "قاتلة هذه الرواية !", actually means "This novel is a killer !", a notion implying that the *novel* is so good that it can "kill off any other novels". Therefore, the true sentiment of the expression and sentence is positive. This example showcases that, without the help of special computational models for dealing with Arabic metaphors, general sentiment analysis models, including machine learning, would tend to fail.

Another example of sentiment analysis error is in the sentence "من أخطر السير الذاتية على الإطلاق" which means "one of the more dangerous autobiography". In Arabic language, the word (أخطر) (translation of *more dangerous*) is normally a descriptive comparative noun which does not carry any sentiment orientation. But in this sample sentence, it actually means "outstanding" or "really good" as a metaphorical expression. Therefore, the sentiment of the whole sentence should be positive. However, the Mazajak tool failed to capture the metaphorical sense and classified this sentence as neutral.

A similar example that demonstrates the challenge of metaphor is "النصائح من مركزة كبسولات", which literally means "capsules of intensive and comprehensive advices". Here "capsules" (كبسولات) has the notion of *cure* and has positive sentiment. But the Mazajak tool mis-classified this sentence as neutral, because the word (كبسولات) is generally used in heath context with neutral sentiment. Similarly, the metaphorical expression in the test data "دماغك هيلحس " (licking your brain) expresses a notion of confusion and frustration. But again, the sentiment tool mis-classified it as neutral, as expected, because the word "licking" itself does not have any sentiment.

In other sentences where the metaphors contain words with same surface sentiment as the whole sentences, the sentiment tool could produce correct classifications. For example, with the sentence " صاحب اليد الكبيرة السلطان حامد " (The one who has the big hand, Sultan Hamid) which means generosity, because the word "big" has obvious positive sentiment, the sentiment tool could predict the sentiment of the whole sentence correctly. This further indicates that tools like Mazajak rely on the superficial meanings of words.

The ideal method to deal with such problems would be to first detect whether or not a sentence contains metaphor, then apply different sentiment analysis algorithms and models suitable for each case. Figure 4 outlines a possible solution to the issue of metaphor in automatic sentiment analysis. In essence, we need another layer in sentiment analysis system, which is responsible for detecting metaphors and analyse their sentiment orientation. If no metaphors are included, the input text can be processed with existing sentiment classification models and tools.

A key issue here is how to automatically detect Arabic metaphors. Over the past years, various approaches have been suggested for this purpose. For example, Wilks et al. (2013) suggested a method based on lexicons. In this method, they analyse and compare the frequencies of a verb co-occurring with nouns in a candidate metaphorical sentence. If the verb is not strongly associated statistically with the nouns appearing in the same sentence, then it is considered to be likely a part of a metaphor. For instance,

the verb "shave" tends to be associated with a set of nouns, such as *beard*, *face*, *leg*, *head* etc. However, if the verb "shave" co-occurs with the noun "life", which is rare, then the sentence likely contains a metaphorical expression.

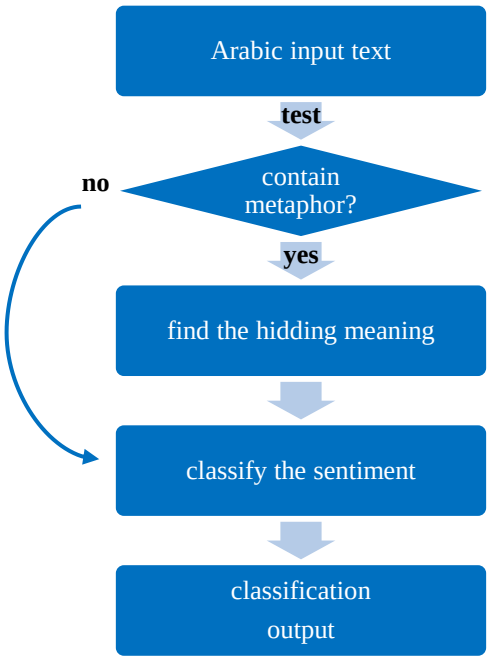

Figure 4: Suggested model outline for handling metaphors in sentiment analysis.

Applying the Wilks et al's hypothesis to our test data, a sentence that contains an expression "هيلحس دماغك" (make you licking your brain) which means something that is confusing and hard to understand. According to the general usage of these words, the verb "lick" has a weak co-occurrence association with the noun "brain". Therefore, we can assume that this sentence is likely to contain a metaphor. Such methods can potentially provide useful ideas for automatic detection for Arabic metaphors, too.

As we have shown, it is an important task to develop Arabic language resources and computing models for dealing with metaphors for Arabic sentiment analysis systems.

6. Conclusion and future work

In this paper, we have discussed the issue of analysing and handling Arabic metaphorical expressions in the context of automatic sentiment analysis. We conducted an experiment in which we tested a state-of-art Arabic sentiment tool with a test data of Arabic sentences, half of which contain metaphorical expressions. We observed a significant performance degrading of the tool when dealing with sentences containing metaphors. Our experiment result evidently shows that, in order to develop an efficient Arabic sentiment analysis system, we need to develop Arabic corpus resources and models that are capable of precisely identifying the hidden sentiment orientation of Arabic metaphors. In particular, we urgently need to develop methodology and tools for detecting Arabic metaphors and identifying their true sentiment polarities in given contexts.

7. Acknowledgements

We thank The Northern Borders University of Saudi Arabia and The Royal Embassy of Saudi Arabia Cultural Bureau for funding the first author's PhD project in Lancaster University, UK, which involves this work.

8. References

Abdul-Mageed, M., and Diab, M. T. (2012). AWATIF: A Multi-Genre Corpus for Modern Standard Arabic Subjectivity and Sentiment Analysis. In *LREC* (Vol. 515, pp. 3907-3914).

Al-Harrasi, A. N. K. (2001). *Metaphor in (Arabic-into-English) translation with specific reference to metaphorical concepts and expressions in political discourse* (Doctoral dissertation, University of Aston in Birmingham).

Alkhatib, M., and Shaalan, K. (2016). Natural language processing for Arabic metaphors: a conceptual approach. In *International Conference on Advanced Intelligent Systems and Informatics*, Springer, Cham. pp. 170-181.

Alowisheq, A., Alhumoud, S., Altwairesh, N., and Albuhairi, T. (2016). Arabic sentiment analysis resources: a survey. In Proceedings of *The Social Computing and Social Media: The 8th International Conference, SCSM 2016, Held as Part of HCI International 2016, Toronto, ON, Canada, July 17–22, 2016*. Springer International Publishing. pp. 267-278.

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

Aly, M., and Atiya, A. (2013). Labr: A large scale arabic book reviews dataset. In *Proceedings of The 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (Vol. 2), pp. 494-498.

Beust, P., Ferrari, S., and Perlerin, V. (2003). NLP model and tools for detecting and interpreting metaphors in domain-specific corpora. In *Proceedings of the Corpus Linguistics 2003 Conference*, pp. 114-123.

Elarnaoty, M., AbdelRahman, S., and Fahmy, A. (2012). A machine learning approach for opinion holder extraction in Arabic language. *arXiv preprint arXiv:1206.1011*.

El-Masri, M., Altrabsheh, N., Mansour, H., and Ramsay, A. (2017). A web-based tool for Arabic sentiment analysis. *Procedia Computer Science*, *117*, pp. 38-45.

Farha, I. A., and Magdy, W. (2019). Mazajak: An Online Arabic Sentiment Analyser. In *Proceedings of the Fourth Arabic Natural Language Processing Workshop*, pp. 192-198.

Faycel, D. (2012). Food metaphors in Tunisian Arabic proverbs. *Rice Working Papers in Linguistics*, (Vol. *3).*

Gholami, S., Montashery, I., and Khorrami, M. (2016). Metaphors in Arabic and English texts: A case study of translation of metaphors in the English versions of Al-Sahifa Al-Sajjadiyyah. *Journal of Literature, Languages and Linguistics* (Vol. *21)*, pp. 56-66.

Jakobson, R., and Halle, M. (1971). *Fundamentals of language*. The Hague: Mouton.

Lakoff, G., and Johnson, M. (1980). Conceptual metaphor in everyday language. *The journal of Philosophy*, *77*(8), pp. 453-486.

Maamouri, M., Bies, A., Buckwalter, T., and Mekki, W. (2004). The penn arabic treebank: Building a large-scale annotated arabic corpus. In *NEMLAR conference on Arabic language resources and tools* (Vol. 27), pp. 466-467.

Raii, J. (2009). Metaphor in day-to-day Arabic speech: a conceptual approach. *Tishreen University Journal for Research and Scientific Studies-Arts and Humanities Series*, *31*(1), pp. 175-193.

Refaee, E., and Rieser, V. (2014). An arabic twitter corpus for subjectivity and sentiment analysis. In Proceedings of The *LREC* 2014. pp. 2268-2273.

Rushdi-Saleh, M., Martín-Valdivia, M. T., Ureña-López, L. A., and Perea-Ortega, J. M. (2011). OCA: Opinion corpus for Arabic. *Journal of the American Society for Information Science and Technology*, *62*(10), pp. 2045-2054.

Wilks, Y., Dalton, A., Allen, J., and Galescu, L. (2013). Automatic metaphor detection using large-scale lexical resources and conventional metaphor extraction. In *Proceedings of the First Workshop on Metaphor in NLP*, pp. 36-44.