The Impact of Online Metaphors on Automatic Arabic Sentiment Analysis

Israa (Mousa) Alsiyat, BSc, MSc
School of Computing and Communications
Lancaster University

A thesis submitted for the degree of
Doctor of Philosophy

August, 2025

Ethics Declaration

This research has been conducted by the ethical guidelines and regulations set by Lancaster University. Ethical approval for this study was obtained and reviewed from the Lancaster University Faculty of Science and Technology Research Ethics Committee prior to the research conducting, with reference number FST-2025-4936-RECR-4. The research adheres to integrity, confidentiality, and respect for all participants involved. All participants were fully informed about the purpose and scope of the study before participating. Informed consent was obtained from each participant, ensuring their voluntary participation and the right to withdraw at any stage without consequences. Measures were taken to guarantee the anonymity and confidentiality of participants, all data being securely stored and used only for research purposes. Throughout the research process, all efforts were made to minimize potential risks to the participants. No personally identifiable information has been disclosed and strict compliance with data protection regulations has been maintained, following the GDPR guidelines. Furthermore, the research adheres to the highest academic integrity and transparency. No instances of data fabrication, falsification, or plagiarism have occurred. All sources and contributions have been appropriately acknowledged. By including this ethics declaration, I affirm my commitment to conducting research, ensuring that all processes align with established ethical standards.

Translation Approach

This thesis follows the Encyclopedia of Arabic Language and Linguistics resource for translation as necessary and available. However, the resource does not contain the new Arabic vernaculars that have been invented recently in online writing. For example, the word الألش *'al-alš'* has no resource for translation and meaning in the above resource. So, the word is interpreted using the annotators' annotations. All translations in this thesis follow the literal translation to show the metaphor in the context. While the accurate translation, which shows the aimed meaning of using the metaphor, was translated based on the annotators' annotation. For the transliteration, the same resource was followed to transliterate the acknowledged Arabic terms, the new terms that have semantic voice such as فسسس *'fsss'* as a notion of *'disappointment'*. The transcribed Arabic metaphor terms provided in the appendix.

The Impact of Online Metaphors on Automatic Arabic Sentiment Analysis

Israa (Mousa) Alsiyat, BSc, MSc.
School of Computing and Communications, Lancaster University
A thesis submitted for the degree of *Doctor of Philosophy*. August, 2025

Abstract

In this study, I investigate the unique use of Arabic metaphors in online communication. In such contexts, metaphors are often conveyed through semantic and symbolic phrases, with opinions frequently reduced to a single keyword. This concise expression suits the fast-paced nature of online interactions, where metaphorical language is commonly used to share viewpoints. As metaphor usage increases online, deeper interpretation becomes essential to uncover the sentiments behind these expressions. To explore this, I compiled the Arabic Online Metaphor Corpus (AMC), a foundational step in evaluating the impact of metaphors on sentiment. My research focuses on how Arabic online metaphors influence sentiment analysis, particularly through their semantic and symbolic nature. However, a major challenge lies in the absence of effective tools for annotating general Arabic texts. Moreover, the unique metaphorical structures found online must first be analyzed before annotation is possible. To assess AMC's impact, we employed a state-of-the-art Arabic semantic analyzer. The limited availability of Arabic sentiment analyzers posed a significant obstacle. To address this, we used the Mazajak sentiment analyzer and additional tools tested on datasets tagged using the Arabic semantic tagger. This approach aimed to explore how metaphorical language could be analyzed and applied to sentiment classification. Our experiments demonstrated the potential of semantic annotation in accurately identifying metaphorical sentiment. We evaluated the performance of different strategies using F-score, precision, and recall metrics. As a result, this research has led to the creation of the first Arabic online metaphor corpus, an initial design for metaphor-based sentiment classification, and the evaluation of sentiment prediction tools using AMC. These contributions represent a significant advancement toward the automatic recognition and interpretation of Arabic metaphors and their associated sentiments in online discourse.

Publications Derived from My PhD Research

Israa Alsiyat and Scott Piao (May 2020a). "Metaphorical Expressions in Automatic Arabic Sentiment Analysis". English. In: *Proceedings of LREC2020 Conference*. The 12th Edition of the Language Resources and Evaluation Conference, LREC2020 ; Conference date: 11-05-2020 Through 16-05-2020. European Language Resources Association (ELRA), pp. 4911–4916. URL: https://lrec2020.lrec-conf.org/en/

Israa Alsiyat, Scott Piao, and Mansour Almansour (July 2023). "Arabic Metaphor Corpus (AMC) with Semantic and Sentiment Annotation". English. In: The twelfth International Corpus Linguistics Conference, CL2023 ; Conference date: 03-07-2023 Through 06-07-2023. URL: https://wp.lancs.ac.uk/cl2023/

Israa Alsiyat (Apr. 2025). "Arabic Metaphor Sentiment Classification Using Semantic Information". In: *International Journal on Cybernetics and Informatics (IJCI)*. URL: https://ijcionline.com/volume/v14n2

Acknowledgments

I dedicate this work to my mother and father, to my two children, and to all those who contributed to bring this research to light.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Arabic online metaphor and sentiment

بيض means (*Eggs / bīd*) is a word that is found to be a noun in regular writing. However, in online writing, it is an Arabic metaphorical word that indicates an opinion towards something. Specifically, the term in a metaphoric context means *'old'* or *'bad'*. The term reflects the change in Arabic metaphor usage in online writing. While the term has to be defined to identify the sentiment, the Arabic language has no reliable resource to acknowledge Arabic online metaphors. Also, the meaning of the same term changes based on the context of the metaphor. Thus, meaning is one of the major factors in identifying sentiment.

The online writing style, which is the way of expressing opinions, has changed depending on the platform and the means of sharing opinions. SA has developed as a means of analyzing opinions expressed online. The opinions take different forms online to be interpreted into sentiment. Opinions started to be expressed on a scale from zero to five, represented as stars. The stars produce an analysis known as a sentiment score. The opinion developed to be stars accompanied by a brief explanation of the product called a review. So, the sentiment changed to analyze the sentiment behind the text not only as sentiment score, but as polarity. The sentiment later developed to analyze various types of text despite the opinion form of writing. The Sentiment Analysis (SA) is developing as the methods of expressing opinions change in online communication. Recently, opinion expressed visually using Bitmoji has become one of the ways to express an opinion in online communication. Shiha and Ayvaz (2017) discussed the impact of the bit emoji on SA. Although Shiha and Ayvaz (2017) did not discuss that a bit emoji could only be expressed as an opinion to identify sentiment, the researchers did discuss how a Bitmoji in association with text can be used to identify sentiment. They also designed a model to identify sentiment based on data collected from Twitter and SentiWordNet.

While SA approaches are still affected by the type of text to be analyzed. These challenges could related to the writing means, the language structure, and the aspect of the language to be analyzed. Those challenges have their own characteristics. For example, writing comprises two ways: formal and informal. Formal writing includes the regular structure of the language so that sentiment will be specified according to that regular structure. So, the Sentiment on the language levels should be specified prior to the sentiment identification, which are phonetics, morphology, syntax, lexicology, semantics and figurative. Formal text is found in news reports and articles. Informal text, however, needs more analysis of the structure in order to identify sentiment, including the standard level of the language. For example, online text can use one metaphor word to express an opinion. This means that informal text has no standard criteria to follow to identify the sentiment. In addition, the sentiment classification for formal writing could be more complex if the text has particular parts of speech but with specific language rules because the text contains ambiguities that need to be clarified. In contrast, the sentiment method should adapt to the unfixed structure such in informal text. Another form of the informal writing style, for the Arabic language there is a new way of online communication called Arabizi and there are new dialectal Arabic metaphorical terms used in online communication such as جامده, which means *'solid'* in literal sense and *'very good'* in metaphorical sense, which is the subject of this current study. Identifying sentiment in informal text is therefore more challenging.

As a consequence of the change in the form of their opinion, the sentiment method was developed to cover different aspects of the online text. For example, machine learning is one of the techniques used, and it based on trained data. Further techniques are used to tackle the lack of labeled data and find the sentiment based on learning from big data such as CNN (Convolutional Neural Network). Mixed methods and resources can be used to identify sentiment. For example, Farha and Magdy (2019) used CNN and LSTM in their model to predict sentiment through a web base sentiment analyzer.

The previous discussion highlights that online writing is mostly unstructured, unpredictable, and constantly changing, which makes designing a tool that can adapt to such unpredictable text challenging. Additionally, it lacks a foundational knowledge base for analysis, making it difficult to design a schema for annotation. However, a schema could be designed based on the practical aspects of metaphor identification. Consequently, the analysis would be based on the frequency of the appearance of online Arabic metaphors in online contexts.

1.2 The current Arabic sentiment analysis

Little attention has been paid to Arabic SA compared with English in terms of creating a reliable lexicon and an annotated corpus Alowisheq et al. (2016) which is publicly available for research and development. Also, researchers have developed thecurrent SA tools on the market for English Rushdi-Saleh et al. (2011b). Examples of resources for Arabic SA include Abdul-Mageed and Diab (2012) and Alhazmi et al. (2013) which are not publicly available. Moreover, some researchers have used Arabic datasets translated from English to detect Arabic SA, such as Elarnaoty et al. (2012). Furthermore, it is that it creates an algorithm that can perform accurately on Arabic SA, given that Arabic has distinct dialects and is highly nuanced.

Regarding the progress of SA research into the Arabic language, one of the more precise and efficient approaches is Aspect-Based Sentiment Analysis (ABSA). ABSA is an effective method of measuring the level of sentiment within a text. It assesses the sentiment in specific entities with particular aspects of a given entity, differentiating ABSA from other methods. In general, the other SA methods such as document-level and sentence-level approaches, determine the overall polarity of documents and sentences, without specifying any aspects or features applied to texts. ABSA for the Arabic language was introduced in 2015 by Al-Smadi et al. (2015), who annotated ABSA employing a human annotation approach using the BRAT (Rapid Annotation Tool) Stenetorp et al. (2012).

In the document-level method, the whole document is considered a single entity and analyzed as a whole. Sometimes, however, the outcome produced by this approach is inadequate. A document that is positively opinionated about an entity does not necessarily imply that the author has positive views about all the features of that particular entity. Similarly, a document that is negatively opinionated about an entity does not necessarily signify that the author is wholly negative about all the features of that entity. Also, in an opinionated text, the author can express positive and negative opinions about the same entity and its attributes. Elarnaoty et al. (2012) are an example of researchers who have studied SA at the document level.

At the sentence level, the document is broken down into sentences. Each sentence is then treated as a single entity and single sentences are analyzed one at a time. The SA result generated by this approach is better than the SA result generated by the document-level method because it is more refined and detailed. The majority of present techniques seek to establish the overall polarity of a document, paragraph and/or sentence regardless of the entity being expressed Farra et al. (2010) is an example of such a method. Finding the sentiment in a text in Arabic associated with metaphor is highly challenging because Arabic is an exceptionally figurative language, which means, as Reyes and Rosso (2012) stated: "it often takes advantage of linguistic devices such as metaphors, analogy, metonymy, and hyperbole to communicate complicated meanings".

Whereas SA is one component that researchers analyse in terms of automatic text understanding, another challenge is to understand nuanced linguistics mechanisms such as metaphors, which is the essence of the research in the current study. Metaphors are casual, short, and direct devices used for online communication. The unstructured metaphors that frequently used online and which do not usually follow grammatical rules are cases that NLP is used to address. In particular, Arabic metaphors used online are a significant subject to grasp and tackle, especially with the challenges of distinct unstructured Arabic colloquial terms, which also show a diversity of metaphorical expression. Metaphor is frequently used in Arabic to express a persuasive and robust opinion. In addition, Arabic has composite metaphors used in the online context, Arabic metaphors as a figurative and nuanced form of language. For instance, typical metaphors in Arabic have different categories such as cognitive metaphor, declarative metaphor, dead, cliché, tock, adapted, recent and original as mentioned in Zeroual and Lakhouaja (2018).

Historically, Arabic has a high correlation with the figurative category. Abdul-Mageed and Diab (2012) pointed out that this literary form is considered as an attractive feature of Arabic because Arabic frequently utilises metaphors as a figurative linguistic device (Reyes and Rosso, 2012). In addition, Arabic has multiple dialects, such as Modern Arabic, Classical Arabic, and Gulf Arabic (Zeroual and Lakhouaja, 2018), Which makes the development of algorithms more challenging. Arabic has a higher metaphor utilization than languages because history and culture play an essential role. Neglecting metaphor detecting sentiment has a negative impact on SA accuracy and annotation reliability. As a result, businesses can tend to lose interest, which is caused by trusting analyses of their products.

In an experiment in which I examined the impact of metaphor on sentiment, I collected 20 metaphorical sentences and 20 non-metaphorical sentences from the LARB dataset (Aly and Atiya, 2013) and the HAAD dataset (Al-Smadi et al., 2015). These two datasets were selected carefully after monitoring other publicly available datasets. My purpose was to examine a state-of-the-art sentiment analyzer's treatment of Arabic SA involving metaphors. Although I intended to know that a sentiment analyzer is not tuned to metaphor sentiment analysis. I intended to spotlight the problem and discuss the possible optimal solutions.

The results of the experiment set out below showed a significant change in the polarity prediction of the sentiment analyzer on sentences with and without metaphors involved. The findings show that even a state-of-the-art sentiment analyzer cannot deal with metaphorical expressions. Moreover, metaphor was found to play a critical role in shifting the polarity because metaphors have hidden meanings that are not apparent. For instance, the sentence خيال قاتل, which means *'killer imagination'*, is a notion of an excellent imaginative story. The sentiment analyzer predicted the incorrect polarity for this example in the experiment. The following figures illustrate the sentiment analyzer's performance on the metaphorical and non-metaphorical expressions.

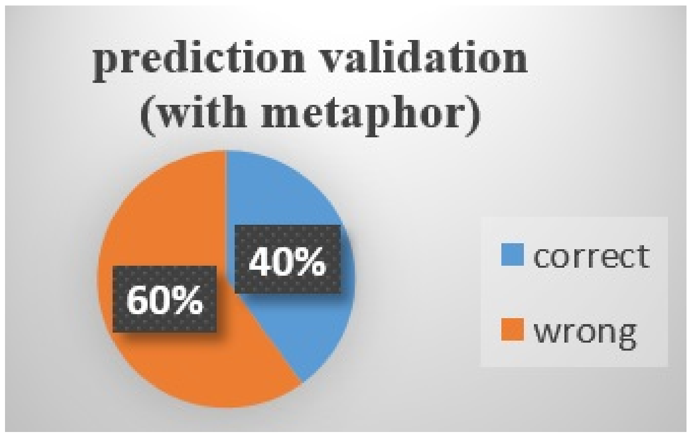


Figure 1.1: Experiment result with metaphor (Alsiyat and Piao, 2020a)

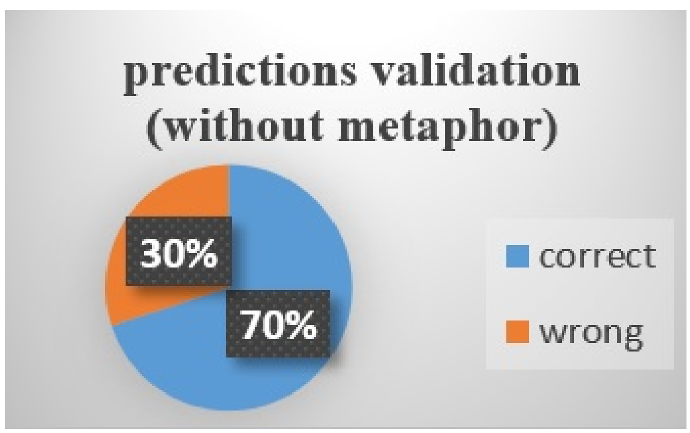


Figure 1.2: Experiment result without metaphor (Alsiyat and Piao, 2020a)

Based on the evidence briefly given above, I decided that the solution was to build an Arabic metaphor corpus annotated with sentiment, meaning, and context. In addition, a tool associated with the state-of-the-art Arabic sentiment analyzer should be designed and tested to investigate the impact of Arabic online metaphors regarding sentiment using different methods. Precisely, the Arabic metaphor corpus uses the designed tool, and state-of-the-art

Arabic sentiment analyzers are used to show the impact of Arabic metaphor in identifying sentiment. The results from the methods should then be compared with the gold standard to show the accuracy of the automatic method. The model's performance was assessed using the standard measurements: precision, recall, and F-score.

This thesis contributes to identifying Arabic sentiment with metaphor in the online context. However, there is no Arabic corpus available that is annotated for metaphor. Additionally, there is no data on Arabic metaphors in the online context annotated with sentiment. Therefore, gathering Arabic metaphors in the online context and building an Arabic metaphor corpus are the main aims of this research. Since there is no efficient Arabic annotation tool for online Arabic metaphors, the annotation is done manually using an XML editor called Oxygen[1].

Building the corpus is the essential first step to identifying sentiment, showing the impact of Arabic metaphors on sentiment, and testing the accuracy of the methods. We annotate the corpus with semantics using the semantic tool El-Haj et al. (2022) after designing tools to identify the sentiment. The tools5.4 and 5.6 were designed to identify metaphors based on existing Arabic resources, specifically the Arabic semantic tagger (El-Haj et al., 2022). Additionally, the tool will be compared with another new Arabic sentiment analyzer.

1.3 Research motivation

In the previous section, I have described the background work for this study and pointed out the limitations of previous studies and the challenges and significance of my research direction. I have demonstrated the impact of metaphor on sentiment identification using an automatic Arabic analyzer (Farha and Magdy, 2019), as metaphor is one of the most used cognitive devices. The cognitive devices are simile, metonymy and metaphor.

In addition, the performance of the state-of-the-art Arabic sentiment analyzer is satisfactory in terms of predicting the polarity of sentences without metaphor, but the performance of the sentiment analyzer (Farha and Magdy, 2019) is inadequate in terms of polarity prediction with sentences containing metaphors. Even though the sentiment analyzer did not contain resources or algorithms for dealing with metaphors, my purpose was to test how neglecting metaphor detection significantly affects the result. The result of ignoring metaphor in detecting sentiment will negatively affect the trustworthiness and reliability of sentiment identification, affecting the tool's prediction accuracy and reliability. For instance, the wrong analysis for a company that wants to know about opinions of its particular products or services would lead to false opinions about their products and thus cause the company to face financial loses. In addition, a false opinion about a product will mislead potential buyers into making a wrong decision to purchase.

[1]https://www.oxygenxml.com/

1.4 Research questions and objectives

Based on the initial investigation of metaphor using Mazajak in Alsiyat and Piao (2020a) and the evidence and discussion set out in the previous paragraphs, I will achieve the research objectives and address the research questions discussed below.

1.4.1 The research questions

The research questions concerns data analysis to observe online Arabic metaphor patterns and structures to build a base knowledge for Arabic metaphor corpus. In addition to using the Arabic metaphor corpus to show the impact of the metaphor on sentiment using existing Arabic metaphor tools. So, the main objectives of this research is to show the novelty of the Arabic metaphor corpus by describing the data structure. In addition to building the corpus as a foundation for Arabic metaphor identification purpose. Hence, show the impact using the existing Arabic sentiment analyzers tools. The objectives can be divided to sub-tasks to achieve the main aim. For example, evaluate the existed Arabic tools was essential to reflect on the impact of sentiment on Arabic metaphor. The research questions (RQs) fall into two types, one of which will be answered by analysis of data observations and the other on practical evidence.

The following research questions correspond to the data collection chapter:

RQ1 What is the most frequently appearing structure of online Arabic metaphors? How can these pattern extract sentiment and identify metaphors to design a schema for annotation and feature extraction in future studies? Arabic online metaphor is ambiguous and the frequent pattern involves context words with polarity in the literal sense that describes metaphor. Arabic online metaphor is an informal form of writing, which means that there is no fixed structure. As an example of the Arabic online metaphor new structure, such as a metaphor with no context provided. In addition, there is a new Arabic dialectal metaphor appears in the form of proverbs which can be described as one sentiment with one meaning.

RQ2 What are the Arabic metaphor terms that identify metaphor in the online context? As discussed above, the Arabic online metaphor is essentially a spontaneous form of writing, which identifies the text as unpredictable and informal in terms of its structure and meaning. The characteristics discussed above clarify differences in the online writing style from the relationship between changes in the online writing style and in sentiment detection. For example, the metaphor could be interpreted as voice semantics to express opinion to detect the sentiment. Such as فسسسس and يخخخخخ means *'Yikhkhkhkh'* and *'fsss'* in the literal sense. In a metaphorical sense they mean *'bad/stupid'* and *'disappointment/sad'* respectively, based on the annotators' annotations. Those new words are used in online contexts as metaphors using sound and pronunciation to express opinion. The transliterated for those words followed using the ALA-CL(American Library Association- Library of Congress), a method to translate the non-Latin words into Latin words. Also, I double check the words in `https://transliterate.arabicalphabet.net/?text=%D8%A7`, which is a website to transliterate Arabic words into English using the glossary of the Arabic words that include sound on the English letters.

RQ3 Does the data analysis for online Arabic metaphors align with the newly created annotation scheme? How can the annotation schema be adjusted for the unpredictable style of online Arabic metaphors to effectively capture the practical purpose of metaphor annotation prior to the sentiment annotation? This is reflected in the newly created schema and the challenges faced during the annotation, which is influenced by the other studies such as VU Amsterdam Krennmayr and Steen (2017) for metaphor annotation and practical purpose to identifying metaphors in online context. The schema was designed to identify the metaphor in an online context before the annotation. Because the Arabic online metaphor has multiple characteristics and ambiguity to be identified and analyzed to annotate the accurate sentiment. The sentiment annotation added to the corpus in Excel format.

Those research questions will be addressed in the chapters discussing corpus annotation and the impact of metaphor on sentiment.

RQ4 How can the existing schema for the English language for metaphor contribute to designing the Arabic online metaphor with sentiment schema? Multiple studies have discussed the schema of metaphors for annotation in English and Arabic such as Krennmayr and Steen (2017) and Abugharsa (2022). However, not all of them fit for this research aim, while those research influences the newly created schema for Arabic online metaphor.

RQ5 What are the web-based Arabic sentiment tools available that can be used to show the impact of the Arabic Metaphor Corpus (AMC) on sentiment? What is the impact of metaphors on sentiment annotation using the AMC in different automatic Arabic sentiment tools? This question reveals the lack of development in Arabic sentiment analysis as one of the NLP problems to be solved. The impact is shown in the differences in metaphor prediction and in the tool evaluation, highlighting the inadequacy of Arabic sentiment analysis tools in predicting sentiment in relation to metaphors using the Arabic Metaphor Corpus (AMC).

RQ6 How can a tool be designed to classify sentiment in the absence of available Arabic sentiment analyzers that incorporate the Arabic Metaphor Corpus (AMC)? This discusses the tool specifications for identifying the sentiment of reviews using the semantically tagged Arabic Metaphor Corpus (AMC) from AraSAS. However, the tool was evaluated to achieve better sentiment classification for the AMC.

RQ7 What is the best method to show the impact of metaphor between the sentiment annotation tools used, which are assessed by the gold standard annotation using statistical information from the standard measurement F-score?

1.4.2 Research Objectives

This project has the following research objectives (ROs):

RO1 To extract Arabic terms containing online metaphorical expressions from a given effective text that is publicly available.

RO2 To produce a new annotated Arabic metaphorical expressions corpus from an online context.

RO3 To find the available web base sentiment Arabic tools to show the impact.

RO4 To Show the impact by evaluating the Arabic sentiment tools predictions for sentiment using the AMC.

RO5 To annotate and evaluate the AMC with semantic using the Arabic semantic tagger El-Haj et al. (2022).

RO6 To design a tool to find the sentiment using the AMC with assistance of the AraSAS El-Haj et al. (2022) as a consequence of the lack of the sentiment classification feature to the Arabic semantic tagger.

1.5 Research Contributions

I shall address the SA associated with metaphors, which can be considered as a text classification problem. As discussed above, SA can be annotated at multiple levels in a given text, making it a much more difficult. Moreover, recognizing metaphors in real-world discourses can be challenging because Arabic language perception is largely intuitive. And there are no resources annotated for Arabic online metaphor with sentiment for proper interpretation and detection of Arabic metaphor. This PhD research project has the following novel attributes:

1. The corpus will be the first Arabic language resource annotated with sentiment information-on Arabic metaphor expressions with context, meaning, theme, metaphor type and semantic.
2. The first designed tool classifies the sentiment in association with semantic information using the Arabic metaphor corpus (AMC).

3. The first research evaluates and assesses the state of art online Arabic sentiment tools using metaphor information to show the impact of metaphor on sentiment.

4. The corpus will provide a new type of language resource for training and evaluating SA algorithms and tools involving sentiment and metaphors.

1.6 Methodology

As discussed above, this study is designed to investigate Arabic metaphors regarding sentiment analysis in an online context. As has been explained, the Arabic metaphor is being newly investigated in this research. There is no previous work to follow for Arabic metaphors in regard to identifying sentiment. In contrast, previous studies for the English language have investigated approaches to identifying metaphors with sentiment using previously built resources. Metaphors in Arabic will be explored by building annotated data for the first time because there are no resources annotated for Arabic metaphors. The principal stage of this research is therefore building an Arabic metaphor corpus. In addition, although there are various methods for detecting English metaphors, using the English language approach is not a realistic solution to identifying Arabic metaphor. As explained, metaphor identification and sentiment classification need to be performed using reliable Arabic metaphor data. Still, there is currently no such data for identifying Arabic metaphor in relation to sentiment. Also, due to the language differences, the English approaches might not be accurate for the Arabic language. We tested Arabic metaphor data using automatic Arabic sentiment systems, but there are few available Arabic sentiment analyzers to use, such as Mazajak Farha and Magdy (2019) and Sentistrength Thelwall et al. (2010). The tool designed in this thesis is to identify sentiment using an Arabic semantic tagger AraSAS El-Haj et al. (2022) tags on an Arabic metaphor corpus. Also, the Arabic semantic tagger was used to identify sentiment and to detect metaphor in future work. Also, the Arabic semantic tagger was used to identify sentiment and to detect metaphor in future work. Also, human annotation is not just used as a gold standard for comparison; it is used to identify online Arabic metaphor as Arabic metaphor has a new structure. So, the sentiment identified by human as Arabic metaphor term might be not recognized by the automatic Arabic systems. For example, the word بيض means *'Eggs'*, which means *'bad atmosphere'* in a metaphorical sense, automatic systems(Mazajak) recognized the polarity as neutral, whereas the word has a distinctly negative polarity in the human annotation.

The methodology devised for this study has therefore been chosen as a foundation from which to identify sentiment and metaphor in Arabic texts. As already explained, metaphor detection with sentiment needs reliable resources, but there are no such resources for analyzing Arabic metaphors. In addition, the corpus created has a new structure that is different from the regular one, which means that it needs more analysis. The data for the new Arabic metaphor corpus has to be chosen carefully. New metaphorical terms must be selected based on social media's most frequently used Arabic informal terms. The chosen methodology therefore has to start by building an Arabic metaphor corpus: it is consequently- hence necessary to:

- Collect reviews that contain informal metaphorical terms frequently used in social media from a large-scale Arabic dataset;
- Specify the new Arabic metaphor terms in the collected reviews.
- Collect the reviews by choosing Arabic native speakers.
- Specify the dialect of the chosen Arabic metaphor terms.
- Analyze the Arabic metaphor structure and pattern to specify the Arabic metaphor nature in an online context and specify the accurate annotation scheme.

The annotation scheme was specified by the data analysis to build the Arabic metaphor corpus in an online context, following the below steps:

1. Find two native Arabic speakers to annotate.
2. Annotate the Arabic metaphor terms from the collected reviews that contain the new Arabic metaphor in an online context.
3. Find the sentiment for the metaphor expression section and the overall review separately.
4. Annotate the context and the part of speech for the metaphor section and context as well as the theme, meaning and metaphor types.
5. Annotate the sentiment of the Arabic metaphor expression and the sentiment of the overall reviews as gold standard annotations.
6. Produce the newly created Arabic Metaphor Corpus (AMC) annotated with metaphor, sentiment, meaning, context, theme, type and part of speech.
7. Test the corpus reliability by applying the AAI.

Three different methods will also be used to show the impact of the (AMC) on sentiment using a state of art Arabic sentiment analyzer. The methods include the semantic information associated with designing an initial tool for identifying the sentiment. It is therefore necessary to

- Annotate the Arabic metaphor corpus for sentiment using Mazajak (Farha and Magdy, 2019), which is an online Arabic sentiment analyzer;
- Tag the Arabic metaphor corpus for semantic using an Arabic semantic tagger (El-Haj et al., 2022) to build a sentiment tool for identifying sentiment based on the semantically tagged AMC;
- Test the tool's performances using the standard measurements in comparison with the gold standard annotation. The standard measurements are the precision, recall and F-score against the gold-standard annotation;
- Compare the F-scores of the four methods to show the impact of the AMC on sentiment. The results will be presented as statistical information.
- The impact is discussed by the differences in the method's predictions on sentiment using the AMC. In addition, the comparison of the automatic and manual annotation resulted from the AMC annotations as statistics.
- The methods were classified as methods considering metaphor and methods not considering metaphor. Because the metaphor sentiment information used in some methods, not the other.

Each stage of the chosen methodology depends on the previous stage, and each stage has its own specifications and distinctive analysis. The data collection is novel regardless of the corpus building. The novelty of the data is demonstrated by the analysis of the Arabic online metaphor patterns because the data analysis is necessary to understand and perform accurate sentiment and metaphor detection in general, and the data source for the Arabic metaphor corpus is the large-scale Arabic SA data collected. For example, the data analysis proves that metaphor cannot be affected by the context. This is because the online Arabic metaphor could appear as the same term, which is usually known from the context. For example, *'Your eggs are on the table'* is literal whereas *'Your atmosphere is eggs'* is metaphorical. So, the context distinguishes literal from metaphorical sentences. In the case of online metaphor, however, a metaphor could appear alone without any context. The particular word could be followed by some context, but a reader might not be able to interpret the metaphor. For example, the same word *'Eggs'* comes in the following sentence: بيييييييض. من وجهة نظرى يعم, which means *'Eggs. from my point of view my uncle'*.

The word *'Eggs'* in this example is followed by a full stop, signifying that the opinion ends after the metaphorical word. Such analysis for a metaphor structure in online writing is new and therefore needs an appropriate tool to be designed. In this study, an online Arabic metaphor is investigated to produce data analysis for the practical work. We extracted features from the data to analyze metaphorical words with their supporting contexts in order to identify the sentiment involved because there are no existing techniques for the reasons given above. The analysis could be promising for enabling classification using the specially designed tool and could also be useful for establishing accurate criteria for annotation. For example, after observing many reviews, it was possible to choose the metaphor that has the main opinion.

For this project, I created, investigated, and tested a newly annotated metaphorical corpus with SA information for the Arabic language. Next, I designed an Arabic SA tool based on existing Arabic resources using this new corpus.

1.7 Thesis structure

Following this introductory chapter, the remainder of the thesis will be organized as follows.

1.7.1 Chapter 2: Literature review

Presents an review of previous studies conducted for Arabic and English metaphor identification and Arabic SA. I shall compare and contrast the existing studies to identify the research gaps. The literature review covers studies in overlapping fields to show the progression of Arabic and sentiment metaphors. Some of the challenges involved are known from previous studies, such as the distinctive Arabic dialects and morphology, and new challenges were faced in the use of new informal terms of Arabic online metaphor with no reliable source for discerning meaning. Other problems are the lack of algorithms explicitly developed for Arabic and the issue of using the English language approaches for Arabic. The literature also offers explanations of the concepts of Arabic metaphor and SA. The discussion reveals the reliability of the resources on which English language studies rely to identify metaphor and sentiment. A survey in section 2.3.1 introduces the aspect level of SA.

> Israa Alsiyat and Scott Piao (May 2020a). "Metaphorical Expressions in Automatic Arabic Sentiment Analysis". English. In: *Proceedings of LREC2020 Conference*. The 12th Edition of the Language Resources and Evaluation Conference, LREC2020 ; Conference date: 11-05-2020 Through 16-05-2020. European Language Resources Association (ELRA), pp. 4911–4916. URL: `https://lrec2020.lrec-conf.org/en/`

1.7.2 Chapter 3: Data collection

I shall discuss the data collection, the Arabic sentiment data sets and the chosen data to collect metaphorical reviews in section 2. In section 3, the criteria for data collection will be discussed based on the examination of the chosen dataset. Additionally, I shall describe the structure and meaning of the Arabic online metaphor with translations. The chapter presents a novel categorization of Arabic online metaphor whether or not there is an available source. The terms are grouped in semantic topics based on their literal use. Some of the meanings of those terms are verified by the findings from a questionnaire completed by 107 Arabic native speakers in Egyptian and Saudi dialects. Still two annotators authenticated the remaining terms. The chapter concludes with a discussion of the limitations of the study and an explanation of the purpose of collecting metaphorical reviews rather than metaphorical sentences.

1.7.3 Chapter 4: The corpus annotation chapter

The corpus annotation chapter then provides annotation guidelines and schema. Designing the schema starts with XML tags but does not follow any of the standard schemata such as SemEval and the MIPVU protocol, however, it was slightly influenced by the VU Amsterdam Krennmayr and Steen (2017) in terms of the tagging type (XML) because the purpose is not to add the new corpus to any existing corpus system such as SemEval. The intention is rather to build an Arabic online metaphor corpus for practical and research purposes. Even so, the schema designed for this study was intended to fit the practical part of this research, so the XML schema was converted into an excel sheet for the practical work. Also, the vu Amsterdam specified the overall metaphor annotation for each sentence, whereas in the new annotation, we specify the metaphorical expression and word with the part of speech. To assess the reliability of the annotation, the inter-annotator agreement (IAA) was calculated. In the section on annotation challenges, the challenges for metaphor, sentiment and meaning are discussed. It was found that the meaning affects the manual annotation of sentiment. For example, the same metaphorical terms in different contexts had different meanings. The first Arabic metaphor corpus annotated with the overall metaphor sentiment with meaning, context and genre is described. The work of this chapter has been presented as an abstract in the CL2023 Conference as:

> Israa Alsiyat, Scott Piao, and Mansour Almansour (July 2023). "Arabic Metaphor Corpus (AMC) with Semantic and Sentiment Annotation". English. In: The twelfth International Corpus Linguistics Conference, CL2023 ; Conference date: 03-07-2023 Through 06-07-2023. URL: `https://wp.lancs.ac.uk/cl2023/`

1.7.4 Chapter 5: Impact of AMC using existed automatic Arabic sentiment analysis

I test the Arabic metaphor corpus using three different automatic methods. Manual annotation and automatic Arabic sentiment analyzers were compared using the standard measurements.The purpose is to show the impact of Arabic metaphor on sentiment using state-of-the- art Arabic sentiment analyzers. In addition, all of the methods are compared with the gold standard, the manual annotations for metaphor and overall sentiment. The performance of the Arabic sentiment analyzers is evaluated and discussed in terms of the improvements that those tools need to use with the Arabic metaphor corpus.

This section describe the gold standard annotation in detail and then the automatic annotations for the Arabic metaphor corpus. For the automatic annotation, the Arabic semantic tagger was used because it has the same categorization groups for the informal Arabic metaphorical terms and has no sentiment classification feature. The tool was designed to classify sentiment based on the semantic tags for the Arabic metaphor corpus and the classification was done after the Arabic metaphor corpus had been tagged. The method was tested using standard precision, recall, and F-score metrics. The results were compared and discussed with the gold standard annotation to show the impact of Arabic metaphors on sentiment using automatic Arabic systems. In addition, the automatic performance was compared with the gold standard annotation using the acquired statistical information.

Israa Alsiyat (Apr. 2025). "Arabic Metaphor Sentiment Classification Using Semantic Information". In: *International Journal on Cybernetics and Informatics (IJCI)*. URL: `https://ijcionline.com/volume/v14n2`

1.7.5 Chapter 6: Conclusion

This chapter illustrates the research summary, highlighting the main idea of this study. It also details the research achievements, describing the aims accomplished. Additionally, the chapter discusses the findings based on observations and results. Suggestions based on these results are provided, reflecting the three main aims of the research: analyzing Arabic metaphors in an online context, building a corpus, and examining the impact of Arabic online metaphors using automatic tools. Some suggestions are implemented and supported by evidence.

For instance, feature extraction resulted from corpus building, and code suggestions were derived from the methods used to demonstrate the impact of metaphors on sentiment analysis. The findings extend beyond the initial targets. The limitations section addresses the constraints encountered and proposes solutions for each chapter. Finally, the future work section outlines the research direction moving forward.

Chapter 2

Literature Review

Corpus building is a process that employs human knowledge to identify text of specific subjects and annotate the data with various useful information, particularly linguistic information. Metaphor is one of the challenging linguistic devices to interpret using computer algorithms. Researchers have therefore competed to design an algorithm that can identify metaphors. The hypothesis Wilks et al. (2013) for the identification of metaphors is based on the agreement between verb and noun with respect to familiarity. For instance, if the verb 'married' occurs with the noun 'brick', the verb/noun agreement is violated and the structure is considered a metaphor. This hypothesis is a possible solution for metaphor detection in Arabic, but the lack of an Arabic metaphor resource makes identifying the metaphor problematic. However, sufficient metaphor corpora are available for the English language, such as Krennmayr and Steen (2017) and S. Mohammad et al. (2016). The existing resources for identifying metaphors in the English language are the lexicon, annotated corpora, and algorithms. The lexicon is a resource containing a large number of words categorized based on word usage and part of speech; some are used solely for one purpose. For example, VerbNet Schuler (2005) identifies verbs with different categories based on their usage. Another example, SentiFig Rentoumi (2012), is a lexicon built for English to identify sentiment in figurative data. Corpus annotation is a considerable amount of data labeled manually to define a linguistic feature. For example, the VU Amsterdam corpus Krennmayr and Steen (2017) was created to identify English metaphors in different genres. Moreover, the metaphor corpus available for English is built in different text domains; for example, POLITICS, ECONOMY, and FOOD are separate metaphor domains such as Ana (1999) and Izwaini (2003). English also has algorithms that help to identify metaphors, such as Word Sense Disambiguation (WSD) and built lexicons such as WordNet Miller (1995).

There is no reliable Arabic online metaphor resource available to identify metaphors the same as those in English. In addition, there are no algorithms that help the automatic identification of metaphors in Arabic. Arabic metaphor in the online context needs a previous annotation. Because there are new Arabic metaphor terms being used online. In this review of the literature, previous studies are discussed to identify and highlight research gaps. The studies compared and contrasted in this review have the potential to be used in the current study. Describing previous research shows the novelty of the current study. Therefore, this chapter will discuss the research gaps identified in previous studies.

2.1 Metaphor studies and sentiment

Studies of English metaphor data will be compared and contrasted with previous studies of Arabic metaphor, which will clearly show the research gaps identified. The discussion covers the studies most related to the current research on Arabic metaphor with sentiment. It will cover three specific subjects: Arabic metaphor with sentiment, English metaphor, and Arabic SA. The absence of studies conducted on Arabic metaphor in relation to sentiment as a resource means that the gaps and challenges are spread across these three research subjects. The reliable existing lexicon resources, annotated corpora, and algorithms devised for English facilitate the process of identifying metaphors, such as BNC Leech (1992) British National Corpus, WordNet Miller (1995), VerbNet Schuler (2005) and TreeBank Marcus et al. (1993). As an example of existing algorithms, Word Sense Disambiguation (WSD) helps identify sentiment and metaphor in English in Rentoumi (2012). Still, although WSD has many applications for Arabic, it does not have any applications for metaphor Hadni et al. (2016).

Previous studies have discussed metaphor conceptually and computationally from different perspectives. For example, S. Mohammad et al. (2016) discussed the relationship between metaphor and emotion using data from WordNetMiller (1995), and S. Mohammad et al. (2016) followed the annotation approach Miller (1995) to identify the target and the domain for metaphor. In addition, Rentoumi (2012) discussed sentiment driven by metaphor, principally the correlation of sentiment and metaphor using WSD. That study also used SemEval 07 data annotated with sentiment to collect a thousand article titles from news reports and discussed the basic concept of metaphor, which is different from the metaphor that is the topic of the current study. That study also discussed the theoretical aspect of the concept of English metaphor. The methodology consisted of WSD, sentence-level polarity assignment (SLP), and sentence-level polarity detection. Rentoumi (2012) mentioned above collected metaphorical data annotated with sentiment and identified metaphor using a previously designed algorithm. Although those algorithms apply to the Arabic language, the newly devised Arabic metaphor corpus has a different structure from the regular Arabic metaphor, so the application's accuracy could be impossible unless the algorithm is adapted to the Arabic online metaphor. As mentioned, the newly created Arabic metaphor corpus was collected from the online context, which has a different structure.The online context has an unpredictable writing style that differs from the regular Arabic metaphor structure. For example, online metaphors in Arabic use the same term to express opinion, and the challenges of the writing style of Arabic online metaphors affect both the annotation of metaphors and the sentiment. For example, punctuation and typography affect the annotation and the tool to detect the metaphor. For example, during the annotation, the position of punctuation marks affects the meaning of the metaphor, as in بيييييييض. من وجهة نظرى يعم *'eggs. from my opinion, man!'*. The metaphor expression بيييييض *'Eggs'* appears as one expression followed by a full stop, making it understood as an opinionated review. Another example of the effect of punctuation is when punctuation is replaced by a letter to tag the Arabic metaphor corpus. All this means that online text has more factors to be considered than just metaphors. To illustrate, Rentoumi (2012) has a similar concept of showing the impact of metaphor in English using different methods. However, that study used previously annotated data with sentiment SemEval 07 Strapparava and Mihalcea (2007) with a regular metaphor structure. In contrast, the current method involves building a reliable Arabic online metaphor corpus annotated with both sentiment and meaning, hence the need to design and test the data using state-of-the-art Arabic taggers. In addition, Arabic and English have different structures, and adjustment could be crucial in such cases.

In addition, more studies discuss the metaphor conceptually and computationally from different perspectives. For example, S. Mohammad et al. (2016) discussed the relation between metaphor and emotions using a data-driven WordNetMiller (1995). S. Mohammad et al. (2016) follows the Miller (1995) annotation approach to identify the target and domain of the metaphor. Rentoumi (2012) discussed the sentiment driven by metaphor. In general, the study discusses the correlation between sentiment detection and the existence of metaphors using WSD. In addition, the study used SemEval 07 data annotated with sentiment to collect one thousand pieces of information from the news. Also, discuss the basic concept of metaphors, which is different from the metaphor aimed at this research.

In addition, the study of Rentoumi (2012) discusses the theoretical aspect of the concept of English metaphor. The proposed methodology contains Word Sense Disambiguation (WSD), sentence-level polarity assignment (SLP), and sentence-level polarity detection. The study Rentoumi (2012) described above collected metaphorical data annotated with sentiment. In addition, metaphors can be determined using previously designed algorithms. However, these algorithms could be applicable to the Arabic language. However, our Arabic metaphor corpus has a different structure from the regular Arabic metaphor. So, the accuracy of the application could not be achieved unless the algorithm adjusts to adapt the Arabic metaphor online.

2.1.1 Arabic metaphor resources

There are sufficient available datasets built for Arabic metaphors in different fields. However, the Arabic metaphor data created for the current study were on a specific domain and field with limitations enabling it to be used in this research. Alowisheq et al. (2016)

stated that:"There are no reliable and adequate lexicons or annotated corpora for the Arabic language," making the metaphor extraction process from Arabic texts challenging. However, there are resources for Arabic metaphor identification created for the linguistic and social science fields Gholami et al. (2016) and Faycel (2012). These studies are mentioned as available resources for building the Arabic metaphor corpus. Regarding Arabic translation, Gholami et al. (2016) investigated Arabic metaphor translation into English following Jakobson and Halle's Jakobson (1956) theory of translation carried out by two native speakers of both languages. This dataset of translation does not fit the current study as the data were built for a religious domain so the metaphor has the Arabic classical structure whereas the current study is focused on online metaphor which is widespread in social media. That study was limited to the religious domain whereas the current study concentrates on the terms for social media which could be used in any domain. For instance, the previous example بيض *'Eggs'* which is one of the metaphorical terms in my new corpus (AMC) that the term could be used on the X platform to describe something. Still, this study does not use evidence from the X social media platform because of X's privacy restrictions. Similarly, Al-Harrasi (2001) applied Arabic metaphor translation into English to political speeches, analyzed and explained metaphor conceptually using translation as evidence. Another linguistic study by Faycel (2012) employed a corpus-based approach to study Arabic Tunisian metaphors about food in the proverb domain because Arabic proverbs are full of metaphors as they reflect Arabic history and culture. Faycel (2012) classified the proverbs in the food domain from human physiological and psychological perspectives and for social science purposes. In addition, the nature of the data was that it had fixed opinion because each proverb has a background culture or story so their metaphorical sentiment polarity is fixed. So, making a base resource for Arabic metaphors with sentiment for Arabic proverbs would be useful, but automatic metaphor identification could not be used. Even so, it could be used to train the tool, but not for the automatic identification to identify metaphors like online metaphors because, as has already been explained, Arabic online metaphor uses the same terms with different meanings and sentiments, which makes it more challenging. Raii (2009) examined Arabic metaphors in dialectal discourse in the linguistic field and discussed changes in Arabic metaphor structure in different contexts. However, the researchers did not mention the data resource collection other than speech but discussed the nature of a random structure which did not have the same regularity as written text. The online context has been and uses metaphors as part of its basic structure (Raii, 2009). That study categorized verbal Arabic dialectal metaphor into different categories based on the metaphor usage. For example, they showed that a metaphor can occur as a verb to describe the onset of a new season, as in ايلول صار عاللبواب which means *'Yaloul has come to the doors'*. That study had a similar concept to the changeable Arabic metaphor structure but did not discuss Arabic metaphor based on basic and reliable information as the data source was not provided.

The only study that addresses Arabic metaphor computationally without incorporating sentiment analysis is (Abugharsa, 2022). The approach proposed by Abugharsa (2022) is effective within the context of the Libyan dialect and poetic domain; however, it may not be applicable to online Arabic metaphors, which often exhibit unfixed and dynamic structures. Moreover, the focus of my work differs significantly, as it incorporates sentiment analysis within the metaphor classification task. While Abugharsa (2022) employ a binary classification method for metaphor detection alone, our tool Alsiyat (2025) is, to date, the only system that integrates both metaphor and sentiment analysis for Arabic. Our approach shows promise in reducing human intervention in the pre-annotation process by leveraging semantic information from the Arabic Semantic Tagger El-Haj et al., 2022 to classify metaphor-associated sentiment (AMC). Importantly, the AMC framework was essential for enhancing the accuracy of sentiment classification.

In other languages, metaphor classification often relies on various supporting resources. For example, metaphor detection in Polish utilizes robust, language-specific resources (Wawer et al., 2017), whereas studies on Chinese metaphor often translate texts into English to make use of existing English-based tools (Peng et al., 2018). However, such translation methods are limited in accuracy and fail to capture important linguistic features, including aesthetic and cultural nuances. In the case of Arabic, comparable reliable resources for metaphor classification are still lacking.

2.1.2 English metaphor detection

In this section, I shall discuss metaphor detection in English and some of the metaphor detection techniques for other languages to demonstrate the new advanced techniques currently in use. The advanced algorithms and the reliable resources designed for the English language show the lack of Arabic language resources and algorithms.

Regarding English metaphor detection, computational studies are divided into two general approaches: using resources (lexicons or annotated corpora) and developing classifiers that do not rely on resources. To my knowledge, no work on detecting and analyzing Arabic metaphors in Arabic discourse has been published. Furthermore, no studies have examined English metaphors as a corpus with sentiment, meaning, context, theme, and metaphor type information. As mentioned earlier, the number of reliable English lexicons facilitates the process of analyzing and identifying metaphors in English, whereas there are very few reliable Arabic lexicons to use for identifying metaphors. Having an Arabic metaphor resource is the main step in identifying metaphors, and the algorithm should be designed to fit the Arabic online text. The structure of an Arabic metaphor corpus could affect the algorithm performance in identifying Arabic metaphors. There is, however, an algorithm built specifically to help to identify metaphor based on the size of the data, this is Word2Vec. Word2vec can predict the metaphor based on the frequent occurrence of a metaphor in specific contexts, but the prediction cannot be achieved unless there is a reliable resource annotated for the Arabic metaphor to train it. Arabic metaphor in the online context needs a previous definition because there are new Arabic metaphor terms used in the online context.

In addition to the metaphor resources for English, there have been different attempts to achieve automatic metaphor identifications. Next, we discuss studies conducted computationally for the English language in regard to the identification of metaphors and consider their methods to identify the differences and similarities between the various applications. As explained above, studies of English metaphor detection can be divided into two general approaches: supervised and unsupervised detection. It was suggested by Rai, Chakraverty, and Devendra K. Tayal (2016b) that linguistic metaphors can be detected using natural language processing (NLP) by analyzing the properties of the surrounding words. They specifically proposed using conditional random fields (CRF) for the accurate detection of metaphors. In that study, a comparison approach was used comparing CRF classifications with previously used methods employed by [Dunn (2013), Klebanov et al. (2015) and Hovy et al. (2013)] each of which used a different approach to analyze the detection of metaphors. The experiments used the VU Amsterdam metaphor corpus as the data set (Krennmayr and Steen, 2017). The CRF was interpreted using the syntactic, conceptual, effective, and word embedding characteristics taken from the MRC psycholinguistic database (MRCPD) and WordNet-Affect. One concern about the comparisons is that the data set was not the same for all the studies: Klebanov et al. (2015) used the VU Amsterdam Metaphor corpus, whereas the others did not. The results would have been more reliable had the data set for all the comparison models been the same. Furthermore, it is essential to note that the current study was focused on metaphors in the Arabic language, while the research by Rai, Chakraverty, and Devendra K. Tayal (2016b) looked at metaphors in English. However, the methods proposed in this approach could be recreated and justified using Arabic data. However, as already explained, the lack of reliable Arabic resources is problematic. In addition, Arabic metaphors can differ depending on the context. The Arabic metaphor corpus compiled for the current study has an unstable structure as the metaphors were collected from the online context, which has a free and informal writing style. So, if the method could be adapted, a justification might be needed, unless the metaphor has a standard structure, which is found in classical Arabic texts. The classical Arabic metaphor context is found in the text of the Qur'an, which is miraculous and inimitable, yet unchangeable. This means that because the Qur'anic metaphor appears in fixed writing, machine learning could be trained to identify and adapt the metaphors. However, the text of the Qur'an is considered challenging for translation to be understood in English (Zeroual and Lakhouaja, 2018). The results showed that the CRF classifier proposed by the researchers was more accurate and precise than other classifiers. When using the VU Amsterdam metaphor corpus, the accuracy rate was 92%.

Do Dinh and Gurevych (2016) investigated automatic metaphor detection. The researchers recognized that typical metaphor detection systems used selection preference violations or concreteness ratings. These are based on hand-coded rules specific to the language that is to be used. They used word embedding trained on large corpora as a new approach and found that this approach produced results that were comparable with other systems but did not require any extra resources.

In the experiments used by these researchers, Word2Vec trained word embedding was used. The test data were from the VU Amsterdam Metaphor Corpus (VUAMC). The system was trained on specific genres such as academic, fiction, conversation and news. It was recognised during the experiments that conversational texts were the most challenging in terms of the detection of metaphors. Overall, the findings showed that the system worked well with English data but needed improvements for use with other languages. The system performed favourably compared to other systems which require additional resources. However, the system was limited to the word embedding coded with it. More research should be done to apply technology to enable the system to adapt and continually add new word embedding automatically. The researchers suggested further research into incorporating more advanced structures such as Recurrent Neural Networks and Long-Short Term Memory Networks into their system (Do Dinh and Gurevych, 2016). The studies discussed above dealt with the actual application of metaphor detection using various advanced techniques for English regardless of sentiment. The advanced techniques were neural network, CRF, word embedding and machine learning, but those studies were restricted to reliable resources such as VU Amsterdam, TreeBank, VerbNet, and WordNet, so their focus was on detecting metaphors, not on building a resource. Moreover, studies with other languages have been conducted to identify sentiment in metaphorical data. For example, Peng et al. (2018) is one of the studies which used the advanced technique of Long Short Memory network LSTM. The study used a manually labelled dataset of metaphor (target) and context. The researchers trained the annotated dataset using word embedding of unlabelled Chinese data to identify sentiment. However, the word-embedding data had no source to acknowledge the data type, so the word embedding used as metaphorical data to recognize sentiment in new annotated data could not do so. The LSTM technique was compared with manual annotation and the findings proved the significance of the contextual information in identifying metaphor and the importance of the LSTM as a baseline for future work. Even so, that study showed that the technique could not distinguish between literal and metaphorical information to identify sentiment. Also, similar to the current study, they used this method to demonstrate that context information is necessary to identify a metaphor. The current study, however, has proved that metaphor can be recognized without context in a known domain, although the context will be annotated as base knowledge for future practical purposes. The previous analysis mentioned for the study was discussed despite the structure of the Chinese language. The study applies an experiment to identify metaphors in the Chinese language after translating the text into English. From the discussion above, studies on identifying metaphors in English have relied on existing resources built for the English language. Even unsupervised studies have relied on existing algorithms designed for English metaphors. This means these algorithms might show different results and accuracy when applied to Arabic online metaphors. For example, Do Dinh and Gurevych (2016) used word embedding with the WordNet lexicon as an unsupervised method. However, this study still relies on resources like WordNet, which are not available for the Arabic language. There is no similar resource for extracting information from Arabic online metaphors.

Arabic has no reliable resources for metaphor identification, which raises difficulties for the detection of metaphors in Arabic. The automatic identification of Arabic metaphor is therefore challenging, so the ideal approach to identifying Arabic metaphor is to build a specific resource before attempting automatic metaphor detection. In contrast, the previous studies sought to solve the scarcity of language resources by designing unsupervised classifiers for detecting metaphor. Moreover, the existing studies focused on identifying and finding the meaning (synonym) for the purpose of translation. In regard to the studies exploring Arabic metaphor from the conceptual prospective, Zeroual and Lakhouaja (2018) discussed the different types of standard Arabic metaphor in different contexts. The online Arabic metaphor has a dynamic structure which fits the nature of online communication, and the challenges of translating metaphor were discussed in terms of the dialects and the Arabic types. Zeroual and Lakhouaja (2018) discussed each Arabic metaphor type in different Arabic forms such as MSA, classical, and dialectal and suggested a method for identifying Arabic metaphor using the right questions to advance the metaphor for Arabic in the NLP field. Metaphor in the Qur'an is one of the classical Arabic forms, and the context is equally important in MSA Arabic. Zeroual and Lakhouaja (2018) stated that: "the Arabic metaphor needs deeply structured knowledge in order to interpret the analogy in relation to the concept." They also discussed Arabic metaphor in the Qur'an as a unique source from which to construct a translation system. Arabic metaphor in the Qur'an has a fixed structure in terms of corpus building because the Qur'an is not changeable, whereas Arabic metaphor has a variable structure. In Alsayat and N. Elmitwally (2020) discussed the challenges of Arabic sentiment analysis on different linguistic levels. They emphasized the figurative level as a consequence of the lack of resources for Arabic sentiment analysis. There are sufficient datasets available for studying Arabic metaphor in different fields, but the Arabic metaphor data was in a specific domain and field with limitations preventing it from being used in the current study. For example,Faycel (2012) studied Arabic metaphor in proverbs in the social science field. The proverbs were in the food domain in the Tunisian dialect. The acquired dataset had the spontaneous nature of the online context, which is similar to our Arabic metaphor corpus. However, the Arabic metaphor corpus built in the current study has multiple genres in different types of Arabic metaphor, whilst the metaphors in Faycel (2012) were in proverbs. Likewise, in Raii (2009)'s study, the dataset collected had spontaneous metaphors, which is more similar to our corpus data, but that dataset was collected from verbal speech in daily life and without any source mentioned or analyzed from the conceptual prospective. Also, the data collected in Arabic dialects reflects the changes in Arabic metaphor in everyday life. The findings of that study showed the data in different categories based on the apparent use of metaphor within a particular context. Also, the Arabic metaphors in Raii (2009) were analyzed for the linguistic field. Even though these are all potential studies for use in building the current corpus, they still lack the type of metaphor which is the focus of the current study, which is Arabic metaphor in the online context. In addition, they studied only specific domains such as food and religion. In regard to translating Arabic metaphors, Gholami et al. (2016) investigated Arabic metaphor translation into English from Islamic books. As already mentioned, they followed Jakobson (1956)'s theory of translation by two native speakers of both languages. That dataset is one of the contributions to the translation of Arabic metaphors, but that work was limited to a specific domain, which was religious books. Religious works are written in classical Arabic, which is different from the form used online. In addition, work has been carried out in the field of linguistics for psychological and physiological analysis. Similarly, Al-Harrasi (2001) applied Arabic metaphor translation into English to political speeches and investigated Arabic metaphor in the linguistic field. So although there are Arabic metaphor datasets available, their usefulness in the current study is still limited for the reasons given above. All of the current corpus was drawn from the book domain and gathered from a particular data set; it is still an online context which is similar to the social media (Twitter) context. So, our corpus might open a new perspective to be used and analysed.

The studies discussed in this section can be described as resources for concepts and awareness of the latest Arabic SA resources and techniques. Also, as mentioned above, Arabic SA is still developing compared to English. One of the Arabic SA data sets was used for the current approach, but the required data type was scarce. It was decided to collect online metaphor and annotate with sentiment, so the largest Arabic SA data sets were considered a priority, as they have more potential to contain online metaphor than other datasets. For example, Al-Ayyoub et al. (2017) annotated laptop reviews on ABSA, which is one possible data set available for use for this research, but the data in that study came from 1028 reviews on the literal meaning, so it does not have the same variety of LARB (large-scale Book Reviews) (Aly and Atiya, 2013). However, a very recent study conducted for the identification of Arabic metaphors with pre-annotated data without integrating sentiment classification is Abugharsa, 2022. Abugharsa, 2022 conducted a binary classification for metaphor identification without sentiment classification following a certain method used to adapt the LSTM (Long-Short-Term Memory) method of identification. The schema followed during the metaphor annotation is a standard schema for LSTM to identify metaphor. In our schema for this research, we annotate all aspects that could identify metaphor in online context.

2.1.3 Aspect level sentiment (ABSA)

ABSA in general involves three processes: extracting the aspect term, the aspect category, and the sentiment polarity information (Pontiki et al., 2016). In particular, due to the lack of freely accessible data sets created for ABSA and the immature state of investigation of Arabic SA Al-Smadi et al. (2015), the research in Arabic ABSA began only recently in 2015. However, there have been few works done for the ABSA in Arabic language, such as

Al-Smadi et al. (2015), Obaidat et al. (2015), Al-Smadi, Talafha, et al. (2018) and Ruder et al. (2016).

The previous works shown in Table 1 focused on applying various Arabic ABSA approaches to several types of Arabic writing such as MSA and dialectal. These works were conducted for ABSA whether or not they created a new corpus. The result was a series of technical developments rather than incorporating any new Arabic characteristics or creating a new data set. In my opinion, the explanation for this could be that researchers focused on replicating the research methodology developed for the English language instead of utilising Arabic characteristics. In detail, the ABSA for Arabic so far has mainly employed methods such as human annotation, SemEval schemata, NNR, SMV, STLM and deep learning. As a result, the SA technology progressed but the work did not explore new significant features of the Arabic language, such as metaphor. Also, as will be explained below, some datasets were duplicated, which shows the scarcity of such datasets. This has meant that the same annotated data set for ABSA was used in different approaches to identify sentiment. In addition, the schemata used for annotation have been taken from the English language schema Table 2.1. For example, Al-Smadi et al. (2015)'s work was the first research into ABSA for Arabic which applied human annotation using the SemEval-2014 schema with the BRAT tool (Stenetorp et al., 2012). Their dataset was a human-annotated Arabic dataset of book reviews, named HAAD, derived from the Large-Scale Arabic Book Reviews (LARB) set (Aly and Atiya, 2013). Obaidat et al. (2015) upgraded the ABSA for aspect category extraction T3 and aspect category polarities T4. The study relied on an earlier dataset HAAD Al-Smadi et al. (2015). The emphasis in Obidat was on identifying T3 and T4 by performing various lexicon-based approaches to address the shortcomings which appeared in every strategy. As an example, Al-Smadi and Obidat (year?) used the same HAAD dataset, pointing to the need for a publicly accessible dataset for Arabic. In Al-Ayyoub et al. (2017) collected Arabic Laptop Reviews (ALR) using the SemEval16-Task 5 schema for annotation. They used an assessment technique which gave an insight into n-grams and used a support vector machine (SVM) classifier to enable the researchers to measure and analyse their frameworks. ABSA made progress even though there is a shortage of accessible Arabic corpora, unlike the situation with English. To the best of my knowledge, there are four well-formed ABSA datasets available, HAAD, news of the Gaza conflict, ALR and the Arabic hotel review (SemEval16- Task 5). This literature review has shown that the works discussed above centred on breaking down the sentiment of Arabic texts and made use of ABSA by performing different approaches. Nonetheless, none of the previous studies sought to investigate 'between the lines' and show the figurative element of the Arabic language. That was clear when the SVM could not recognise the literary text during the annotation of ABSA for Arabic. The current study is therefore designed to investigate SA from the figurative perspective of Arabic and will focus on metaphors. I shall therefore focus on the application of Arabic metaphor with SA. The ABSA dataset mentioned above has the possibility of being used for this research, but it is small compared with the three biggest Arabic SA sets (Elnagar and Einea, 2016; Elnagar et al., 2018; Aly and Atiya, 2013). This means that the data does not have the diversity of the data in LARB to identify metaphor.

2.1.4 Document level and sentence level SA

In this section, I will discuss some of the SA studies which have generated Arabic data sets annotated at the sentence / document level, specifically the studies which used social media text regardless of the type of written Arabic. Almuqren et al. (2017) reviewed corpus annotation for Arabic SA and covered the studies conducted on Arabic for corpus annotation from different perspectives, such as the level of annotation and the tools used for annotation. Following that review, I shall discuss those studies and add further state-of-the-art sentiment studies. Studies of Arabic sentence-level annotation have used various techniques for different purposes. For example, Al-Subaihin et al. (2011) studied MSA Arabic using a game to build a lexicon for sentiment annotation. Their new method of annotation was to extract a pattern of the players using the crowd-sourcing concept but in a more entertaining and comprehensive manner by extracting the frequent sentiment annotations (the patterns) used by the players. The overall polarity was calculated after creating a lexicon from the game. Even though this method had more details than crowd-sourcing as the text went through multiple stages of annotation, it still had the usual crowd-sourcing drawbacks, which are finding available participants and the unpredictable time to finish the annotation. The challenges of annotating micro-blog text, which has words and phrases with no indication of sentiment, was discussed, as were the challenges faced during the system design. The challenges included the problem of informal online text, which has unpredictable phrases and words with no specific meaning or sentiment. The data used for that study have characteristics similar to those acquired in the current study, but the annotation showed the challenges of online data annotation, whether or not it includes metaphor. Duwairi and Qarqaz (2014) annotated social media Arabic text at the sentence-level using crowd-sourcing. Data were collected using a Twitter and Facebook crawler tool. A crawler is a crowd-sourcing annotation tool that is used to annotate the collected tweets. A supervised method was used on different machine learning algorithms to classify sentiment, which were NB, SVM, and KNN. SVM showed better performance in classifying sentiments than the other algorithms. Similarly to the previous study, the challenges in annotation of sentiment in online text were discussed, but the challenges were related to the writing style and not the level of ambiguity. For example, Arabizi is a type of online writing that changes Arabic text to English letters and numbers. Abdul-Mageed and Diab (2014) focused on building lexicon resources for Arabic, namely AWATIF Abdul-Mageed and Diab (2012) and SANA Abdul-Mageed and Diab (2014). However, each study collected a different type of Arabic, the former was MSA and the other combined MSA with dialectal Arabic. In SANA, they labeled the data set that was obtained from various resources, manually and automatically. In AWATIF Abdul-Mageed and Diab (2012), the authors discussed the challenges faced during the annotation process. Specifically, they showed the difficulties that annotators faced in labeling sentiment for a text that has a figurative part of speech, such as an idiom, metonymy, or metaphor. For example:

فبين الحين والحين
أجدني مكبلًا بطلاسمك الأبدية

'One time and another'
'I find myself shackled by your eternal spells'

This example explains one of the annotation challenges, which is metaphor. The example refers to classical (regular) metaphor which is found in poetry and literary works. This regular type of metaphor which makes it difficult in Arabic to at least find the meaning was challenging for the human annotators to identify the sentiment. Metaphor in social media text, however, has new and random means of communication. Designing a model to detect Arabic sentiment in such text is therefore not easy. Abdul-Mageed et al. (2014) built an Arabic language classifier for SA specifically for social media text and also built a lexicon of different Arabic dialects collected from different resources. This lexicon contained 3582 terms tagged for sentiment. The researchers discussed the challenges of Arabic text such as the rich morphology, dialects and the problems of identifying sentiment in the Arabic dialects on the three linguistic levels. They also discussed the challenges of finding the sentiment in short social media texts without mentioning significant features of the Arabic text in that specific context. Even though they did not discuss the type of the short text, that study did address the challenges of classifying sentiment for the Arabic language in social media. Social media text has more features than the dialectal and the regular linguistic properties. For example, they discussed the regular problem with identifying Arabic sentiment in dialects but did not mention dialects, the writing style or the invented terms in social media. Even though those types are not the main purpose of the current study, that discussion showed the difficulties which could be faced in identifying sentiment in Arabic used in social media. Rushdi-Saleh et al. (2011b) sought to solve the lack of the available Arabic data to perform SA by creating the Opinion Corpus for Arabic (OCA) from five hundred Arabic reviews collected from different online resources. The data were balanced as they considered half of the reviews to be positive and the other half negative. This consideration was not based on supervised annotation because they used machine-learning algorithms using Rapid Miner. Rapid Miner is a data science platform on which the performance of machine-learning algorithms can be validated. They stated that: "Arabic resources that focus on analyzing and mining opinions and sentiments are very difficult to find". So, the Arabic language still suffers from the lack of available resources for SA, which is one of the challenges which the current study is designed to address. Although some efforts have been made to create an Arabic SA corpus, the data available still lack the criteria and specifications of this research. So one of the first priorities was to create a reliable base dataset for Arabic metaphor in association with sentiment. Rushdi-Saleh et al. (2011a) sought to tackle the problem by conducting a study on an opinion corpus drawn from Arabic movie reviews and the web pages dedicated to them. In order to determine if there were issues which arose in translation, they translated the OCA Rushdi-Saleh et al. (2011b) into English using EVOCA and then compared the sentiment annotation results from OCA and EVOCA using machine-learning algorithms. They also compared the results with those of similar experiments conducted for English. In 2014, an analysis of Twitter for Arabic SA by Refaee and Rieser (2014) found that there had been enormous academic interest in this topic given the political unrest in the Middle East because governments and political analysts believe that examining Twitter feeds from this region can provide more insight into the overall mood of the average person in the region. To determine the legitimacy of this claim, Refaee and Rieser (2014) analysed a "newly collected data set of 8,868 gold-standard annotated Arabic twitter feeds. The corpus was manually labelled for subjectivity and sentiment analysis (SSA) (K= 0.816). In addition, the corpus was annotated with a variety of linguistically motivated feature-sets that have previously shown a positive impact on classification performance". Refaee and Rieser (2014) recruited a group of native Arabic speakers to compare the findings of the corpus with what they could themselves perceive in the selected Twitter feeds, and concluded that the corpus generated a high level of accuracy in SA. In addition, the study included a discussion of the online feeds structure as one of the challenges for annotation. They described the tweeters' writing type as free writing, which means that it is changeable. Also, the discussion included the ambiguity of semantic text such as sarcasm. Online semantic text has unclear polarity for sentiment annotation.

2.1.5 Sentiment analyzers

Sentiment analyzers are tools that classify and predict the polarity of text input. There are a significant number of web-based sentiment analyzers for English language, compared to Arabic language such as Chamlertwat et al. (2012) Thelwall et al. (2010). There are two available Web-based sentiment tools: one predicts the sentiment and the other classifies based on labeled data, named the lexicon-based sentiment tool. However, some papers consider the crowd-sourcing approach as a Web-based sentiment tool. Although the tool does not predict the sentiment, rather classify. The two mentioned have different approaches in sentiment classification and prediction. For example, the Mazajak sentiment analyzer Farha and Magdy (2019) built based on a deep learning approach to predict new text. The El-Masri et al. (2017) classifies the sentiment based on existing data with no prediction feature. In regards to Arabic, there are few to use to predict the polarity. Since this research started with testing the Arabic metaphor information in automatic tools for Arabic sentiment analysis Alsiyat and Piao (2020a), which is the essence of this research. So, this section discusses the available Arabic automatic sentiment analyzers. In our approach, the state-of-the-art Arabic sentiment analyzer is crucial to predict the sentiment using the Arabic metaphor corpus. In addition to the Arabic semantic tagger as a resource to predict the sentiment and potentially metaphor.

Mazajak tool Farha and Magdy (2019) is an online sentiment analyzer based on deep learning in Arabic. The system was tested using different data sets, such as SemEval 2017 and ASTD. The system was designed using state-of-the-art techniques to identify the polarity. Specifically, the algorithm is designed by preprocessing the AraVec Soliman et al. (2017), which is 67 million Arabic tweets. The model represents the sentences in multidimensional vectors using word embedding to process the data to the LSTM (Long Short Term Memory) and CNN (Netural Network). CNN is for feature extraction, and LSTM is for considering the contexts of each target. The pattern passes through the softmax layer to find the probability of the sentiment. Means, the pattern is the words with similar contexts have similar polarity. The model was linked to an online API with different ways of predicting the sentiment. The first way is a dialog box as text input to predict the sentiment with the option to train the sentiment analyzer for the new text. The second way is through a dialog box to type the tweeter account name and retrieve the first three hundred sentiment tweets. The last way is to upload the tweets as a text file, and the system returns the tweets with the sentiments for each tweet in Excel file. Mazajak shows good performance with the testing data sets, which are ASTD and SemEval 2017. However, the system is designed to learn the new words based on the crowd-sourcing concept, which could be a solution to find the sentiment for the metaphor. But this solution is not effective and reliable, as the system needs large metaphorical data to teach the system to predict the sentiment with metaphor. However, there are no such big data for Arabic metaphor available to use in such case. In addition, the system is not designed to detect the metaphor before the sentiment, as the metaphor drives the sentence polarity.

Another multilingual sentiment analyzer including Arabic is sentiStrength Thelwall et al. (2010). SentiStrength predicts sentiment by the degree of negativity or positivity of the text in multiple languages. However, the sentiment analyzer does not predict the neutral polarity. For example, if the Arabic text is positive and negative minus one, means the text is not positive and not negative, which is neutral. So, the system does not classify the text as neutral. It is designed based on machine learning algorithms to classify sentiment. Compared to state-of-the-art techniques such as CNN and deep learning, machine learning does not have the learning and adaptation feature of new data. That means that machine learning needs supervised data, while the other techniques could accept learning from raw data. In addition, this sentiment analyzer made to detect the emotions strength. They explained that the obvious emotion words such as love could come as negative based on the context. They present an example, 'Love will tear us apart', which holds a level of ambiguity. precisely, the example has metaphor. So, the system could not detect the sentiment for such an example. They show evidence for the English language, which is the main language used in the system. So, the evidence is suppose to be applicable for other languages in the system as well.

The rest of the Web-base Arabic sentiment tools built on the basis of the concept of annotation. For example, El-Masri et al. (2017) listed some of the Arabic web base tool for sentiment. However, those tools created for crowd-sourcing for sentiment annotation such as Al-Subaihin et al. (2011) not predicting. Also, in El-Masri et al. (2017) you can find profit-oriented Arabic sentiment analyzers. However, the sentiment classification for commercial Arabic sentiment analyzers is unknown. The El-Masri et al. (2017) is a web-based tool for lexicon-based sentiment. The tool used one hundred fifty two thousand four hundred fifty five labeled tweets for training. However, the tool was not tested because the experiment focused on building the lexicon. In addition, the tool is still not available for use for further modifications. Means the tool did not predict the sentiment for new data. Therefore, the tool might not be classified as an Arabic sentiment analyzer. El-Masri et al. (2017) has a user interface with multiple features. The features designed as dialog boxes to insert the data and choose the range of tweet dates, pre-processing types, the sentiment classification method, and the machine learning classifier. However, as mentioned, the tool is not available for testing and evaluation.

The previous online Arabic systems were scrutinized to be aware of the Arabic sentiment analyzers. In addition, examine the best resource fit for our purpose. SentiStrength Thelwall et al. (2010) does not have the criteria to use for Arabic metaphor corpus. The SentiStrength measure the sentiment degree using sentiment score for emotional text. However, the builder for SentiStrenght explains in an example the inability of the system to detect the sentiment for the figurative aspect of the text, even if it has the same emotional word. The example explained previously has ‘love’ in a figurative context. We aim to use the state of art Arabic system that is based on advanced techniques. New techniques such as CNN, deep learning, and LSTM may have the ability to know the Arabic metaphor. However, these techniques are not designed to identify the metaphor. For example, Mazajak builds based on one of the advanced techniques, which is deep learning. In addition, the system has a correction section to allow users to add the right sentiment annotation. Even tough, the system could not predict the correct sentiment for the Arabic metaphor. The Arabic Web base created by El-Masri et al. (2017) is not available for use. So, because of the lack of the available Arabic sentiment analyzers. Designing a sentiment tool is crucial. As discussed, there is a lack of available Arabic sentiment analyzers. The state of the art Arabic semantic tool El-Haj et al. (2022) was used with the designed sentiment tool to classify the sentiment. Even though the semantic tools have no features to classify the metaphor or sentiment, the tagger still has potential to detect the metaphor based on the semantic tags.

2.1.6 Discussion

In this chapter, the resources and concepts were discussed and explained to explore metaphor and sentiment for the Arabic language. The chapter discusses some studies of the three overlapping subjects of this investigation. The subjects are Arabic metaphor, English metaphor, and sentiment analysis. In addition, the studies discuss the gaps for those subjects as Arabic metaphor and sentiment that is still in progress compared to English. In addition, the chapter discusses some studies for Arabic metaphor in social science and linguistic fields. In addition to the conceptual analysis, that addresses the metaphor as one of the NLP problem to solve. Because there are no studies to build an Arabic metaphor corpus and sentiment detection for Arabic language.

The previous literature demonstrate the gaps and challenges to build and investigate the Arabic online metaphor with sentiment. The challenges appear in the foundations of this research, the data, and the methodology. The lack of reliable and available Arabic metaphor resources is one of the biggest challenges in this research. As mentioned above, previous studies available for Arabic metaphor in different fields do not meet the goals set for this research. So, the Arabic metaphor corpus with sentiment information that includes meaning and context was formulated. Also, since this is the first study to investigate the Arabic metaphor with sentiment, there is no previous methodology or algorithms to follow to detect the metaphor. Even if this research uses the English metaphor methodology, the English language has supportive algorithms and resources that are used to detect the metaphor. As mentioned above, Rentoumi (2012) uses WordNetMiller (1995) and SentiFig, as well as the WDS algorithm. In addition, Wilks et al. (2013) used VerbNet for automatic English metaphor identification. So, The automatic identification for English metaphor based on reliable lexicons. Not to mention the English metaphor resources built for the English metaphor such as Krennmayr and Steen (2017), which are used for training and testing. In regards to the methodology followed to identify the English metaphor despite sentiment, they conducted based on strong and reliable foundations. As discussed in this literature, the approaches linked multiple resources and techniques to detect the metaphor. For example, Do Dinh and Gurevych (2016) used a neural network with word embedding linked with VU Amsterdam Krennmayr and Steen (2017) to identify the metaphor on the word level. So, Krennmayr and Steen (2017) facilitates the search for the metaphor at the word level, as the data were annotated at the word level for each sentence in XML tags. Other studies discuss the gaps of the different advanced techniques for identifying English metaphor only. Arabic language studies discuss only conceptual challenges for metaphor detection with no actual application. For example, the Rentoumi (2012) research literature discusses the gaps in previous approaches to applying the English metaphor with sentiment.

In regard to the sentiment for the Arabic language despite the metaphor, the sentiment for the Arabic language still lacks resources compared to English. Not to mention the challenges for Arabic language challenges in an online context. The challenges are the distinctive Arabic dialects, the complex morphology, and the writing styles. In terms of metaphor and sentiment, there is no study that combines sentiment, metaphor, meaning, and context for the Arabic and English language. So, this chapter highlighted the gaps for Arabic sentiment analysis in general. In addition to the gaps, this research is subjected to metaphor and sentiment analysis. The chapter discusses the problems of identifying the sentiment of a metaphor in online text similar to social media texts. The discussion illustrated as the lack of the available resources and the ambiguity of the online text hold in meaning and sentiment, respectively. Also, the sentiment annotation challenges for regular and metaphor text. In addition, the discussion shows the novelty of the Arabic metaphor corpus.

Studied on ABSA for Arabic metaphor			
The Researcher	The Research	the Scheme/Approach	Arabic type
(Al-Smadi et al., 2015)	Human annotation of Book reviews	SemEval-2014 (XML)	MSA (HAAD dataset).
(AL-Smadi et al., 2015)	News of Israel-Gaza conflict in 2014 from Social media	SemEval-2014 (XML)	MSA.
(Obaidat et al., 2015)	Enhancing the Determination of Aspect Categories and their Polarities in Arabic Reviews using lexicon-based	the news data set following the Gaza attack in 2014	classical
(Pontiki et al., 2016)	SemEval-2016 Task 5: Aspect-Based Sentiment Analysis (multilingual including Arabic)	SemEval16-task 5 (Arabic hotel review) In-house tool for annotation and Multilingual	MSA Arabic
(Ruder et al., 2016)	nsight-1 at semeval-2016 task 5: Deep learning for multilingual aspect-based sentiment analysis	Deep Learning (CNN)	SemEval16- Task 5 data set.
(Al-Ayyoub et al., 2017) ABSA on laptop review	SemEval16- task 5	SVM classifier	dialectal.
(Al-Smadi, Talafha, et al., 2018)	Using long short-term memory deep neural networks for aspect-based SA of Arabic reviews	short-term memory and deep neural networks	SemEval16- Task 5.
(Al-Smadi, Qawasmeh, et al., 2018)	deep Recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' review NNR vs. SVM for ABSA of Arabic hotel reviews	NNR+SVM	Dialectal (Arabic hotel reviews).

Table 2.1: ABSA studies for Arabic Language

Chapter 3

Data Collection

This chapter describes the process of collecting the Arabic online metaphor data. The frequent observed behavior of the Arabic metaphor in the online context will be discussed, as will the analysis to define the scheme for corpus building including data collection. The novelty and details of the data collection process will be explained in the following sections. The data analysis presented in this chapter is essential for building an accurate annotation scheme and for developing an accurate algorithm for identifying Arabic online metaphors in the future. It also shows the observed changes in the Arabic online metaphor. The frequent occurrences of the Arabic online metaphor specified the features that could be used for the computational aspect of the analysis. The discussion also shows some of the linguistic structures of Arabic online metaphors. Although data collection could be included in the same chapter as corpus building, the analysis of data collection and the discussion that follows is sufficiently important to be discussed in a separate chapter.

The regular structure of the Arabic metaphor is opposite to that of a simile Alsiyat and Piao (2020a). More precisely, the Arabic implicit metaphor contains a logical object with an illogical trait. For instance, explaining the bravery of someone, a logical object, using an illogical trait of another absent object. For example, الجندي يزئر في المعركة means *'The soldier roared in the war'*. In the sentence, the solider (a logical object) is said to roar, which is an illogical trait because roaring is a lion trait. However, the sentence used the lion trait, which is an absent object, to show the courage and strength of the soldier. In another example, the explicit metaphor structure is the usage of an illogical subject within a logical

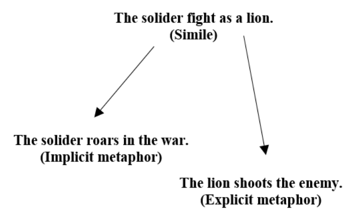


Figure 3.1: Arabic metaphor Definition

context: الاسد يصيب العدو means *'The lion shoots the enemy'*. These examples show the two primary types of Arabic metaphor, implicit and explicit, although there are more types of Arabic metaphor such as dead, cliché, stock, adapted, recent, and original (Zeroual and Lakhouaja, 2018).

The Arabic metaphor in the online context has a spontaneous nature; it is brief and impromptu to adapt to the speed of online communication. The informal Arabic online metaphor is not only different from the regular structure, it has a variety of Arabic forms, which make it a new Arabic online metaphor. For example, the Egyptian dialect uses new terms such as مالوش حل, which means *'has no solution'*. Another example, مش طبيعي means *'not normal'*, which is used to express a positive attitude towards a book. These two terms are prefixed with negating words in the form of a phrase. The phrases in the new data set that follow this negation provide an indication of positivity.

Metaphors in the online context occur in different categories for different purposes. For example, the metaphor بيض that means *'Eggs'* is used sarcastically as a food term to express an opinion; the term indicates a bad atmosphere. Because an egg has a rancid smell, the term is given a negative polarity to indicate the bad smell of the egg. From the previous discussion, we can group the Arabic metaphor terms into categories. This chapter discusses the source of the Arabic metaphor corpus and shows the pattern of Arabic online metaphors. In addition, the discussion shows the similarity of the Arabic online metaphor with the context of social networks.

3.1 Why Arabic online metaphor?

In this section, terms used online are discussed to clarify the type of Arabic metaphors collected and the meaning of online terms used to describe the structure and similarity of other Arabic metaphors that can occur. Online metaphors are now widely used after the wide spread of Arabic data; they are specifically used in social networks, which is the form most often used to express opinions, particularly in light of the large amount of data exchanged on social networks. Also, the Arabic online metaphor corpus is new and unique as it contains terms which have no reliable source for their meaning. For practical purposes such as machine learning applications, the most desired data used are similar to the social networks, which are the data used often online. The regular Arabic metaphor could have been used as a source for this research, but the data would not have any novelty, but would simply mean building a regular resource without any new analysis. For example, the Arabic metaphor in poetry has a fixed structure because of the classical Arabic which is used, which is the ancient structure of the Arabic metaphor. Another example is the Arabic metaphor in the Qur'an, a context which is miraculous but fixed in its structure.

The metaphors used in this research were used on an online book review website, the Good Readers website. So, the data collected has similar characteristics to the social media feeds. However, we use online terms to generalize the type of data. In addition, online terms are used to describe any data used on the World Wide Web regardless of the place or platform of the data used. The metaphors were collected using specific criteria to distinguish the differences and the changes between the regular formal Arabic metaphor, the online Arabic metaphor, and the formal Arabic metaphor in the online context. Although we were seeking to collect new Arabic metaphors online, we were restricted to the LABR data set (Aly and Atiya, 2013). For example, we wanted five particular dialectal Arabic metaphors such as بيض/*Eggs*, but we only found two examples of these terms in the LABR data. In addition, the criteria used for data collection targeted new Arabic metaphor terms on-line, which were similar to the terms used on social media platforms. Collecting and selecting the new Arabic metaphor terms was one of the guidelines for data collection. The terms in the online version were carefully collected to fit the criteria. The new corpus derived from LABR contains many online Arabic metaphors which demonstrated distinct features from traditional or formal metaphor usage contexts. The distinctive features for the informal and formal metaphors were represented by the structure and were new terms without any reliable source for their meaning.

3.2 Why LABR?

There are sufficient datasets available for Arabic sentiment analysis publicly available. The largest-scale Arabic sentiment datasets were targeted for data collection, and Arabic sentiment datasets were also targeted because sentiment is an extension of the computational aspect of this research. We therefore chose the sentiment in the large Arabic datasets as a main goal. The Arabic sentiment data set, which is annotated at the sentence level, was therefore used to compare the sentiment efficiency with the Arabic metaphors. The features, the discussion, and the annotation were therefore performed based on the computational aspect of this research. The linguistic analysis of the dataset in the discussion shows the significance of this research.

The two largest available Arabic sentiment data sets have limitations which had to be considered. BADR Elnagar and Einea (2016) and HARD Elnagar et al. (2018) are large datasets of Arabic sentiment. BADR has 500,000 Arabic book reviews annotated for sentiment at the sentence level using a machine learning algorithm. The data set tested for the annotation shows the Arabic language challenges, whereas HARD is largely written in the Gulf dialect of Arabic. Both data sets were suitable for this study despite being written in specific types of Arabic for different domains. For example, most of the Arabic dialect written in HARD is the Gulf dialect, whereas BADR has MSA and mostly dialectal Arabic.

LABR Aly and Atiya (2013) is the third largest sentiment analysis data set for Arabic (see Table 2). It contains 63,000 book reviews written in colloquial and MSA Arabic and is annotated for sentiment at the sentence level. The diversity of the written Arabic forms (dialectal and MSA) in the LABR dataset would widen the analysis in this study because of the many different dialects in it, such as Egyptian, Saudi and Tunisian.

In addition, the other available datasets were written solely in MSA and other types of Arabic were not available, such as Maamouri et al. (2004) and Abdul-Mageed and Diab (2012). In addition, the other available sentiment dataset has fewer sentences than LARB, which decreases the probability of having an appropriate number of metaphor sentences. In this step, metaphorical expressions were collected from publicly available data in the LABR. LABR was used to demonstrate the impact of metaphorical expressions on the prediction of the sentiment analyzer for LREC2020 Alsiyat and Piao (2020a).

- We used sentiment data sets to collect metaphorical data, even though there are Arabic metaphorical data sets.

- Mention the available Arabic metaphor datasets. (studies conducted with limitations and differences)

3.3 Online context of metaphors in LABR

LABR stands for Large-scale Book Reviews for an Arabic dataset. Its source is a social networking website where readers list their opinions about books, so the entries have an online writing style. Although the website is not considered a social media platform, which has more free writing and communication, the data nevertheless have the social media communication style. Some examples from the new corpus of social media writing features are typically brief and opinionated. The metaphors extracted from LABR therefore have the required social media writing style. One of the examples is تحفة, for which the hidden meaning is *'amazing'* and the literal meaning is a 'vase'. This is a one-word expression that is used to express the positivity of an item. It is also challenging to identify, even when it is supported by the context, because such expressions are opinionated and expressive, so the supporting context gives an emphasis to the opinion without necessarily explaining the word. Defining those expressions in a resource is therefore essential to detecting the metaphor computationally. Although the LABR contains informal metaphors, it also has formal writing of metaphors in the opinions of the books. Formal writing has the full structure of the Arabic metaphor usually written in MSA. An example of the formal metaphor context is يجعلك تلتهم الكتاب في وقت قصير which means *'makes you devour the book in a short time'*. The sentence has the full structure of Arabic metaphor, and 'devour' is a metaphoric expression that indicates reading the book with excitement.

3.4 The data collection process

The criteria were established to meet the aim of the study and to be extended for the next work. Metaphorical reviews were carefully chosen from LABR. The principal criterion was to pick metaphor expressions that have polarity. For example, we ignored sentences that included a metaphor as a quotation from a book. Also, from the long reviews, we picked the part that combined metaphor and sentiment. For example,

عجبنى أوى الجزء الأخير اللى بيقول فيه
واستقر الشاطئ وسمعت طقطقة
مكنة الماء، واحسست ببرودة الماء في جسمي

'I like the last part the most, which said:
and the beach settled and I heard the water
machine whirring and I felt the coldness in my body'

In this review, the reviewer gives a quotation from the book that contains a metaphor. But the metaphor part of the review does not have sentiment or give an opinion about the book. Reviews containing facts, however, which are considered as a natural polarity, were included in the dataset. We also ignored sentences that had two contradictory polarities with two different genres of a book, but we did include aspects of the sentences which contained metaphor. We wanted to have diversification in collecting the same metaphor terms with differences in the types of metaphor, different literal meanings, genres, contexts, and different polarities.

The sentence رائعة.وصادمة لكل الدافنين رؤوسهم في بحار الرمال, which means *'amazing. and shocking for all who buried their head in the seas' sands"*, is an example of a case that has two different opinions on a book with a fact at the end of the sentence that includes a metaphor. We ignored such cases, as the metaphor does not have sentiments that contribute to the main opinion of the novel. In addition, we include metaphors that provide an emphasis of the main opinion. For example, in the sentence تحفة. بحر ممزوج من الخيال, which means *'amazing. a sea mixed with imagination'*, the main opinion is represented by تحفه, which literally means *'Vase'*, but the writer followed his opinion with another metaphoric expression for the purpose of hyperbole. Another condition, We avoided single-expression reviews which contained a metaphor that we had already collected. For example, we collected بيض, which means *'Eggs'*, as an Arabic contemporary metaphorical expression in a different context, but we ignored it if it occurred without context. For the metaphor terms collected, we used a reliable dictionary to validate the meaning to ensure the validity of the metaphors collected. For example, the annotator chose شائكه, which means *'suspicious'* not *'barbed'*.

We assigned native Arabic speakers to collect the data after providing them with guidelines and verbal instructions. The data set was reviewed in multiple rounds with the annotators to eliminate redundancy. We also discussed the agreement of the selected metaphorical expressions in some sentences.

3.4.1 The Data collection criteria

The Two Arabic native speakers were asked to search through the LABR for the most popular Arabic metaphor terms used in social media. So, they manually searched using the 'Find' feature in the LABR data text file. In addition, the LABR was segmented and distributed among Arabic native speakers in order to avoid repeating the same choices. In addition, they were asked to select formal metaphors.

We intended to build a balanced dataset, but we were restricted to the dataset and the availability of metaphorical expressions. We collected 1000 reviews and found that 500 of them had two hundred different metaphorical expressions. These two hundred were divided into two groups each of one hundred formal and informal metaphorical terms. The collected terms were categorized and organized into different types based on their use, the part of speech, and the literal meaning. The informal metaphorical terms, which are not common in formal writing or in a conventional dictionary, were classified based on their use. Formal metaphors were categorized on the part of speech because the formal metaphor terms are known in formal writing, but the context defines the metaphor of those terms. As the domain of the dataset was books, most of the terms were assumed to be relevant to a book, novel, or story. So, the domain defined the illogical usage of those terms.

The informal metaphors were grouped under food, weapons, verbs, adjectives, compositions, medical, illness, drugs, personification, and offensive categories. Similarly, formal metaphors were grouped under categories of verbs, proverbs, nouns, and adjectives. For example, in informal groups, سلاح وقنبلة means *'weapon/bomb'* which was categorized in weapons, which included any term used to refer to combat. The food group, which includes the terms egg, cream, delicious, honey, sugar, scrumptious , flavor, freshness, meal, onion, taste, appetizers, spices and fatty.

Each term had to have a minimum of five different sentences, meanings, or genres. Although some terms were found only in a maximum of two sentences in LABR, we nevertheless included them in the dataset. As has already been explained, the data collection process was assigned to Arabic native speakers and the guidelines were explained to them verbally and in writing:

- Collect the Arabic dialectal metaphor used often in social media, such as بيض meaning *'Eggs'*.

- Collect known metaphors in formal writing.
- As much as possible, avoid long sentences and avoid sentences that contain metaphors in quotations.
- Collect sentences that combine metaphor with sentiment in the main opinion about the book.
- As much as possible, collect five different sentences of the same term to have a balanced dataset.
- Avoid redundancy.
- Collect the sentences and highlight the metaphor in each one.

We were restricted to the number of metaphors available in LABR. In addition, the data were revised and gathered to meet the criteria.

- We used the dictionary to determine the meaning of the metaphors (for example, if there was a word that seems to be a metaphor, we checked the familiarity of the word in a metaphorical context using a reliable dictionary (for example, the *al-ma'ānī* Arabic online dictionary). In addition, Standard Arabic metaphorical terms were verified using an Arabic online dictionary, but only to identify their literal meanings, not their metaphorical ones. The metaphorical meanings were instead validated by annotators and questionnaire participants. Some terms appear in Modern Standard Arabic (MSA) rather than dialectal Arabic, yet their online usage differs from their standard meanings. For example, كحولية *'koholiyyah'* literally means *'alcoholic'* in MSA and metaphorically *'nostalgic'*, which has a different meaning than recognized. To ensure that such terms were not inherently metaphorical, we conducted additional checks, though only for a few cases. Our findings revealed that these terms had no metaphorical meaning in the dictionary. This indicates a novel way of expressing metaphor in online contexts, where MSA is used instead of dialectal Arabic.
- Avoid collecting reviews with examples from the book that do not convey any opinion or sentiment.

3.5 The Arabic metaphor data description

In this section, the dataset is described to show the differences that could affect the practical application of identifying online Arabic metaphor. The dataset contained thousand Arabic metaphor sentences. It was intended to be balanced, but we were restricted to the availability of the targeted metaphors. The dataset was divided into five hundred informal metaphors and five hundred formal metaphors. Five sentences were selected for each term. Although there were sentences with terms that appeared only once in LABR, those terms were chosen as new Arabic terms in the online context, which is similar to social media.

The dataset contained some metaphors that have the same meaning but with a different morphology. For example, أخطر و خطير means *'the most dangerous' and 'dangerous'* and sentences containing them are من اخطر السير الذاتية and رواية خطيرة. Both examples mean *'one of the most dangerous novel'* and *'a dangerous novel'*. In addition, the same Arabic metaphor word can have different meanings. For example, the term الثمالة means *'until intoxicated'*, which is interpreted as *'lost'* in the following sentences:

رواية رمزية مبهمة من الألف إلى الياء.
لم يرق لي الترميز المعقد و اللذي وصل حد الثمالة

'A symbolic vague novel from A to Z'
'I do not like the complicated symbolic that arrives to the intoxication'

Another example of the same term with different meanings is الثمالة literally means *'intoxication'* in another sentence, which means *'joy'*:

مُبكية. ومؤلمة. وراااائعة حدّ الثمالة.!

'crying. hurtful. and amazing until intoxication'

As stated above, the dataset was intended to be balanced, so we looked for the same metaphor term in different sentences. However, the way in which the term is written affects the result. For example, the term قشطه means *'cream'*, but can be written in different ways. Some terms written in dialects can have different suffixes to refer to the femininity and masculinity of the described target. As an example, قشطه is written as اشطططططططططططططط and بوشين, which means *'two faced'* in dialectal pronunciation with elongation to emphasize the opinion, قشطة with a different suffix. In addition, writing in Arabic metaphor terms جامدة ، جامده means *'solid'* with the feminine suffix with slight differences in writing ةه. The جامد و جامده does the same.

Some sentences contained more than one metaphor, so we chose the ones based on our criteria and the ones which would balance the dataset. For example, تحفه means *'masterpiece'*, and it appeared multiple times in different sentences, and many of the sentences collected contained the same term. The online Arabic metaphor structure is changeable, which means that it does not follow the usual Arabic metaphor structure. In addition, the structure (as well as the terms) is constantly updating and changing based on the nature of the online communication. For example, كحوليه means *'alcoholic'*. The dataset is a showcase of Arabic online metaphors, using new dialectal terms in different sentences for the online context. The term كحوليه, which means *'alcoholic'*, is considered a new Arabic metaphor term even though it appears in MSA because of the unusual use of the term in this structure in this context. In addition, كحوليه is used in Arabic to describe a drink, but never as a metaphor for a story.

The same metaphorical terms were collected in different sentences with different meanings or genres. For example, the literal meaning can be annotated differently in different sentences for the same term. For example, تحفه metaphorically means *'amazing'* in the sentence:

كتاااب تحفه جداا بيحكي مشاعر وافكار
طبقه معينه من الناس بطريقة عفويه اووي

'very amazing book it talks about feelings and
thoughts of certain level of people in very spontaneous manner'

Whereas it is used to mean *'amazing'* in the sentence. The data were described in terms of morphology, meaning, and structure that could affect the practical application of the identification of the Arabic metaphor. Although data were collected for Arabic online metaphors, which is similar to the social media context, the practical application still made it challenging to identify the online Arabic metaphor.

3.6 The Arabic metaphor in the online context

The difference in terms of metaphors in the online context and in the regular context can make it difficult to gather all the linguistic aspects to evaluate it, because regular Arabic metaphors occur in multiple types, as mentioned above, with different structures, and the informal metaphor has a dynamic structure.

This section discusses the difference in the new Arabic online metaphor corpus in terms of meaning, structure, and context. The differences in informal metaphor writing affected the data collection of Arabic metaphor terms. As discussed above, the data collection process was carried out using guidelines that explained how to select the new Arabic metaphor terms that are frequently used on social media. However, the terms were affected by the method of writing used on social media, which involves elongation and typos. In addition, informal writing can be dialectal, not following the grammatical rules of the Arabic language. For example, the difference in writing the term جامدة and جامده, which means *'solid'* is different, in which the letters ه and ة have different grammatical rules. This had an effect on the process of collecting the LABR reviews.

As mentioned in the previous section, the criteria prioritized sentiment. The new dataset showed a variety of sentiments for the same metaphorical expressions, despite the fact that they had a different morphology. For example, فظيعه means *'terrible'* but the term can be negative or positive in different contexts. The term is used metaphorically as a positive indication of the distinction of a work, but is used as a negative indication to emphasize a negative opinion. For example, in the sentence كئيبة بطريقة فظيعة, which means *'depressing in horrible way'* the term is used as negative.

We observed patterns in the context dataset to find features to identify metaphor and sentiment, which also enabled us to notice linguistic patterns, and we found that some categories had specific patterns. For example, in the category of *'illness'*, some terms were prefixed with particular verbs that could also be considered metaphors. In the example يجيبلك / تجيبلك means *'will brings you/ brings you'*, one of the verbs occurs several times in different terms in the illness category. The other terms in the same category were preceded by regular verbs تجيبلك غثيان اكتئاب، صرع ، تخلف عقلي literally mean *'It gives nausea, mental disability, epilepsy, depression'*

In addition, there were some informal terms that always contained particular words, which can both be known as Arabic metaphors in an online context and could occur in phrases. And those phrases were categorized as composition metaphors in the category we made for the metaphor terms (see Table 4.2). Some of those terms were categorized as composition metaphors, and others as personifications. The terms were categorized according to the type of main words, which were considered metaphors. Because this type of metaphor comes as two words together. For example, سهلة الهضم ، غسيل دماغي ، خبط

لزق. means *'easy to digest, brain washing, hit and past'* and حتى النخاع و حد الثمالة means *'till the marrow'* and *'till intoxication'*. The example مجنونة was used to describe the superiority of a book, while مجنونة, which means *'crazy'* as a metaphor, is rarely used in formal writing or contexts.

1. The dialectal metaphor written in social media and in the regular context (dialectal pronunciation) can elongate (repeated) letters to emphasize the word. For example, قشطه اشطططططططططططططا ، و بوشين which means *'has two faces, and cream in different forms of writing'*, but is written in dialectal pronunciation.

2. Online metaphors use irregular terms to express confusion, love, sensation, and satisfaction; they use *'illness'*, *'weapons'*, *'drugs'* and offensive terms, whereas formal metaphors have recognizable terms that show their use in regular contexts and give the full structure. For example, بعدَ الإنتهاءِ من هذه الرواية شعرتُ (بدوخةٍ) جمالٍ كبيرة and it means *'after I finished from this novel I felt overwhelming dizziness of beauty'*

3. Composite metaphor terms comprise a combination of two terms to produce a single meaning. For example, دمه خفيف، شهاجتي مجروحه، سهلة الهضم.*'light blood, my testimony is wounded, easy to digest'*

4. There are differences in meaning between the same terms in metaphors written in MSA from those in dialects. For example, مخيف means *'scary'*,and could be interpreted as a regular adjective in a normal context, but in the online context it could be interpreted as a metaphor if it is accompanied by another positive adjective. ليست خيال ولا حقيقه ترتيب الاحداث فيها مخيف *'It is neither fantasy nor reality; the sequence of events is terrifying'* كتاب خرااااافي بكل ما تحمله الكلمة من معنى حقيقي لدرجة الخيال مخيف لدرجة الضحك *'A phenomenal book in every sense of the word, so real that it feels like fantasy, so terrifying that it becomes laughable'*

The dataset contained verbal discourse to express metaphor, which is similar to the study by Raii (2009). That study showed a time category for metaphor in speech with verbal discourse in Arabic dialectal forms. For example, the study has a metaphor to express time such as اجا تشرين, which means *'November comes'*. The metaphor occurs as a verb in the dialectical discourse represented in the word اجا which means *'comes'*. The dataset has يجعلك which is *'makes you'* in verb form to express the metaphor informally, although the metaphor in the data has the formal form. This means that the Arabic metaphor could be expressed informally as a metaphor but written in formal writing, which is MSA.

3.7 Informal metaphor

Arabic online informal metaphors have no source to explain the actual meaning because those terms are invented to fit the online environment. This section explains the literal and hidden meanings of informal Arabic metaphors and their context. The dataset contains new dialectal terms that have no source in the official Arabic dictionary. The meaning and dialect of some of these terms were authenticated using a questionnaire.

Metaphors in the online context use voice semantics to describe the writer's opinion. For example, فسسسسس and يخخخخخ (*'Yikhkhkhkh'* and *'fsss'*) are notions expressing *'disgust'* and *'disappointment'*.Although the informal Arabic metaphor has a known structure following the pattern of the data, it is still presented in the online writing style. For example, the structure of informal metaphors online does not always follow the main structure for regular metaphors. An example is جنتنى حرااااااااااااااااام عليك which means *'you drive me crazy shame on you'*. The hidden meaning of the informal term الش is *'lying'*, but the dictionary definition of the literal means a city in Spain, though it is written with a different pronunciation. So, the term has no source in Arabic that can define the correct hidden meaning.

بطيخ means *'watermelon'*, which is a fruit of pumpkin species that occurs in many types, usually spherical or ovoid in shape with a thick green or yellow skin. However, the actual usage of 'watermelon' is a notion of 'meaningless talk', which has no reliable source to explain the hidden meaning. Also, the term بطيخ/*'watermelon'* in this form often comes as a single term after expressing the main opinion. However, the same term in a different form occurs as an adjective احلامها البطيخيه and means *'her watermelon-y dream'*. The term is translated as *'watermelony'* (an adjective), an example of the Arabic online metaphor form, while the regular part of speech for 'watermelon' is a noun. Also, the translation of the Arabic informal term in the sentence ثم لّخمتني رحاب بسام في عالمها فأحببت أحلامها

البطيخية وذكرياتها ومشاكلها وحكاياتها means *'Then, Rehab Bassam thrust me into the chaos, and I grew fond of her watermelon-y dreams, her memories, her problems, and her stories'* is not as precise as the translation offered by Google translation, which is *'her watermelon-y dreams'*. Therefore, the Arabic online metaphor needs a reliable source that can clarify the hidden meaning of the metaphor in translation.

From the previous discussion, it can be seen that authentication was crucial in data collection. In addition, interpreting the literal meaning is fundamental in our work. The literal meaning of each term is given in an online dictionary. The literal meanings of informal metaphors fall into two types: those which have as source and those which do not, based on the authentication. In addition, there are words that have voice semantics, which means using the voice in writing to express opinion.

3.7.1 Terms with no source

In this section, I shall discuss the meanings of the informal Arabic metaphor and the approach to specifying the meaning. During the data collection process, the chosen metaphor sentences were examined, and this process led to the realization that there is no source for some Arabic online metaphor terms. As mentioned above, there is no reliable source for online Arabic metaphors with hidden meaning. In this section, the lack of a source is for the literal meaning of some Arabic online metaphor terms, and although there is a literal source for some Arabic metaphorical terms, they do not apply to phrases which mean something different from the original. For example, مالوش حل means that *'has no solution'*, which has no literal meaning in the Arabic dictionary. This term was written in an Arabic dialect, which means it is new in online communication. Another example of a term that has an original source in Arabic but has no related meaning is خبط لزق, which means *'knock'* and 'past', two separate terms that have a source in the Arabic dictionary but not as a phrase. However, some of the mentioned terms do not have a source in the Arabic dictionary that indicates the same meaning, such as the word الألش, for which the literal meaning is the source of the term, not the hidden meaning. The previous example of *'watermelon'* بطيخ is such a term; it refers to a fruit but يطيخ has a hidden meaning of *'nonsense'* or *'inaccurate'*. So, the hidden meaning is the main point behind the term.

The absence of a source discussed in this section refers to different cases of Arabic metaphor terms in the dataset. First, terms used as metaphors that are not derived from the literal meaning, and second, when the term used as a metaphor is opposite to the literal meaning. Third, the metaphor occurs as a semantic voice, and fourth, the term is present in the Arabic dictionary, even with the illogical structure of the metaphorical sentence. The online metaphor in the online context has not just a new structure; it has new terms invented despite the fact that they have a source in the Arabic dictionary with a different meaning. Most of the Arabic online metaphors have no reliable source in terms of meaning.

Some terms, however, have no source for their correct meaning, and the metaphorical meaning is not derived from the literal meaning. New informal terms written in a dialect are considered new and without a source. To clarify, in the English metaphorical sentence 'you are a rock', the rock is an object indicating strength, whereas the literal meaning is a solid mineral material forming part of the surface of the earth. So, the rock as an example of strength is used to describe a person's trait, which is strength, but the literal and metaphorical meanings are not related in the term هايف in Arabic, which means *'silly'* based on the annotation of the annotators.

Also, a term could be represented by a known term in Arabic but comes with a new structure, an example of which was discussed above: كحوليه, means *'alcoholic'* which is in the Arabic dictionary but comes with a new structure to mimic online writing. An example of a term without source is هايف, which in a literal context means a *'tall women'* as a good trait, but used metaphorically it means a *'silly person'*. In addition, the word الروشنه means *'a man who has lost his mind'* in the literal sense, while metaphorically it conveys the notion of *'coolness'*. Our approach in collecting the informal metaphor was, therefore, to check the source of each word. Some unfamiliar terms appeared as metaphors, while the dictionary shows their familiar meaning in the context. For example, the term صارخ which means *'obvious'* and not *'screaming'*,as in the sentence.

وعاد بنا الكاتب بقوة الى مسرى ومجريات الأحداث .
مريم وليى مثال صارخ لما عانته المرأة الأفغانية من عنف داخلي خانق.

'The writer brought us back to the course and unfolding of the events'
'Mariam and Laila are a striking example of
the internal suffocating violence endured by Afghan women.'

مثال صارخ means literally *'screaming example'* as the literal meaning in this sentence, although the illogical explanation of the word مثال *'example'* is used in the example. Thus, the familiar word صارخ in the dictionary is literally used. Also, the semantic voice terms in online Arabic have neither a literal nor a metaphorical source. Semantic voice terms are written to express an opinion rather than being spoken. As mentioned above, فسسس *'fsss'*, is an expression used verbally to describe *'disappointment'*, but it is used in writing in an online context. Such semantic voice terms have no literal source from which to derive a rather intuitive understanding of them. The semantic voice terms have to be authenticated to know the meaning and dialect.

In addition, online Arabic metaphors can appear as slang. Some of these phrases could be interpreted as metaphors from the context. For example, in غسيل مخ/*'brain wash'*, the word غسيل means *'wash'* is a regular verb followed by the illogical noun برين. This concept is similar to the English verb violation in Wilks (1978). The data have cases in which a phrase is interpreted as a combined metaphor comprising two words. For example, the slang term خبط لزق/*'knock past'* occurs as a phrase that combines two verbs, but if computationally applicable, the phrases have to be separated to identify the Arabic metaphor.

Slang terms have no source to describe the literal meaning from which they are derived. For example, مالوش حل means that *'has no solution'*, which may appear as a negative term. However, the hidden meaning could contain as positive an indication as *'creative'*. It could also occur as a negative or neutral connotation depending on the context. In addition, غسيل مخ means *'brain washing'*, an expression that has two terms to describe the meaning. Furthermore, the term خبط لزق means *'knock'* and *'past'*, which indicates *'random work'*.

Proverbs are a rich source of metaphor. Some proverbs are recognized and known in Arabic. The challenge Faycel (2012) discusses the meaning of Arabic metaphors in proverbs in food terms. The new dataset has a similar type of data. For example, دس السم في العسل means *'deceiving'*, but our data contain new Arabic dialectal proverbs in a different category, which is an example of the lack of a reliable source with literal and hidden meanings to identify it as a metaphor. Although those proverbs are treated as one phrase in terms of meaning, the labeling which we applied was based on the practical work and the segmentation of the terms to identify the context. Thus, identifying the metaphor was based on the before and after terms. For example, وجع البطن ولا رمي الطبيخ is one of the online Arabic proverbs with metaphor.

3.8 Data collection challenges

The guidelines were described above in detail. The data collection process will be described next. We searched LABR using the *"find"* feature to find the reviews that contained Arabic metaphor terms that were similar to those used in social media. We wanted to find five reviews for each term.

However, we faced difficulty in finding the terms due to the informal online writing style. For example, as already explained, in writing the same informal metaphor term as جامدة and جامده, there are slight differences in the letters ه and ة, which affected finding the required number of reviews containing the same metaphor. Because the Notepad is case sensitive. So, we considered the differences in the online writing style to find the required number of reviews. In addition, the elongation described above affected data collection. For example, the new informal Arabic metaphor term found in social media بيض *('Eggs')* was written as بيييييييض *'Eggs'* with repeated letters to emphasize the opinion.

This exaggerated elongation affected finding the metaphorical term as the Notepad that was used is case sensitive. Those terms were therefore searched with the differences in writing style to facilitate and expedite the data collection process. However, the writing style is still changeable, which means that the LABR could have more informal terms written in such a style, but we cannot predict the online writing style. So, we follow all possible ways to find similar new Arabic metaphor terms. So, due to the unpredictable writing style, we could not have the balanced dataset which we had originally intended by finding five reviews containing each term. So some metaphorical terms occur in only one or two reviews in the dataset. For example, the proverb وجع البطن ولا رمي الطبيخ/*'Stomach pain is better than throwing away the food'*. appears only once in the LABR data, and the term بيض /*'Eggs'* has only two metaphorical appearances. The above challenges discuss the known social media text challenges, which are the elongations and morphology. In addition to the Notepad case sensitivity. In addition, the terms appear once in all LARB Aly and Atiya (2013).

3.9 Results

As explained in the preview sections, I collected thousands of book reviews that contain Arabic metaphor data. Although a large amount of data was collected, the main opinions contained in those reviews are metaphorical. The reviews collected contain sentences of metaphor. There are context words that explain the other factors of the text that affect the sentiment, which is a reflection of the practical work to identify the sentiment. For example, in some of the reviews, the ambiguity of metaphorical sentiment was interpreted from the supporting context of literal words. The supportive context can be clarified from the frequency of the occurrence of the online metaphors. Metaphoric words could be classified as negative due to the fourth or fifth order of words before or after the metaphor, despite the meaning of the metaphor and the indication of negativity.

In addition, metaphor terms were categorized on the basis of the literal use of the metaphor. The categorization started with formal and informal Arabic metaphor terms. The informal categories were constructed on the basis of the literal use of the words. For example, terms with an indication of alcohol were classified under *'drugs'*, despite the hidden meaning of the metaphor. For example, كحولية was included in the *'drugs'* category. The formal topics were categorized according to the regular structure of Arabic words. For example, the word خطفت which means *'kidnapped'*, a notion of captivation, was classified as a verb. Formal metaphor terms are terms used in official writing, such as news reports and articles, and are usually written in classical Arabic and MSA. Informal metaphorical terms have an irregular structure of Arabic metaphor and are written in dialectal Arabic. In addition, some of the informal terms were new in Arabic, which means that they had no reliable source in terms of meaning.

3.10 Dialect authentication

This section discusses the questionnaire to authenticate the Arabic dialect for the new metaphorical terms. However, we did authenticate some of the new Arabic metaphor terms as a sample. Some of the new Arabic metaphor terms were collected later to balance the dataset, and afterward we found some redundancy in the collected reviews. At the same time, this research was carried out for data analysis of the new Arabic metaphor terms with no aim to specify the language. In other words, this research aimed at the practical aspect, and the corpus has been adjusted to fit the practical aim. However, more precise work will make the corpus more coherent. But this corpus, with this specification set up, is enough for this research. In addition, some of those words are used in multiple dialects, as social media writing is not exclusively for a certain Arabic dialect. For example, I may use the Egyptian dialect in writing and speaking to show my opinion sometimes. In addition, from the practical perspective, the dialect is not affecting the meaning, either the sentiment or metaphor, to identify them. So, we authenticate all the meanings of the metaphorical terms, as the metaphor affects the annotation.

Therefore, a questionnaire was devised for some of the new informal metaphorical terms. Other new informal terms were collected later and authenticated by the annotators. The questionnaire had multiple choice questions for meaning and for the Egyptian and Saudi dialects for informal terms. An option was offered ‘Other’ if the information provided was not applicable. Some of the informal metaphorical terms were not authenticated for the dialects. Dialectal authentication was needed to specify the data for the corpus, which will show the corpus as a coherent work. However, as mentioned, this was not a main category for this corpus.

3.10.1 Questionnaire Results

We received 107 responses to the questionnaire regarding the Arabic dialect. From the responses, we observed that some terms were annotated with the option "Both," indicating that the term can be used in both the Saudi and Egyptian dialects. As the percentage of choice was high for a certain choice.

For example, the term تحفة, which literally means *‘vase’*, was annotated metaphorically as *‘beautiful/nice’* and was identified as belonging to both dialects. This term received a 44% selection rate, equally divided between the choices of the "Both" and the "Egyptian dialect". This suggests that the term is commonly used across dialects. Another example of a term used in the cross dialect is شهي, which means literally *‘delicious’* and metaphorically annotated as *‘seductive’*. The percentage of the two choices of the ’Both’ and ’Saudi dialect’ were equal to the %44 percentage.

3.11 Participants profiles

3.11.1 Respondents

To support the authentication of Arabic dialects (Saudi and Egyptian), a questionnaire was distributed to Saudi and Egyptian audiences. Participants were selected based on their native proficiency in their respective dialects and their ability to recognize dialectal characteristics. Participants over 18 years of age included both male and female participants with diverse educational backgrounds. Their contributions were used solely for linguistic verification purposes, ensuring anonymity throughout the process.

3.11.2 Data collectors

The data collection process involved two anonymous participants with computer science backgrounds, one with a post-graduate degree and the other an undergraduate student. Both participants were responsible for collecting Arabic online metaphors, leveraging their technical expertise and familiarity with online content. The postgraduate participant contributed knowledge to find the Arabic online metaphor according to the data collection criteria, while the undergraduate participant focused on manually gathering the metaphors from the LARB Aly and Atiya (2013). Their combined efforts ensured a comprehensive and accurate dataset, enhancing the accurate choice of metaphorical expressions in Arabic dialects.

3.11.3 The annotators

The data annotation process was conducted by two anonymous annotators, both native Arabic speakers with a background in computer science. Their role was crucial in the construction of the Arabic Metaphor Corpus, as they were responsible for annotating multiple categories with metaphorical expressions while also identifying the associated sentiment. Using their knowledge of the Arabic language and technical expertise, they accurately analyzed and categorized the collected data to ensure accuracy and consistency between different types of metaphors. Their combined efforts contributed to the creation of a structured and reliable corpus that captures the new structure of Arabic online metaphors and their sentimental nuances. Anonymity was maintained throughout the research to protect the integrity and confidentiality of their contributions. All annotators are above 18 years old.

Arabic metaphor

Dialect and meaning authentication for online Arabic metaphor

* حدد اللهجه للمصطلح التالي مع المعنى: (تحفة). مثال: (تحفة).بحر من الخيال الممزوج بالتاريخ -1

: اختر من الخيارات التالية اللهجة أولاً ثم المعنى الجازي علماً ان المصطلحات هي استعارية علماً أن الامثلة كتبت للتعليق على الكتب

- [ ] اللهجة السعودية
- [ ] اللهجة المصرية
- [ ] كلاهما
- [ ] ولا واحدة
- [ ] المعنى:
- [ ] رائعة/جميلة
- [ ] سيئة/رديئة
- [ ] اخرى

Figure 3.2: Questionnaire form

Chapter 4

Corpus Annotation

This chapter discusses the corpus building of the Arabic Online Metaphor (AMC) E.1 with sentiment and other categories. The categories are specified on the basis of the practical purpose of extracting metaphor information and sentiment. The chapter also discusses the annotation process in detail, which includes the schema and the metaphor annotation, different structures, and challenges in the online context. The discussion section discusses evidence of the changes in the form of Arabic metaphors in the online context. Also, the use of contemporary Arabic metaphor to indicate an opinion.

4.1 AMC construction: Introduction

In our corpus, we find that metaphor in an online context is affected by meaning and literal sense. For example, we found that the meaning must be annotated to understand the polarity based on the context. Once the annotators understand the meaning, the sentiment can be specified. In practice, to identify sentiment with metaphor, a process for classifying metaphors and word disambiguation is required, which was applied in the SentiFig system for the English language in Rentoumi (2012). So, the annotation for meaning was essential before the sentiment annotation.

When it comes to constructing an annotation schema, the structural differences in languages have led to different methods of constructing a corpus. From the perspective of computational linguists, differences in the construction of a corpus depend on the preferences and features of the language to be specified (Garside et al., 1997). However, this is not the case for the unpredictable Arabic online metaphor, as there are no specific features to follow. The schema was designed purely for practical purposes. Hence, the schema reveals the pattern and the feature extraction. To clarify, previous studies have annotated metaphors using a standard concept such as MIPVU (Metaphor Identification Procedure VU University Amsterdam). Some studies used the SemEval-07 schema to annotate metaphor data such as S. Mohammad et al. (2016) to add new annotated data to the same group of data. Our schema was designed specifically for Arabic online metaphors in its specifications. The schema was designed not to add to a specific lexicon for Arabic but for practical use for the computational aspect of the study.

Although the new schema was influenced by VU Amsterdam, it is different from that schema based on the purpose and the structure of our data. For example, we used the OXygen tool for annotation, whereas XML was used for their annotation. The VU Amsterdam Corpus annotated each word of each sentence by lemma and part of speech. The annotation has different tags for literal and unknown metaphors. Also, the VU Amsterdam corpus annotates all the words of each sentence and identifies metaphors through the annotation, whereas our annotation involves extracting the metaphorical context of each review and specifying the metaphor terms for annotation. One of the research aims is to design a schema to annotate Arabic informal metaphorical terms with sentiment because building an Arabic metaphor resource is the foundation for further work with an advanced technique for training and testing.

The vast majority of studies conducted on the Arabic metaphor discussed the metaphor conceptually only. These studies discuss the challenges and suggest approaches without experiments. However, many studies on identifying English metaphors have used supervised and unsupervised datasets. Regarding conceptual studies in Arabic, there are limitations when it comes to building a figurative dataset. For example, Alsayat and N. Elmitwally (2020) discussed the identification of Arabic sentiment analysis on all linguistic levels. That study built an Arabic figurative dataset for hyperbole and simile by annotating about a thousand sentences using manual annotation by two Arabic speakers. Although that study discussed examples of simile in regular text, annotations were performed in Qur'anic text. The Qur'anic text is miraculous and has a fixed structure and sentiment, as discussed in the previous chapter. The data involved in the study of N. S. Elmitwally and Alanazi (2020) did not have a data collection source, but the examples represented in the study showed that the data came from Qur'anic text. In addition, no previous studies have been conducted for the construction of the Arabic metaphor corpus. The limitations of the datasets which have been built are that they were solely for linguistic studies and for specific domains in particular dialects. For example, Faycel (2012) conducted a study to build an Arabic metaphorical proverbs dataset in the food domain for social science and psychology in the Tunisian dialect. In addition, Reyes and Rosso (2012) collected Arabic everyday metaphorical terms, proving that the metaphor is not only used in poetic and rhetorical contexts but could be used informally in everyday speech.

In Raii (2009) dataset was divided into multiple categories such as terms for location, time, and politics. Similarly, Gholami et al. (2016) conducted research on metaphor translations from Islamic books into English, which explained and analyzed the metaphor conceptually for the linguistics field. Previous studies collected data comprising regular Arabic metaphors, which means MSA, dialectal Arabic, and classical Arabic metaphor. The dialectal and MSA in Reyes and Rosso (2012) have a certain format, which is a metaphor that comes as a verb itself rather than as an object or as a symbolic term. For example, مر تشرين بسرعه means *'October pass fast'* in a literal sense; they consider the metaphor in the verb (pass) as logically months cannot pass. In addition, they are in speech in day-to-day life, which may be similar to the social media ones, but not in the same types and variety structure of this research. In addition, Gholami et al. (2016) has the type of text that is written in classical Arabic or MSA and is the standard structure of the Arabic metaphor. Whereas the metaphors collected for this study have different forms starting from one term to dialectal proverbs.

For example, the sample data contained expressions that were considered as metaphors but which did not have the regular structure of an Arabic metaphor. The irregularity in those expressions could be seen in the absence of the forms which construct standard Arabic metaphorical sentences. Therefore, the linguistically built metaphorical Arabic datasets in previous studies which could have been used and tested were not appropriate for the purpose of the current study, as stated above. In addition, they are not accurate for any study in identifying the sentiment and the metaphor, unless there is a previous labeling.

We therefore designed an Arabic metaphor schema and created a unique annotated metaphorical dataset with sentiment, meaning, and context for the Arabic language by analyzing the new Arabic online metaphor structure. The new dataset can be considered as the foundation for the current study, but it is also a potential source for testing and developing work in the future. In addition, the new schema has the possibility of being used as an additional annotation tool.

This chapter describes the core work of this research. There are sections and sub-sections explaining the design of the new schema and the annotation in the context of the challenges and uncommon metaphor structures which were encountered. The annotation work shows the challenges of manually annotating Arabic metaphors and also shows the impact of the Arabic metaphor on the identification of sentiment. The impact was shown by evidence from Arabic online metaphor examples.

In the results section, we present the statistical information of the generated AMC corpus. The table 4.1 has the frequent number of different ranges of review lengths in the AMC. The table divided the AMC into different categories based on the reviews' lengths and the polarities. Such statistics are important to show the majority lengths of the AMC reviews 4.4. In addition, the table shows the highest percentage of the polarity taken from the gold standard annotation 4.1. Moreover, it shows the percentage of the highest category range of lengths. The length affects the annotation decision. So, the percentage used to show that the AMC contains only %7 of the lengthy reviews. In the conclusion chapter 6.1, I shall discuss the limitations of the current study, one of which is the specific domain of the dataset, which was book reviews. Other limitations are the specific type of Arabic metaphor and small amount of Arabic metaphor reviews.

4.2 Arabic Metaphor in online context

The annotation produced the Arabic Metaphor Corpus (AMC) E.1 as a result. The AMC was annotated mainly for sentiment and metaphor information, while the other categories of meaning, theme, context, and part of speech were added to identify the Arabic metaphor in an online context for practical purposes. Metaphor, in general, is one of the most widely used linguistic devices according to Jakobson (1956), but the Arabic metaphor has previously been widely used in poetry, which considers the elegance of the Arabic language.

The structure of the Arabic metaphor has three pillars. Arabic metaphor by definition has a structure opposite to that of the Arabic simile (Alsiyat and Piao, 2020a). A simile has three essential elements: likening, device, and trait. Metaphor can be defined through simile only when the tool is deleted, which defines metaphor. The first pillar is likening *al-mushabbahu* things or concepts and has an object which is projected as a simile onto either a thing, a concept or a person. The second is *Adātu at-tashbīhi* means the simile tool is a device used to link the two elements of the simile together, which are مثل means *'like'*, ك means *'as'*, تشبهين *tushbihīn* means *'to look like'*. *al-mushabbahu bihi* is also a thing or a person that is related to the first pillar.

The Arabic word is a simile, except for the simile device *Adātu at-tashbīhi* and the other pillars replaced with a trait that indicates an absent object. In Alsiyat and Piao (2020a), Arabic metaphors are grouped into two main types as shown in Figure 4.1.

However, Arabic metaphors can have different structures depending on the type and context. For example, metaphor used in the Qur'an has a different structure from metaphors used in a regular context, as the Arabic Qur'anic metaphor is miraculous, whereas the metaphor has a regular structure in regular Arabic writing. In the online context, as discussed in the previous chapter, the metaphor has changed to an expression or a phrase used to express an opinion. Arabic metaphors have changed to adapt to the online practice of communication, which means that metaphors have become short for rapid use. As a consequence, the structure of the Arabic metaphor in the online context changed to be represented as one symbolic term to adapt to the rapid use. For example, the Arabic metaphor in the online context uses semantics to describe opinion. Clarify a symbolic term such as food, drug, and weapon to express an opinion.

The problem arises in identifying Arabic metaphors in practice as there is no reliable resource for the Arabic language for metaphors expressed with sentiment. In addition,

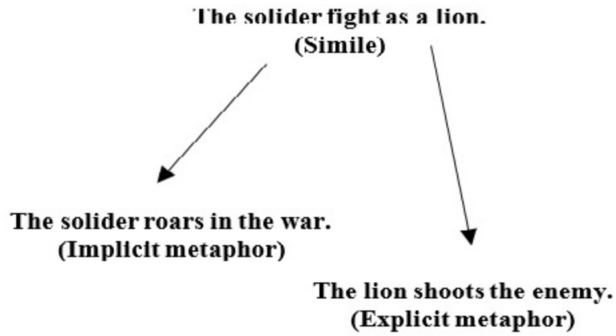


Figure 4.1: Example shows the differences between metaphor and simile

there is no reliable Arabic lexicon resource that can be relied upon to identify a metaphor, such as English metaphor identification. English metaphor identification is based on a lexicon categorized by part of speech, which meets the requirement of violating verb/noun agreement, such as WordNet and BNC.

In comparison, Arabic has large lexicons, but they were built for Arabic sentiment analysis. Another problem caused by the lack of resources built for Arabic metaphor is the identification of sentiment in a metaphorical context. Furthermore, as discussed in the previous chapters, the Arabic online metaphor has its own specification for being interpreted as a figurative feature of the Arabic language, not to mention if it is in an online context. So, for the identification of sentiment, an online metaphor needs to be interpreted in advance. Most of the previous studies on automatic metaphor identification in relation to sentiment for the English language rely on annotated data. However, advanced methods need a large amount of data to predict metaphors without prior annotation. Advanced methods such as word embedding, neural networks, short-term memory (STM), and machine learning algorithms.

4.2.1 Why using the traditional method of annotation rather using an annotation tool?

"There is no publicly available text annotation tool that supports Arabic text annotation efficiently" (Al-Smadi et al., 2015). The previous statement is applicable to the automatic annotation tool as well. This means that there is no publicly available Arabic annotation tool to automatically annotate an Arabic text efficiently. The reason is that no tool can automatically collect the sentiment, part of speech, meaning, and other categories. There are Arabic annotation tools for certain types of text, such as semantics, morphology, and sentiment. The tools are not targets for our purpose of annotating Arabic metaphor. The aim is to find an Arabic annotation tool that can be fed with guidelines to annotate the Arabic online metaphor similar to BRAT, which is made for the English language. MADAD Al-Twairesh et al. (2016) is a web-based Arabic annotation tool. It has a friendly user interface and features to add the annotation schema. It is one of the tools that may fit the annotation for the Arabic online metaphor. However, the tool is not available for use or testing. AraBERT Antoun et al. (2020) is one of the annotation tools available for Arabic text for sentiment, named entity recognition, and question answering. The tool is made for classification, but it is not for manual annotation to use for Arabic text.

Find a tool to manually annotate the Arabic text with regard to metaphor and sentiment from English language tools. BRAT for example Stenetorp et al. (2012) could be one of the options, but it does not support the Arabic language. So, it cannot annotate the Arabic text using Arabic tags, nor automatically recognize the Arabic text type. Although there are some attempts by developers on GitHub to adapt the BRAT tool for the Arabic language, it is still not sufficient to use. It has not been officially added to the tool website. However, some studies have used BRAT (Stenetorp et al., 2012) to annotate Arabic text for the analysis of aspect-level Arabic sentiment analysis (Al-Smadi et al., 2015). They used a SemEval2014 configuration file for annotation to create the annotation schema, which is not applicable to our research. The Arabic metaphor requires a new schema to accommodate its unpredictable structure. Therefore, our schema has been adjusted to fit only the cases necessary to show the impact of Arabic metaphors in the online context on sentiment. For example, we added another layer of annotation during the process to determine the text that clarifies the metaphor term, specifying whether the context is before or after the metaphor. However, annotation schemas from the English language are used to annotate. So, based on the previous discussion, the Arabic Metaphor corpus was annotated without an annotation tool and following a newly created scheme using the XML editor Oxygen. Basically, the XML tags were created prior to the annotation; then the schema was extracted. Then we extracted and analyzed the features of the Arabic online metaphor to show the impact of metaphors on sentiment.

4.2.2 The Experiment

In this section, the annotation process and schema design are explained, as well as the target dataset, which was LABR, and the manual selection of Arabic metaphorical reviews. As has already been explained, LABR is a publicly available dataset containing 63,000 book reviews annotated at the sentence level. It is one of the largest Arabic datasets for sentiment analysis. The fact that the LABR contains book reviews was one of the reasons for choosing it as a source because people tend to write more about books in reviews, as shown in the repeated lengthy reviews in the dataset. Reviews containing metaphors were carefully selected from the dataset following our criterion for selecting metaphorical expressions that had polarity. For example, we ignored sentences that contained metaphors as a citation from a book. Also, from lengthy reviews, we selected only the part which combined metaphor, sentiment, and the main opinion. For example,

عجبنى أوى الجزء الأخير
اللى بيقول فيه استقرّت السماء واستقر الشاطئ
واحسست ببرودة الماء في جسمي
وسمعت طقطقة مكنة الماء

'I like the last part the most, which said:
"and the beach settled and I heard the water
machine whirring and I felt the coldness in my body"'

In this review, the reviewer gave a quote from the book that contained a metaphor, but the metaphor part of the review did not contain sentiment or give the writer's opinion of the book. Reviews that contained facts, which are considered as natural polarity, were counted in the annotation. Also, we ignored sentences which had two contradictory polarities for two different aspects of the book. However, we did include parts of the sentences that contained a metaphor. We wanted to have one polarity in each sentence, with versatile cases and types of metaphor.

The sentence رائعة.وصادمة لكل الدافنين رؤوسهم في بحار الرمال *'Wonderful. And shocking to all those burying their heads in the seas of sand'* is one of the cases which had two distinct opinions of the book, with a factual statement at the end of the sentence, which included a metaphor. We ignored such cases because the metaphor part did not have any sentiment. The manually collected and annotated dataset accurately shows the metaphorical Arabic discourse used online.

4.2.2.1 Designing the annotation schema

The annotation schema was designed to interpret the Arabic online metaphor with sentiment in order to extract information and show the impact of metaphor with sentiment. The task involves finding the accurate elements and attributes to fit the aim of this research. Because the annotation is based on extracting information. So, the annotation elements and categories should fit the aim. In our case, the schema was designed first to interpret the text to extract information and show the impact. For example, the metaphor should be interpreted based on context and sentiment as well. After the annotation, the feature extraction was extracted from the annotation. While the impact of metaphor is shown during the annotation as well by the different sentiment annotations of the same metaphor in different contexts. The elements of the schema involved the before and after context to specify the metaphor, hence sentiment. But during the annotation, the meaning plays a significant role in interpreting metaphor and sentiment, which will be explained in detail in the coming sections.

The design of the schema started from the previously mentioned concept of fitting the purpose of the annotation. These rules for semantic annotation were sufficiently detailed to enable them to be used with any language for general semantic annotation. The rules in general can be categorized for different purposes, so the overview was that semantic annotation should not follow the rules, but should be based on factors such as the language, aim, and domain. However, we did use some of those rules since we had one semantic field on which to work, which was metaphor. The schema was primarily designed for identifying metaphor and was divided into the two types of metaphor, formal and informal. Informal metaphors usually occur in a sentence as an expression of an opinion which defines the sentence polarity, whereas formal metaphors have the full regular structure of Arabic.

The design of the annotation schema was influenced by Krennmayr and Steen (2017), but was modified to fit the computational aspect of this study. For example, the context before and after a metaphor term was added to identify the metaphor, while Krennmayr and Steen (2017) annotated the entire sentence and identified metaphors during annotation. Specifically, the VU Amsterdam Corpus Krennmayr and Steen (2017) annotated the sentence and theme in metaphoric form if there is a metaphor in the sentence. So, the corpus was specifying whether there is a metaphor in each sentence or not. Although our reviews collected all contain metaphor annotated with metaphor term, sentiment, theme, context (before and after), part of speech, and meaning. So not all sentences annotated only the metaphor expression from each review collected with its context.

The schema was designed to fit the purpose and the structure of the Arabic metaphor in the online context. However, the schema may not be useful for metaphor in other languages unless the structure of online metaphors in those languages is similar to that of Arabic. We designed and annotated the schema using the Oxygen XML editor. This is an application for generating an XML file which can be assigned to a friendly user interface. The schema was designed to identify metaphorical words. It was divided into three basic attributes, metaphorical meaning, metaphoric meaning and context. The metaphorical terms have two types: formal and informal metaphor, according to the part of speech. The literal meaning can be identified as a near synonym to the metaphorical words. Also, a literal tag was used to find the metaphor's meaning, which facilitated the process of identifying the sentiment. The inconsistent structure of Arabic online metaphors in XML was shown in an Excel table with specific categories; for example, a metaphor written in an Arabic dialect prefixed with a pronoun to indicate the future tense, as shown in Figure 4.2.

However, the annotation considered the litter هـ in the word هيلحس means *'will'* /(يلحس which means *'licking'*) as a pronoun as it was new to the regular Arabic structure, and the context would make it possible to identify a metaphor from the neighboring words. The genre element was specified to define the topics related to each sentence, which could be a good resource for statistical information. The genre is represented by four domains of the dataset: the book, general, the writing style, and the author (see Figure 4.6).

Those genres were categorized according to the frequency of the subjects used in the reviews. Although there were metaphorical sentences with different topics in the gathered dataset, those sentences were too few to group in another genre. So we chose the theme tag with the topics as attributes with 'Name' type using Oxygen feature to produce an automatic schema. The general genre was most used to describe an overall opinion of a book and the topics which were discussed in it. The general genre was also represented by the reviewer's comment on the incidents, characters, or conclusion of the story. The writing style genre is used for reviews which describe the author's approach to writing. The author genre is used in reviews which criticize the author's work. The genre tag was used to specify root and word id tags. Some sentences had multiple genres with different sentiments, but those sentences were annotated according to the part which contained a metaphorical expression. It should be noted that sentences which contain two genres and sentiments can be a potential topic for aspect level annotation in future work. To illustrate, the schema for each sentence applied to different levels:

- Annotate the metaphorical words according to their type.
- Specify the meaning of a word which is the nearest meaning to the metaphorical word.
- Annotate the context, looking at the three words before and after the metaphors based on the sentence length.
- Specify the part of speech of the above tags.
- The genre tag, which is defined as a sub-tag for each sentence.

4.3 The annotation

This section describes the annotation rules for the two stages of annotation: the metaphor identification annotation stage and the sentiment classification stage with the gold standards for the overall and metaphor, and discusses the details of the annotation for each annotation category: the metaphor type, the hidden meaning, the context, the theme, and the part of speech for each metaphorical section, in addition to the manual annotations of sentiment in the review and the metaphorical section.

4.3.1 Annotation guidelines

In this sub-section, the guidelines provided to the annotators are explained, but the annotation will be described in detail in subsequent sections. The guidelines provided were similar to those provided for the data collection. For example, different Arabic native speakers were asked to collect the reviews that contained new Arabic online metaphors and then to annotate those that appeared in the selected reviews. So the guidelines were similar to those for the data collection and slightly different from those for the annotation. Manual annotation was performed in two stages. The first stage of the annotation was adding XML tags of metaphor, genre, meaning, and context. The second was the creation of an Excel table of overall sentiment and metaphor sentiment. Automatic annotation was evaluated and compared with the manual annotation. The guidelines for the two stages are given here whereas the actual annotation details will be explained in the following sections. The guidelines were:

- The annotators were provided an annotated sample of XML tags for a review with all tags. Also, each tag was explained.
- The reviews provided were not annotated in full; only the section containing a metaphor should be annotated.
- The annotators were asked to select new metaphor terms which combined opinion about the book and the sentiment, as there were cases of metaphor which had no opinion but only a quotation from the book.
- Each tag was provided with annotation instructions. For example, the part of speech (POS) tag was annotated using NLP Stanford POS list, so the POS list was provided for use during the annotation. However, NLP Stanford did not recognize the correct POS for some of the Arabic dialectal pronouns, so the POS was annotated by one of the native speakers rather than both annotators. This means that the POS, genre/theme and metaphor type were annotated by one of the annotators as those categories would not affect the annotation quality but were added as resource for the corpus.
- The reviews were provided in the form of an Excel sheet for the native speakers to annotate the overall sentiment.
- For the metaphor sentiment annotation, the metaphor was labelled from the first tag, so the annotators were asked to identify the sentiment for the metaphor section.

4.3.2 Annotation in general

As already explained, the manual annotation process was done in two stages as we wanted to use them in practice after we had collected the metaphorical data. The first annotation stage involved annotating the metaphor with metaphor type, genre, meaning, and context. The second stage was annotating the overall sentiment of the metaphor section. The annotation started with XML tags which were later converted to Excel tables to facilitate the analysis and extract information. The annotation started with the sentence as a root tag and each target word was manually identified and classified in a metaphor tag according to its type. The literal meaning was derived from the target word (the metaphor) and identified for its POS. The literal tags, as already mentioned, were specified by finding the nearest synonym to the target word. However, some of the informal cases had to be specified as sentences to clarify the meaning. For example, the informal metaphorical expression بعثرتني, which means *'scattered me'*, was identified by literal tags as a sentence جلبث الحنين / *'brings the nostalgic'*. The literal tags were designed not to express the literal meaning of the metaphorical words, but the meaning behind them. For example, تغرد خارج السرب means that the context is unusual, whereas the precise meaning is that *'the birds sing out of the group'*. Furthermore, the genre was annotated according to the topics specified above. Each review was expected to have one topic with one sentiment, but some sentences had multiple genres and sentiments, so the annotation in this case focused on the part which contained the metaphor and ignored the parts which did not. For example, the sentence

رواية مشوقة جدا وغاية في الابداع
بس النهاية حرقت دمي
كانت ممكن تطلع احسن من كده

'so exciting novel and so creative'
'But the end burn my blood'
'it could be better than this'

has two aspects, one expressing a good general idea about the book and the other expressing in detail the bad conclusion of the novel. The genre is specified not just as a classification of the topic but as a previous context of the target word as a formal metaphor. As mentioned before, some reviews had two aspects of sentiment, but others had one expression of neutral sentiment with another sentence of positive sentiment. In our annotation, we took the positive aspect which elucidated the first part, and both sides complemented each other. This case is different as it contains one topic with a different sentiment rather than two topics with two sentiments. Also, the metaphor was our main goal, so we annotated and elucidated the second part.

However, the context tag cannot be identified for the informal words as it is assumed that it is basically a metaphor. In the annotation of one expression, we omitted the context tag. Informal metaphors with one expression were regarded as metaphors because there are no words to show the metaphoricity of the sentence. In each formal metaphorical sentence, the target words were defined and the context words were classified using the three words after the target word. In sentences in which the metaphor occurred at the end of the sentence, the context words were the three words preceding the target words. In exceptional cases where we could not use the three following words and the metaphor occurred at the end of the sentence with only two words of context preceding it, we omitted one of the context tags. It should be noted that this concept was compatible with the generated dataset, but not for any data. When it comes to sentiment detection from the features defined according to the trained dataset, one expression was assumed to be a metaphor but not an expression of sentiment. It was necessary to decide whether it is always positive or always negative. In such cases, the one expression of a metaphor was swapped with the literal word to facilitate the process of identifying the sentiment. So the previous concept is not appropriate for any data as one positive expression such as رائعة / *'amazing'*, which is considered to be a literal expression in another dataset, cannot be classified as a metaphor because of the pre-trained dataset for semantic annotation and word embedding classification. In addition, the classified target words need further classification by tagging the type of metaphor and the PoS. The table above shows the XML representation of the tags mentioned earlier (see Figure 1.1. Problems arose with informal metaphors because a 'one-expression metaphor' has a different meaning and sentiment from what it shows. For example, خطيرة means *'dangerous'* as its literal definition but *'fabulous'* as its metaphorical definition. Another example is بعثرتني, which means *'scattered me'* in its literal sense whereas its metaphorical meaning is positive, which is *'make me nostalgic'*. Also, in the sentence سلاح فعال لمكافحة الاكتئاب, which means *'An effective weapon for combating depression'* that the book is a cure for depression as a motivational implication. However, سلاح meaning *'weapon'* is interpreted as a negative word when it comes to the sentiment. So identifying metaphors including sentiment was not only challenging in terms of finding the meaning and identification, but also for finding the sentiment, which can be identified by using the pre-trained dataset . The challenging aspect of informal metaphors which have one expression is not just the implicit ambiguity, but the fact that they are derived and invented as a fortuitous expression in everyday conversation. Also, the expressions جامده and تحفة mean *'wonderful'* and *'amazing'*. Another case of an informal challenge is when a metaphor comes in a two-word expression because the two words are essential for the term to be classified as a metaphor. In this case, we annotated the two words as a metaphor. In fact, informal metaphors have different cases because of the irregular structure of Arabic metaphors. A formal metaphor has The complexity of identifying a metaphor does not solely depend on the intuitive nature and the ambiguity; it also depends on the complications of the Arabic language structure. In Arabic speech, MSA in particular, metaphor can be used as an interlapping loop of description as it strengthens the author's evidence to convince the reader. For example, the review

هذه الحكاية تخاطب الروح والعقل.
وتمازج بين اضطراب الفكر وضايع النفس
حتى تقطفه كدمعة ثم تعود تلك الحروف لتعالج ذلك

'This story speaks to the soul and brain.'
'and compose between the confusion of thoughts and self doubt'
'until it picks it as a tear then those words comeback to cure that'

has two metaphorical sentences, which is another irregular case for annotation. The sentences were annotated starting with the first sentence, which has a context that is metaphorical, so the context was annotated as a sub-tag metaphor with a literal and another context tag. Metaphor has a compound nature as well, and the sentences in our dataset showed that metaphor can come in a two-word combination. The two words complete the full meaning of metaphor, which is known in Arabic as an expression used in specific forms such as proverbs. The annotation in this case was based on the type of the Arabic sentence. Grammatically, Arabic sentence types are divided into two: the nominal and the verbal type, and the PoS will be classified according to the type of the sentence. However, the type of metaphor will retain the same classification as those expressions known in the language. The reason for this is that those expressions are considered as MSA. Evidently, Reyes and Rosso (2012) noted that the metaphors collected are invented in everyday life, so they are dialectal and not standard Arabic or MSA, and they are used in different contexts with full expression and pillars of the Arabic metaphor. Another compound case is when a review had continuous metaphorical sentences containing two-term expressions, meaning that this case is a combination of two cases, one of which is the two-term expression and the other sequential metaphorical sentences. For example, the review

من ولد الخير ولد له فراخاً تطيربالسرور
من ولد الشر انبت له شجراً أشواكة الحسرة وثمرة الندمفرحم
امرؤا أغضى عن الاخطاء واستمتع لظاهر

'who born the good, it will born for him a chicks flying with pleasure'
'who born the evil, it will plant for him the trees with heartbreak thorns and the fruit of regret'
'God mercy who avid other mistakes and enjoy what is obvious and clear'

has two metaphorical expressions annotated irrespective of the neglected part which does not contain the metaphorical expression. In the annotation, the review is represented by a metaphor tag with nested metaphor tags inside the context tag for the first one. We faced another case of metaphor annotation which is proverbs. Proverbs are frequently used in

Arabic and are based on Arabic culture and history. As already stated, there is a previous study of Arabic proverbs in the linguistic field (Farra et al., 2010), but it was limited to the domain of proverbs. The annotation was based on the overall meaning and the literal meaning of the proverbs. There are usually no context words for proverbs as the whole expression is regarded as a metaphor. This is because proverbs usually signify semantics, which are traits and adjectives, to express the meaning. In the literal sense, the nearest synonym can be identified using single words which carry the main meaning behind the proverb. In our dataset, the proverb وهل يخفى القمر؟ means *'can the moon be hidden?'* which signifies something famous and well-known, so the literal tag specified as 'famous' was applied. For example, the proverb السم في العسل literally means *'the poison hidden in the honey'*, which indicates that bad thoughts in the book are deliberately presented in a good way which cannot be interpreted by others. The parts of speech for those expressions were classified as a nominal or verbal sentence. In the review,

لا أتردد في وضع أعلى التقييم لهذا الكتب وهذه القصة الأدبية الراقية،
، لمست أوتار قلبي وأحببت من خلالها القسطنطينية وجسورها المعلقة

'I do not hesitate to put the highest stars for this book and this high class literary story'
'it touches the cord of my heart and I like through'
'this story the Constantinople and the suspension bridges'

Some sentences used metaphor, which is one of the hyperboles used to highlight the reviewer's opinion, whereas the sentiment can be understood from the context without the metaphor part. In this case, we took those sentences into consideration for annotation. For expressions which were identified as being entirely composed of metaphor, the literal tags were specified by finding the hidden meaning of the expression. For example, تصنع من الحجر شيئا لينا means *'create something soft from a stone'*, which means that the novel was influential. The writer was referring to uncertainty about the novel, so the interpretation had to be general, although the meaning is understood that the novel can affect him/her to be kind or soft. Although we were aware of the MIPVU schema's efficiency, the intention was to have a new schema adapted to the Arabic metaphor structure as well as to achieve the research purpose of metaphor classification.

The MIPVU is the standard method of the metaphor definition for annotation. Means it defines the steps of identifying metaphor this including the annotation assessment, which is known as Inter Annotator Agreement (IAA). Our annotation method certainly follows the standard aspect of this protocol. The annotation follows the identification of metaphor, the inter-rater agreement, and the final assessment for the reliability. However, the details steps of MIP to identify the metaphor have not been followed precisely. As the Arabic online metaphor has a different structure than the English language. Means that the Arabic online metaphor is even new to the Arabic language itself. So it cannot be compared to the English language. However, the annotators were guided by a list of annotation guidelines and instructions to annotate the metaphor and sentiment 4.3.1.

4.4 Corpus annotation categories

The previous section discussed the annotation in general, explaining the cases faced during the process, but the annotation categories were only mentioned briefly. In the following subsection, the annotation categories will be discussed in more detail.

4.4.1 ID number

The id number refers to the sentence number from the corpus. These numbers were applied sequentially to organize the annotation as there were similar metaphor terms annotated in different sentences. The id number specified in the XML tags was
⟨s no= ' '⟩ (see annotation sample 4.6).

4.4.2 Genre/theme

The theme was specified based on the subject of the metaphor term. For example, in Figure 4.6, the theme word was specified as *'book'* كتاب, so the metaphor section, which was قتل وجعه وجماله, which means *'It killed its pain and its beauty.'* is referred to in the book. So the theme is for the metaphor section, which means the word location in the reviews is far from the annotated metaphorical expression 4.4.2. However, the theme was sometimes far from the metaphorical expression. So we annotated the theme word from the general context, which was not necessarily included near the metaphorical expression context. For example, the review theme

مُبكية. ومؤلمة. ورااائعة حدّ الثمالة !
عملٌ مُتقن ، لا يُضاهيه عمل.
أحببتُ الخيل من رواية نصر الله هذه : مُبهرة . وكَفى !

'So . hurtful. and wonderful till intoxication !'
'well versed work, incomparable to any other work'
'I loved the horses from Nasar Allah novel this one: fascinating. period'*

was annotated as general because the only word which indicated the generality of the review was the word عمل, which means *'work'*. So, the annotators specified the theme based on this rule. In circumstances where there was no word in the review to indicate the theme type, meaning that the metaphor expression referred to an absent target, we assumed that the review was about the general domain, which is 'book'. For example, proverbs have no word to specify the theme as a proverb could be stated for any similar situation. We therefore also assumed that the theme was about the book. In this case, we annotated the theme word using the word which most explained the metaphor expression. An example was السم في العسل, which literally means *'the poison in the honey'* which is an indication of deceit. The proverb has no word which specified the book, but since the data were from a book review domain, the proverb's aspect was assumed to refer to a book as well. In addition, the entire corpus has only one aspect, which was book, so the reviews were all about books, but in different categories. So we specified the theme types based on the frequency of the review topic. Although there were reviews which did not directly discuss the book rather than the book content or subject related to the book, the theme types were still specified as general.

4.4.3 Metaphor types

The main metaphor types are formal and informal, but we grouped both types in categories based on the literal appearance of the metaphor term. The categories for the informal terms could be considered similar to the semantic categories, but for the metaphorical terms, the categories were food, drugs, illness, and so on, as previously explained. Formal metaphors were grouped based on the term's PoS. For example, the formal metaphor term in 4.6 قتل means *'kill'* and was specified as a verb. The metaphor in 4.6 was annotated as formal because the term was written in MSA and the review had the Arabic metaphor's full

```xml
<?xml version="1.0" encoding="UTF-8"?>
<sentences>
    <body>
        <s no="1"> هيلحس دماغك <genreT Type="general">هيلحس</genreT>
        <metaphor Type="Verb">
            <Informal>يلحس</Informal>
            <Tword Type="MD">ه</Tword>
            <Tword Type="VBP">يلحس</Tword>
        </metaphor>
        <literal>
            <Tword Type="VBP">يحيرك</Tword>
        </literal>
        <context Type="before">
            <Tword Type="NN">دماغ</Tword>
            <Tword Type="PRP">ك</Tword>
        </context>
    </s>
```

Figure 4.2: Informal Metaphor

structure. In the metaphor قتل means *‘kill’*, it refers to تاريخ means *‘history’* and the cause is دون في كتب التاريخ/*‘it written in the history book’*.

In the review 4.2, the metaphor was specified as informal because it was written in the Egyptian Arabic dialect and did not have the regular structure of an Arabic metaphor. The metaphor specified as هيلحس means *‘will lick’*, which is an indication of confusion, prefixed with ه *‘H’*, which means *‘will’* in dialectal Arabic, and the metaphor referred to an absent target, which is the book. دماغك/*‘your brain’*is a context which specified the metaphor. We considered the different structure of online writing which follows the dialectal pronunciation. For example, in 4.2 the *‘H’* ه in the metaphor was considered as a pronoun during the annotation. However, the pronoun and the part of speech were not included in the corpus excel sheet. Cases of Arabic metaphor dialectal structure appearing during the annotation were observed and discussed as one of the annotation challenges.

4.4.4 Part of speech

The PoS was specified for each annotation category. Stanford could not recognize the negation and the demonstrative pronouns written in dialectal Arabic. For example, مش and دي mean *‘not’* and *‘this’* respectively. So, the Stanford does not have the correct PoS for all the annotation categories as they are written in dialect. For example, in the review دي مش ابلد دي خرابة/ *'this is not a country this is corral'*, the word دي, which is one of the annotation context words, means *'this'*, annotated as a noun using Stanford for PoS. So, the part of speech categories were annotated by the annotator using the part of speech tags from the Stanford tagger. For example, we use *'NN'* to indicate the noun. The *'NN'* used in this form of writing in the Stanford automatic annotation for part of speech.

4.4.5 Metaphorical meaning annotation

The category was annotated based on a metaphor's hidden meaning, and the literal meaning is the translation of each metaphor term, which was identified during the discussion and is shown in Table 4.2. The literal meaning is one word of the metaphor term's meaning, so the annotator used the closest meaning to the metaphor term. For example, in Figure 4.2 the metaphor هيلحس دماغك which means *'will lick your brain'* was annotated as a notion of confusion. There were also cases where the metaphor came as a one-word metaphor and had a meaning for a one-word metaphor. For example, بيض means *'eggs'*, which was annotated as 'bad'. For the literal annotation, the PoS was specified for the literal annotation as well. In addition, all meanings in the metaphoric sense were specified by the annotators.

4.4.6 Context

As already explained, the context is the three words before or after the metaphor term, and the type of the context is the location of the context words in relation to the annotated metaphorical term. Punctuation was considered in the context annotation. Punctuation affects a metaphor's meaning and specifies the type of the metaphor term. For example, if a full stop comes after a metaphor, it specifies a stand-alone metaphor. An example is بيييييييض. من وجهة نظرى يعم/*'Eggs. from my opinion my uncle'*, where the full stop comes after *'eggs'*, which is the metaphor. The context words were specified based on the words used to interpret the metaphor, whether they come before or after the metaphorical words. The context word's PoS specified this.

4.4.7 Sentiment analysis annotation

The Arabic sentiment analyzer was used to predict the sentiment. Mazajak is an automatic Arabic sentiment analyzer with a friendly user interface. It is a web-based analyzer; it accepts Arabic text only as input in a dialog box. Then it predicts the polarity with a choice of correcting the polarity if it is wrong. اعتقد اني كسرت صيامي على بصل/ *'I think i broke my fast on an onion'* is an Arabic metaphor expression, and the automatic sentiment analyzer Farha and Magdy (2019) predicts it as positive polarity. If a sentence starts with a positive opinion of the writing style followed by a contrary opinion for the idea, the review ends with a negative metaphorical expression. Logically, the annotation for the overall sentiment is neutral since the sentiment analyzer is not designed to predict the sentiment aspect.

However, Mazajak sentiment analyzer seems to predict the polarity of the first section of the sentence. This means that the sentiment analyzer cannot see the metaphorical expression which occurs as an illustration of the reviewer's opinion. It predicts the sentiment using the literal words with polarity to specify the sentiment. Although the correct annotation was neutral, the sentence was more on the negative side. So the sentiment analyzer could not identify the aspect level, which is the number of negative/positive terms in each sentence, to identify the polarity accurately and, most importantly, predict the metaphor section. The metaphor is the last main opinion, which comes as negative towards the end of the sentence and shifts the sentence's overall polarity. The Mazajak Arabic sentiment analyzer neglects the fact that a metaphor expression affects the accuracy of the prediction. Another example demonstrates the previous fact using the automatic Arabic sentiment analysis:

الكتاب بجد جامد جداااا الجزء الاولانى ساخر
بس مضحكنيش اوى الجزء التانى اللى بيحكى فيه
عن مواقف حصلتله بجد تحفه وملوش حل

'The book is really amazing! The first part is sarcastic'
'But it didn't make me laugh much, the second part that he tells about'
'It's about real situations that happened to him, truly amazing and has no solution'

The prediction was negative for the previous sentence, whereas the logical overall annotation is neutral. It should be noted that the last section of the sentence has a positive metaphor expression, and we have similar terms annotated. But the previous observation is based on one sentence; however, it could be proof if there are multiple sentences with the same behavior of the sentiment analyzer with metaphor. The Table 4.2 shows the informal Arabic metaphor terms, which usually have a different structure of Arabic metaphor than the regular metaphor written in MSA or dialectal Arabic. Also, the terms are translated into English, assuming that they refer to a book with different categories, which are the author, the writing style, or the general genre.

4.4.8 Gold standard

In general, 'the Gold standard is a trustworthy corpus that is necessary for training and meaningful evaluation of algorithms which use annotation' (Wissler et al., 2014). In our research, the gold standard is trustworthy annotation for a certain aspect of a text that can be used to evaluate any automatic tool. It is not necessarily corpora; rather, it is another human annotation used as a reliable annotation to assess the automatic annotation tool.

The gold standard was annotated by me following the same guidelines given to the two annotators. I started with the overall sentiment analysis for all the reviews regardless of the presence of metaphors, but some short reviews were all metaphorical, so the sentiment was entirely driven by the metaphor. Moreover, as previously discussed, the LABR rating score was used for long reviews which could have multiple polarities. The second round of the gold standard annotation is to annotate metaphor expressions despite the overall polarity; even the short reviews driven by metaphor had a level of ambiguity. The LABR data scores were checked for short metaphors, and the annotation was different from the metaphor and the writer rating. This means that although a metaphor had negative polarity, the rating indicated the positivity of the expression. For example, دى مش رواية دى صدمة عصبية means *'This is not a novel, this is a neurogenic shock'*, which was assumed to be negative. However, the rating score for this review was five, which means that the writer wrote the review positively. So in such cases of ambiguity, I followed the LABR rating score.

4.5 Inter-Annotator Agreement (Artstein, 2017)

This section discusses IAA (Inter-Annotator Agreement), which is the main computational aspect of sentiment annotation, but the raw agreement will be calculated for other categories such as metaphor term selection, meaning and context. The Cohen's Kappa (Cohen, 1960) metric was used to calculate the IAA using the following formula:

$$\kappa = \frac{(p_o - p_e)}{(1 - p_e)} \tag{4.1}$$

Cohen's Kappa here was used for three categories, which are the three polarities: positive, negative and neutral. As already explained, the annotators were given guidelines in the form of written and verbal instructions. The two annotators selected online Arabic metaphors based on the criteria provided 4.3.1. The main criterion was to label new online Arabic metaphors, which is the main purpose of this research. However, sentences that had regular Arabic metaphors were labelled based on the concept of containing an online Arabic metaphor. For example, metaphors written in MSA, which should have the regular Arabic metaphor structure, have the online metaphor structure. For example, كحولية means literally *'alcoholic'*, metaphorically romantic 4.2 is one-word metaphor, which is one of the Arabic online metaphor forms in an online context. The tasks for the annotators were to label:

- the genre tag by choosing the genre type and the most indicative word from the review which identified the genre type. For example, the general genre frequently occurred with الرواية *'the novel'* and الكتاب *'the book'*.
- Metaphors were tagged based their identification as online Arabic metaphors. Precisely, the new Arabic terms used in the online context were similar to those used on social media, For example, بيض/*'eggs'* is one of the metaphors used on social media platforms. The metaphor type was specified as informal if the metaphor had an online structure and formal if it had the regular structure of Arabic metaphor.
- The literal meaning was tagged, which is the metaphor's hidden meaning, based on the closest literal meaning to the metaphor. We used the online dictionary called *al-maʔānī* `https://www.almaany.com/en/dict/ar-en/`here to ensure the familiarity of the word in the sentence's context as there is online metaphor words have no source as discuss in the previous chapter 3. Also, the meanings and dialects of informal Arabic metaphors were authenticated by a questionnaire 3.10. The responses were 107 of 100 informal Arabic metaphor. Some terms were not authenticated all as the data were updated later to remove redundancies. In addition, the annotators authenticated the meaning and the context, which is the main aim of this research. In addition, some of Those words can be used between the Arabic dialects. Even so, we authenticated the dialect of some of the Arabic online metaphors as a sample of the dataset to prove the dialects most used in the new corpus.
- The context was labelled based on the metaphor's location and the sentence length. For example, if a metaphor comes at the beginning of a sentence, the context type is after, and vice versa. If a metaphor comes in the middle, the context type was identified based on the logical order, so that defining the meaning of the sentence could help to interpret the metaphor.

The labelling process was discussed with the annotators based on the previous criteria. The resulting scores of the annotators showed an agreement of the data of 0.16, which is in the range of low reliability. The different annotators labelled the same dataset separately. The Cohen's Kappa IAA was calculated to measure whether the agreement was reliable. The observed agreement was the number of identical labels from the raters calculated as a percentage.

The high IAA score from the annotators' annotation is a proof of the clarity of the guidelines from Artstein (2017)' perspective. Artstein (2017) discussed the process of annotation from the perspective of the clarity of the guidelines. The study determined the relation between the clarity of the guidelines and the reliability of the data by using a repetitive process. This meant that every annotation process with unreliable data needed to be re-written or the guidelines needed to be clarified. However, this did not occur in our case as the data sample had 0.89 IAA, which meant that the guidelines were clarified. But the IAA for the whole of the corpus data changed to 0.16, which is only a slight agreement. The annotation process was affected by many factors. In our case, it is possible that the data were affected by the size of the dataset and the length of the reviews, as discussed before. This means that after we had calculated a bigger size with many metaphor cases, the annotation changed.

Also, the level of ambiguity of the online metaphors presented a problem. For example, some of the online terms could have two possible sentiments but no supporting context from which to clarify the sentiment term, as discussed 4.6 elsewhere in this chapter. An example is هيلحس دماغك which means *'will lick your brain'*, which could be either negative or positive. So one of the annotators could decide that the polarity is neutral, whereas the other might annotate the metaphor as positive. Moreover, the gold standard would regard the term as negative.

4.5.1 Metaphor raw agreement

As discussed above, the raw agreement, which is also called the observed agreement, is the number of compatible annotations between annotators. It was calculated using percentages. The rate of the identical selection of metaphors was 70%. The 30% of unmatched annotations had differences in the size of the metaphor expressions compared with the correct ones. For example, some of the unmatched selections of metaphors were only different in the number of words, but the metaphor was the same.

As an example, one annotator selected a metaphor as خطفت, which means *'kidnap'* whereas the other chose خطفت عقلي خطفا, which means *'kidnap my brain'*, which is the same metaphor but in the form of an expression, not a single word. In this example, there was a partial match between the annotators, but since the point was to have identical annotations, it was considered a mismatch. The annotations of both annotators were accurate. The first annotator considered the verb خطفت which means *'kidnap'* to be a metaphor as the target of this metaphor is the book, which is absent in the context. The other annotator considered that 'brain' would clarify the context and used a metaphor phrase to interpret the metaphor.

Another similar example of a partial match is فتح, which means *'opened'*from one annotator, whereas the other annotator's annotation was فتح لي عالم آخر/*'It opened up another world for me'*. Both annotations were accurate as the first annotator assumed that the verb فتح/*'opened'* referred to the book, which is the general domain for the dataset. The other annotator assumed that the verb فتح لي عالم أخر/*'It opened up another world for me'* was explainable more from the subsequent words, which were also a metaphor. Another case of raw disagreement is full mismatch; this occurs when both annotators make completely different metaphor choices for annotation. An example of a full mismatch is يخرج للنور/*'It comes to light'* from the first annotator and عقبة *'obstacle'* from the second annotator. The mismatch is the different metaphor choices from both annotators, which is 20%. The agreement percentage was calculated based on comparing the metaphor terms chosen by the two annotators.

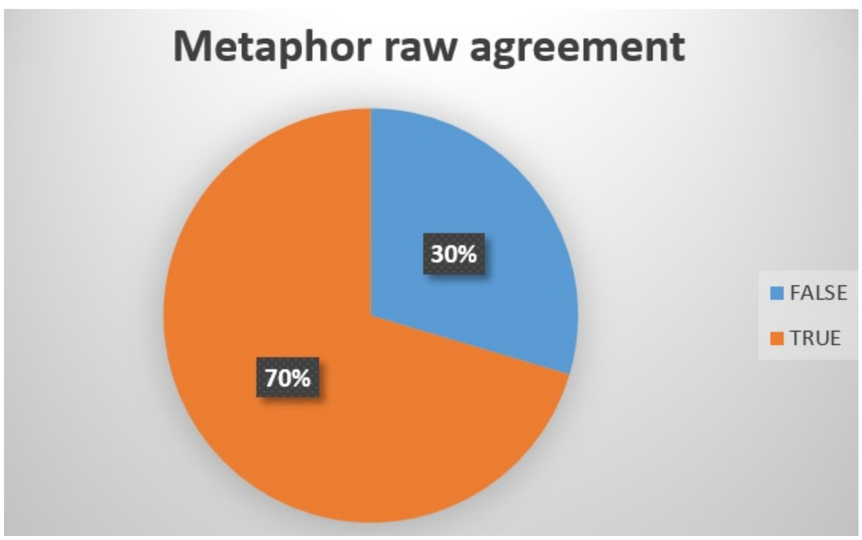


Figure 4.3: Metaphor raw agreement

4.5.2 Corpus statistics for manual annotation

This sub-section describes the general statistics of the Arabic online metaphor corpus. The statistics are about the size of the corpus, the length, the number of tokens, and the percentage of each sentiment category.

Statistics of Data Strcucture of Arabic Meatphor Corpus		
Category	**Value**	**Unit**
Size of corpus	1000	review (mostly sentences)
Language	Arabic	language
Fearture of corpus	manually annotated	sentence sentiment, metaphor (words/MWEs), metaphor sentiment, metaphor meaning, context of metaphor
Average sentence length	107.968	token
Longest review	1804	token
Shortest review	2	token
>=1000 tokens	5	token
<1000 toks and >=500 toks	32	token
<500 toks and >=100 toks	37	token
<100 toks and >=90 toks	24	token
<80 toks and >=70 toks	30	token
<70 toks and >=60 toks	31	token
<60 toks and >=50 toks	57	token
<50 toks and >=40 toks	53	token
<40 toks and >=30 toks	70	token
<30 toks and >=20 toks	99	token
<20 toks and >=10 toks	149	token
<10 toks and >=5 toks	90	token
<5 toks and >=1 toks	63	token
Positive text units	70%	review (mostly sentence)
Negative text units	17%	review (mostly sentence)
Neutral text units	13%	review (mostly sentence)
1-10 tokens	153	15%
10-20 tokens	169	17%
20-50 tokens	222	22%
50 -100 tokens	195	20%
100- 1000 tokens	69	7%
The total number of reviews	1000	

Table 4.1: Statistics of manually annotated corpus.

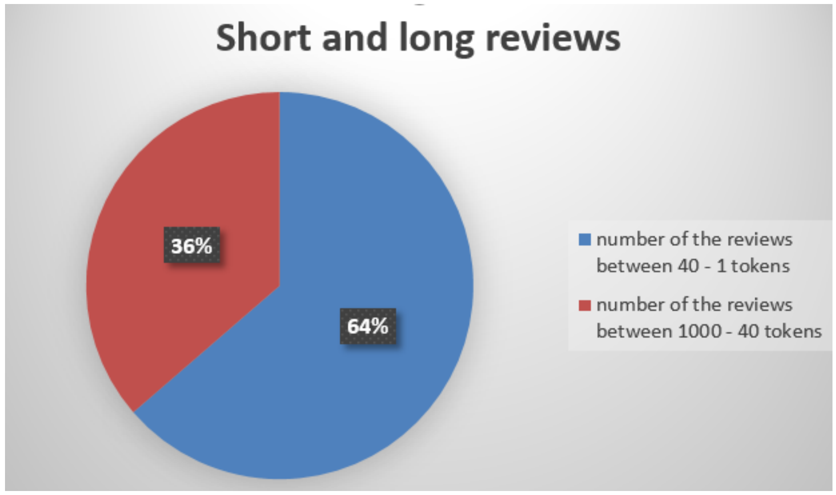


Figure 4.4: Manual annotation statistics

Table 4.1 describes the number of tokens for different ranges. The statistics show those numbers as the number of tokens, which are the lengths of the reviews. The categories are divided with different ranges of lengths as they may affect the sentiment decision because lengthy reviews have many factors which could change the sentiment for metaphor. The category between 1000 to 40 was considered lengthy reviews. The category below the previous category is short reviews. The corpus has a high number of short reviews. This means the approximate proportion of lengthy reviews was only 36% of the total number of reviews, whereas short reviews had 64% of the total number of reviews as shown in Figure 4.4. So, the corpus has more pure metaphoric reviews than the lengthy ones, which are affected by multiple factors.

The figure above shows the short and the long reviews as proportions of the total number of reviews. The smaller proportion of lengthy reviews could nevertheless affect sentiment annotation decisions because the lengthy reviews could contain multiple factors such as different sentiments, different metaphors, and different writing styles. However, most of the lengthy reviews had sentiment combined with metaphor, meaning that the main opinion of the book combined metaphor and sentiment. This example shows a lengthy review which contained only one main opinion which combined a metaphor and sentiment, which comes at the very beginning of the review, followed by a full stop. What is mentioned after the phrase is just clarification of the main opinion. Another example of a metaphor in the middle of the review but with a negative opinion is:

كمية التحشيش ف الرواية بوظت الرئة عندى!!

'The amount of weed in this novel run my lungs!!'

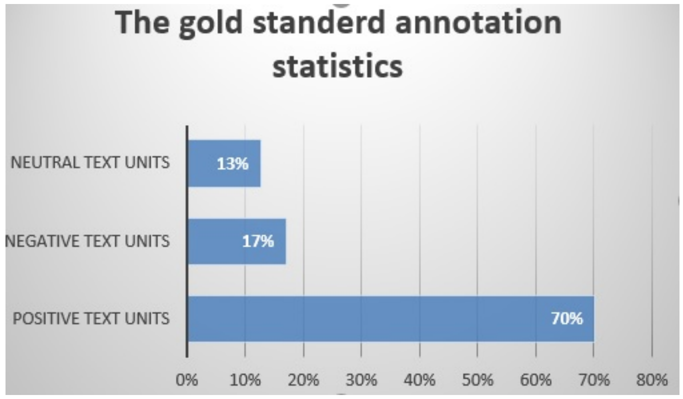


Figure 4.5: GS annotation statistics

In which the metaphor specified is بوظت الرئة, which means *'ruins my lungs'* written in dialect, which means *'painful'* or *'annoying'*. This example has multiple metaphors, but the annotators chose the ones which had a new structure and met our criteria for data annotation. There were some reviews with metaphors which were not combined with the main opinion about the book. The metaphor in this example was specified as تخرج شحنات, which means *'extract electrons',* which is an indication of extracting emotions. However, the metaphor, which is an opinion, in this review was not directly for the book being reviewed, but is an opinion about the book's content. The book content is not directly reviewing the book but rather mentioning the incidents that happened based on the book's story. For example, if the book talks about politics, the reviewer will give an opinion about the state of the Egyptian commoner.

This example is from a lengthy review, which was only a small proportion of the whole dataset. So the overall annotation could be affected by the length of the review. However, as mentioned, the proportion of lengthy reviews was few compared to the short ones. The longest review was 344 tokens and the shortest was two tokens. As discussed in the data collection chapter, lengthy reviews were chosen to balance the dataset. We aimed for the same five terms in different contexts and with different meanings and sentiments if possible as we were restricted to the availability of the terms in Aly and Atiya (2013). So we found some terms which would balance the dataset in the lengthy reviews. For example, the term فظيع means *'horrible'*, and it was chosen in five different contexts, one of which was long.

As was explained above, the term فظيع means *'horrible'* and was an important piece of evidence to demonstrate the change in the Arabic metaphor structure and sentiment. The term signified negativity, but it was used as a positive term in all the contexts that were chosen. So we chose the terms which proved the change in the Arabic metaphor and the sentiment. The proportion of the sentiment categories of the gold standard annotation in our Arabic metaphor corpus was calculated. The positive annotation of the reviews was 70% of the total, which means that most of the reviews had positive polarity. Negative and the neutral annotations correspondingly formed 30% of the total gold standard annotation.

4.6 Challenges to manual annotation

The annotation was done in three stages. The first stage was to annotate the metaphor terms in each review. The second stage was the sentiment annotation for the overall reviews and for the metaphor expressions. The third stage was for the meaning and the context annotation. In each stage, we faced challenges to manual annotation. As mentioned before, the data contained long and short reviews. The short reviews were driven by metaphor whereas the long reviews were driven by metaphor and other text factors. For example, there were some reviews with multiple metaphors, but they did not contribute to the main opinion of the book being reviewed. For example, there is a review with a subject related to a politics book and the reviewer gives an opinion about his country's current state. Such as دي مش بلد دي خرابة, which means *'this is not a country, this is a corral'* in a literal sense, which indicates corruption. So for the long reviews, we tended to choose those in which the main opinion was expressed as a metaphor.

The metaphor annotation was specified, as explained above, using guidelines. However, the metaphor was specified differently if there were two metaphors in the same sentence. For example, in كتاب دمه زي السكر, the first annotator chose زي السكر/*'like sugar'* and the other chose دمه/*'its blood'*. There was a difference in annotating the metaphors, but both metaphor terms chosen by the annotators were compatible with the selection guidelines. A case in which a metaphor occurs in the context of an annotated metaphor is called a 'nested metaphor'. Nested metaphors were annotated as context annotations without specifying the metaphor on the Excel sheet because each review contained one metaphor that specified the sentiment, whereas the other was hyperbole to emphasize the opinion.

The annotators were faced with many challenges during the overall sentiment annotation. The challenges were related to the level of ambiguity of Arabic online metaphor sentiment and meaning. For example, هيلحس دماغك, which means *'it will lick your brain'*, was interpreted as a negative expression by one of the annotators who interpreted it as a notion of confusion, whereas the other annotator annotated the expression as neutral. This means that the expression might be interpreted as a notion of fascination and frustration together. So the level of ambiguity encountered in annotating the sentiment was high, even for a short expression.

However, the phrase could certainly be expressed as a negative or positive expression if it is supported with negative or positive words. The expression in this case occurred as a single phrase with no supporting words. Since the meaning is ambiguous and there are no reliable resources, the meaning must be annotated. As already mentioned, the meaning had to be interpreted from the context because the same term had different meanings and sentiments according to the context. For example, يخربيت/*'May it be ruined'*can be either negative or positive in different contexts. Nested metaphors described above occurred more frequently as formal Arabic metaphors than informal. For example, in the review يربت على القلب فيصب فيه الامل, the terms يربت which means *'dabs'* and فيصب which means *'pours'* are metaphors used in the same context. An example of an informal metaphor in a nested case is كتاب دمه زي السكر. The first metaphor is دمه which means *'its blood'* and the second is زي السكر, which means *'like sugar'*, but the second metaphor is regarded as a simile because it follows a *'like'* word in Arabic dialect.

The example يخربيت الضحك اللي ضحكته means *'ruined the laugh that I laughed'*, which is positive, whereas يخربيت السخافة means *'God ruined the silliness'*, which is negative. The meaning annotated as جميل means *'beautiful'* and تافة means *'silly'*. So the meaning for the same metaphor term annotation was annotated for each word in each context because of the different notions. The Arabic online metaphor had a level of ambiguity in regard to specifying the sentiment, even for the Arabic native speakers, so the annotators faced difficulties in specifying the sentiment. We found that the meaning played a role in specifying and understanding online Arabic metaphors because there is no reliable resource for Arabic online metaphors to give their meaning, and we found during the annotation that the meaning is necessary for understanding a metaphor's context to identify the sentiment.

I asked the annotators to specify the meaning for the metaphorical terms within the context. For example, جدا ع الجرح means *'too on point '*, which was annotated as negative as the meaning is not clear. The lack of clarity in this term was represented in the term despite the context, which means the term in our dataset was followed by a sad Emoji, which could indicate sadness, whereas the phrase implies that the book affects readers in terms of things which they wanted to reveal and discuss. The same term occurs as مقالات ع الجرح /*'articles on the wound'*, which was followed by a winking Emoji, which could indicate that the review is 'what I want to reveal too' and is positive. So the context plays an important role in identifying the meaning and the sentiment of the metaphor. Although, as mentioned before, there are some Arabic online metaphor terms with no context which had an obvious meaning and sentiment if the domain is known, the annotation could be changed based on the context. The term is metaphorical; however, the term in its context could change both the meaning and the sentiment. As has already been discussed, the same term can occur in different contexts with a different meaning and sentiment. For example, the review مقالات ع الجرح which means *'Articles on the wound'* had the same metaphor, and the meaning was annotated as *'painful'*, which indicates the negativity of the expression.

In terms of meaning and writing style, the typographical and the parsing are two of the challenges in understanding a review for annotation. For instance,

تاريخ غفلنا عنه ودوّن
في كتب التاريخ باختصار
قتل وجعه وجماله

'A history we have neglected and yet it has been recorded'
'In history books, in short.'
'It killed its pain and its beauty.'

In this example, the metaphor specified is قتل means *'kill'*, but the context words had to be considered in order to understand the metaphor. The word دوّن *'written'* in the review could be understood as different meanings as the word was parsed with آ.The annotators interpreted the word based on the context, meaning the previous and the following words. Also, elongation in Arabic online metaphor terms is typical in online writing. It does not affect the annotation, but it does affect the data collection, and it acts as an emphasis of an opinion. The annotation for the previous example was as follows:

- The metaphor is قتل which means *'kill'*
- The metaphor was categorized as Verb
- The metaphor type is formal and the metaphor's part of speech was specified as VB, signifying a verb.

```
<s no="2"> تاريخ غفلنا عنه ودوّن في كتب التاريخ باختصار قتل وجعه وجماله <genr
        >كتب</genreT>
    <metaphor Type="Verb">
        <Formal>قتل</Formal>
        <Tword Type="VB">قتل</Tword>
    </metaphor>
    <literal>
        <Tword Type="VB">مسح</Tword>
    </literal>
    <context Type="after">
        <Tword Type="JJ">وجع</Tword>
        <Tword Type="PRP">ه</Tword>
        <Tword Type="CC">و</Tword>
        <Tword Type="JJ">جمال</Tword>
        <Tword Type="PRP">ه</Tword>
    </context>
</s>
```

Figure 4.6: XML Annotation

- The literal meaning specified as مسح means *'erase'* and is a verb.

- The context words are specified after the metaphor term وجعه و جماله, which means *'Its pain and its beauty'* that the metaphor was clarified as a metaphorical term.

Online communication has a spontaneous writing style and the location of the punctuation affects the meaning of a metaphor. For example, if a metaphor term is followed by a full stop, that means that the opinion has been stated. But sometimes the context words written after a metaphor can support the metaphor to clarify the sentiment and meaning even though there are cases where the supporting words and the metaphor have no obvious opinion. For example, in بيييييييض. من وجهة نظري يعم which means *'eggs from my perspective'*, the metaphor is the first word بيييييييض and the full stop comes after it, which means that *'eggs'* is the main opinion, whereas the supporting context, which is *'from my perspective'*, has no obvious sentiment of the metaphor. The punctuation was left in the corpus to show the online Arabic metaphor structure and style of writing. For instance, in the previous example, the Arabic online metaphor could occur as a single expression only to express an opinion. The term *'eggs'* or بيض was found alone during the data collection, but we avoided such cases as they might have made the metaphor identification process more arduous as the metaphor is most likely to be identified from the context or the surrounding words. So we left such cases out of the annotation.

In the *'illness'* category, a phrase such as صدمه عصبيه *'Neurogenic shock'* is an online Arabic metaphor term which has no supporting context to specify the sentiment accurately. The sentiment was therefore scored during the annotation decision. In a sentence assumed to be a negative opinion by Farha and Magdy (2019), the sentence score is five, which is positive, whereas the two annotators assumed that it was negative and neutral, respectively. Another example is هيلحس دماغك, which means *'will lick your brain'*, which was annotated by one of the annotators as neutral, assuming that the novel was likely to challenge your brain or confuse you, whereas the other annotator assumed that it was positive. I assumed that it was negative as an indication of confusion. In the LABR, the rating for this review is three, which means that the reviewer rated the book with a bit of positivity despite the negativity in his review. Moreover, clearly the text is negative if we consider this as textual analysis, but for a book review, it is neutral.

From these examples, it can be seen that Arabic online metaphors can be highly ambiguous considering the many factors which can affect them. The challenges lie not only in annotating the metaphor's meaning, sentiment and context, but also in identifying the online writing style and the means of communication. Annotation is therefore challenging for such metaphors. Also, such contradictions in understanding Arabic online metaphor describe a fair IAA agreement between the annotators, but not the raw agreement.

4.7 Discussion

The annotation examples show the changing structure of the Arabic metaphor in the online context. There are many irregular cases of informal metaphors. For instance, informal metaphorical terms written in dialectal Arabic occur as one term, and more likely come as metaphors in the online context in the book domain even though an informal term could be used in a literal context. For example, بيض means *'eggs'* annotated as an informal metaphor and logically known as a metaphor in the online context, whereas there are particular Arabic informal terms which could hold both literal and metaphorical meanings, but are considered as metaphors if they are in dialectal Arabic in the online context. For example, جامده *'solid'* is one of the metaphorical dialectal terms that can be literal in MSA.

As well as regular cases of the formal use such as proverbs and sequential metaphorical terms, we discussed and modified the annotation to fit each case. When there are no different cases of the formal metaphor, we can deduce that the online metaphor in our dataset comes in two forms, a stand-alone expression and a term with dependent words. However, proverbs are annotated as expressions. For example, the proverb السم في العسل means *'The poison in the honey.'* was annotated as a proverb as proverbs have a fixed structure and the meaning is in the story behind the proverb. However, known proverbs could be annotated more easily than new ones. For example, there are new proverbs written in Arabic dialect which is وجع البطن ولا رمي الطبيخ, which translates in literal sense as *'a pain in the tummy is better than throwing the food'*; an example of a new Arabic dialectal proverb, which means 'pain'.

A stand-alone expression is a contemporary created term (an informal metaphor) used in online communication. A term is described as 'stand-alone' if it has no context words or dependent phrases from which to identify it as a metaphor. It is usually followed by a full stop if it comes with a context, and that was considered in the new dataset. For instance, تحفة/ *'vase'* and بيض/ *'eggs'* are examples of the stand-alone metaphor type. A stand-alone metaphor can occur in MSA as well as dialectal Arabic. For example, the metaphor term كحولية/*'alcoholic'* is written in MSA and بيض *'eggs'* is written in dialectal Arabic.

The target, which is equivalent to the subject, in an Arabic metaphor is omitted even though it has a slight change in the structure when it occurs in the online context. When a metaphor comes as a phrase, the annotation should merge the two terms together in some cases. For example, مالوش حل means *'has no solution'*; it comes as dialectal and is expressed in two words combined. However, the term يلحس دماغك/*'lick your brain'* comes in two words but it can be clear if the phrase comes as one word. For example, the word يلحس is a regular verb, but the metaphor depends on the illogical context, which is دماغك, which means *'your brain'*.

As already explained, the corpus building was influenced by VU Amsterdam's work, but the annotation schema was modified to fit the computational aspect of the present study because the main purpose of the VU Amsterdam corpus Krennmayr and Steen (2017) was to build an English metaphor resource for researchers in linguistics, whereas the current study was designed for use in the computational aspect for the analysis of the linguistic field for online Arabic metaphors. Also, as previously discussed, the VU Amsterdam annotates all the words of a sentence whether they are metaphorical or not. In addition, each word is assigned to a lemma by its part of speech, whereas our purpose was to annotate metaphor terms from their context for metaphor identification. The annotation reveals the data analysis of the Arabic online metaphor as the data analysis is essential for building an accurate tool. It should be noted that Oxygen, which is an XML editor application, was used to annotate the data for the VU Amsterdam corpus.

Our schema design was based on the computational requirements to meet the practical aim of this study of showing the impact of the online Arabic metaphor on sentiment. For example, the context before and after a metaphor term was specified to define an Arabic metaphor. The assumption of the computational aspect is that the metaphor would affect the sentiment annotation using different methods.

4.8 Results

The Arabic metaphor corpus resulted from the annotation described in this chapter. As mentioned before, the Arabic metaphor corpus started as XML annotation of the metaphor expression with metaphor type, hidden meaning, and context. Each tag had its own specification for annotation. The corpus was annotated for metaphor, sentiment, meaning, and context. Also, the theme and metaphor type were specified. However, as already explained, the data were converted into Excel sheets to meet the computational aspect of this study and also because of the inconsistency of the Arabic online metaphor structure. However, the Arabic metaphor corpus annotation reveals the Arabic online metaphor structure. Even though the structure is not static in the online context. The overall structure has nevertheless been discussed based on the frequency pattern of the Arabic online metaphor.

The overall manual sentiment annotation showed high raw agreement between annotators, but high raw agreement is not an indication of the validity of the corpus. The observed agreement showed 51% compatibility of the overall sentiment category from the two agreements. As discussed above, IAAs were calculated for the overall sentiment agreement and showed a low agreement level. A low level of agreement is a common case in linguistic annotation. Also, it indicates that the annotation task was hard (check the citation). However, since our purpose was to annotate for practical use, which was stated in the discussion of practical application of the data as not necessary for successful machine learning (Artstein, 2017), it was discussed that the annotation reliability does not imply the sufficiency of the data for practical purpose. Although the unreliability of the Arabic metaphor corpus implies a high level of ambiguity of the Arabic metaphor, a base knowledge of Arabic online metaphor is necessary.

4.8.1 Extracted features

The pattern extraction is one of the results of the Arabic Metaphor Corpus (AMC). These patterns help define Arabic online metaphors by revealing the surrounding words and their structure. The features extracted from the AMC are used to determine sentiment and identify metaphors, focusing on polarity and target/context words. These features can be employed for further investigation into Arabic metaphors. Our primary objective is to assess the impact of the AMC on automatic Arabic sentiment analysis. Further suggestions for future work are discussed in 6.1.

- The metaphor can be identified using the polarity, and the algorithm will learn from the annotated AMC with sentiment. The terms associated with contradictory polarity (positive and negative words) ممتعه لدرجة مرعبه mean *'joyful till terrifying'* in a literal sense; metaphorically, it is so joyful, it is more likely to be positive and metaphorical. Because the term (for example, مرعبة means *'scary'* is negative in the literal sense, when it is associated with a positive word, it is more likely to be expressed as negative and metaphorical. Another example, رائعة حد الثمالة means *'fabulous till intoxication'* So, the metaphor comes from the negative word associated with the positive word. In a literal context, the same term comes with positive words (ex مرعبة و رائعة)/ *'terrifying and wonderful'*. It is more likely to be positive and literal. We put this assumption as the word مرعبة *'terrifying'* usually comes in a literal sense to describe a negative incident. But when it is associated with positive words, it is more likely to be metaphoric. In addition, the previous example has *'till'*, which often comes with a metaphorical word in the AMC. However, this discusses the Arabic metaphor in an online context, where the text is unpredictable and changeable, as discussed in the previous chapters. So, it could not apply to all social media/online text.

- In addition, the metaphor can be identified using the target words and the contexts following the feature below. The AMC annotated with the context of the metaphorical words. The contexts were specified as a definition of the metaphorical words. So, the contexts specified after or before the metaphorical terms. The contexts can be compared with the same term in different contexts. However, this won't detect the semantic meaning of the sentences. For example, morphology affects the similarity even when the two words are similar in meaning.

- Where the target words same with different contexts.
- Where the target words (different form but same meaning) with different contexts. In this condition, the meaning annotation could be used to spot the term with the same meaning.
- Where the target words similar (different morphology) with similar contexts.
- Where the target words same with the same context. This condition produce same classification for the polarity and the if it is metaphor or not metaphor.

Informal	Category	Translation literal	Translation metaphoric
تحفه	Thing	Vase	fantastic
جامده\ جامد		solid	wonderful
الألش		no source	lie
بيض		Egg	bad
اشطا\ قشطه		cream	lovely
لذيذ\ لذيذة		delicious	beautiful
عسل		honey	beautiful mean in negative context
سكر\ مسكر		sugar or sugary	nice
شهي	Food terms	delicious	entertaining
نكهه		flavor	style
طزاجة		freshness	good
وجبة		meal	information
بصل		onion	bad
مذاقاُ		taste	style
مقبلات		appetizers	additions
دسم		fatty	rich information
البهارات		spices	extra information
الطبخة		the recipe	the story
بطيخ		watermelon	nonsense
خلطة		mixture	the content
طعم		taste	experience
فظيع\ فظيعة		horrible	wonderful
مخيف		scary	majestic
قاتلة		killer	fantastic
خيالي		imaginary	magical
خورافي		legendary	fabulous

Dramatize

اجرام\مجرم فقيع فشيخة\افشخ\فشخ مرعبة فتاك موت جنان خطير قصة جنونية خوقاق\خوقاقي		criminal no source(bubbling) no source scary deadly death heavens Dangerous story crazy no source	 exaggerate beautiful attractive fantastic wonderful amazing fantastic creative rejected wonderful
علاج جرعة مضاد كبسولة تخدّر	Medical	cure dose antibiotic capsule numb	comforting energetic comforting comfortable prevent
يجيبلك مغص مريضة صرع تخلف عقلي ع الجرح بدوخة ملعون ابو الصدق اكتئاب للجرح مجنونة هلس	Illness:physical and mental	bring stomachache sick epilepsy mental retardation on the wound dizziness cursed truth depression Wound crazy Hallucination	painful bad confusion depression on point beautiful despise sad pain miracle empty
سلاح	Weapon	cure	

قنبلة			
دماغ	Personification	brain	smart, genius
حد النخاع		to the bone marrow	very
بوشين		two faced	hypocrite
سكرى\ تسكر	Drugs term	intoxicated	delightful
تحشيش		weeding	nonsense
مخدرات		drugs	tranquilizers
كحولية		alcoholic	romantic
الثمالة		drunk	very
خرا	Offensive	shit	bad
زبالة		trash	bad
زفت			
خرابة			
أكلت	Verbs	eat	affect
يفطس		makes you die	very
تسبح		swim	daydreaming
القاتلة		killer	strong
تلتهم		devour	finish the book quickly
يجيبلك		brings you	brings pain
خضني		shake me	scares me
زغزغني		tickle me	makes me laugh
بتطبطب		dabbing	compassionate
بتسحب		pull	ugly
تحفر		sculptured	affected me
يخرب بيت		damaged house	demolish
هيلحس		lick	surprise
بتطلعني		takes me	

تشدك		pull you	
سوداء		black	sad
خفيف		light	simple
مضروبة	Adjective	hit	bad
نسمة		breeze	short,light
خنيقة		suffocated	ugly
ممطوطه		stretched	long
غسيل مخ !		brain wash	clean
خبط لزق		hit and past	
الرغي الرغي		gossip gossip	
غسيل مخ		brain wash	affected
دماغه عالية	compound	high brain	disregard
خفيف دم		light blood	funny
شهادتى مجروحة		injured testimony	incomplete
ثرثرة مجانية		free gossip	just words
ضحك السنين!!!!		years' laugh	funny
سهلة الهضم		easy to digest	likeable
حد الموت		till death	very
كتاب ملحد	Religious	Atheist	the author

Table 4.2: A sample long table.

Chapter 5

Impact of Metaphors on Sentiment Detection

5.1 Why not using deep learning and large language model?

Based on the technical aspect of LLMs and deep learning: This research addresses a relatively new and underdeveloped area in Arabic natural language processing, which is Arabic online metaphor and its impact on sentiment. To address such a new Arabic online metaphor, pre-annotation is essential for this task. The Arabic metaphor in the online context is new in terms of structure, meaning, sentiment, type, and even includes some that are invented specifically for this context.

Given the lack of existing tools and datasets, it is nearly impossible to advance in this field without a large language model (LLM) specifically built for Arabic metaphors. However, an LLM built for Arabic metaphor cannot be applied directly. This is because the first step in identifying Arabic online metaphor is to develop a word embedding trained exclusively on Arabic online metaphor. Such a model would enable machines to learn and predict sentiment from metaphorical expressions across a large volume of text. Since no such resource currently exists, this remains impossible. In addition, since deep learning depends on LLMs, the application is not accurate either.

Based on the research aim: This research started from the ground up, constructing a specialized corpus for this purpose. Developing the Arabic Metaphor Corpus (AMC) with the required specifications took nearly three years of my PhD. The data and analysis were compiled into structured Excel tables. The central aim was to examine the influence of metaphors on sentiment. Achieving this required building the AMC and developing automatic tools to test the hypothesis. I created two tools in collaboration with the Arabic Semantic Tagger El-Haj et al. (2022), both of which yielded promising results. Although the tools have been published and accepted Alsiyat, 2025, current Arabic large language models lack the ability to interpret metaphors, necessitating prior annotation—a task addressed through AMC.

Corpus construction in this domain is a demanding and time-intensive process, especially given the absence of existing Arabic metaphor corpora or annotation schemas. Thus, the AMC was structured based on key interpretive elements in Arabic online metaphors: metaphor term, context, theme, meaning, type, and semantics. These dimensions are crucial for developing future detection methods. For comparison, the VU Amsterdam Metaphor Corpus took five years and five authors to build, focusing only on identifying metaphors in sentences—without analysis. In contrast, the AMC provides both identification and analytical insight, aiming to demonstrate the impact of metaphor on sentiment.

The only previous Arabic study applied the LSTM method for binary metaphor classification, excluding sentiment analysis. Incorporating sentiment into such models remains challenging. Moreover, the existing tool's metaphor classification accuracy is limited. In contrast, my tool aims to classify sentiment derived from metaphorical expressions using semantic tags— which is promising to identify metaphor without requiring manual human pre-annotation. Although still limited to AMC and in its early stages, it shows strong potential to become the first tool for sentiment classification of Arabic metaphors using semantic tagging.

In English-language research, only recently have deep learning approaches been adopted for metaphor detection, relying heavily on well-established resources like WordNet, VerbNet, and SentiFig. These primarily support metaphor identification, not sentiment analysis (Wilks et al., 2013). More recently, English studies have begun employing large, pre-annotated datasets with LSTM for combined metaphor and sentiment analysis. In contrast, my research introduces a streamlined method that avoids human annotation by relying solely on semantic tags—offering a promising direction for automatic metaphor and sentiment detection in Arabic, where no such tool currently exists.

Deep learning methods require substantial data to achieve high accuracy. For Arabic, even a large word embedding model would not suffice without prior annotation or interpretation. As such, AMC plays a crucial role as a foundational resource. The corpus took three years to develop, incorporating detailed analysis and tool design. While the tool is still undergoing optimization and not yet highly accurate due to time constraints, it represents a strong initial step toward Arabic metaphor and sentiment classification.

This chapter discusses the testing of AMC using different automatic sentiment annotation methods to detect sentiment in Arabic book reviews. The Arabic Semantic Tagger was used to annotate the semantic senses of AMC entries, offering a potential pathway for future metaphor identification through semantic information. Additionally, the three tested sentiment detection methods were evaluated using standard performance metrics, such as accuracy, precision, recall, and F-score, to determine their effectiveness.

We describe the experiments conducted with AMC, with each phase explained through flowcharts (see Figure 5.4). The experimental process included applying AMC data using the Mazajak Arabic Sentiment Analyzer to explore the effect of metaphor on sentiment. Other sentiment analyzers were also tested and compared against AMC's gold standard (GS) annotations. Given the scarcity of web-based Arabic sentiment tools—many of which lack local deployment capabilities, such as El-Masri et al. (2017)—we utilized the Arabic Semantic Tagger (AraSAS) to facilitate sentiment classification.

The results indicate that AraSAS shows potential for identifying Arabic metaphors. However, due to time constraints, further development will be proposed as future work. The experimental tool designed in this thesis demonstrated the capacity of the semantic tagger to classify sentiment accurately, using emotion-based tags. Neutral tags were excluded due to complexity in certain reviews, as discussed in this chapter. The performance of each annotation method was compared with the gold standard annotations, and these comparisons were further supported by statistical analysis. Selected manually annotated examples are included for context, with a detailed explanation of how the gold standard was used to evaluate the automatic sentiment classifiers.

5.2 The Gold standard

This annotation was conducted based on manual annotation. For the manual annotation in general, two Arabic native speakers were engaged to annotate the data set. However, the gold standard was annotated by me. The gold standard is essential for the automatic annotation. The methods below included the gold standard annotation, which was used to assess the overall sentiment from the AMC reviews. The discussion below is about the gold standard annotation process, annotation challenges, and annotation cases. The discussion compared the gold standard annotation with Mazajak to show the differences in sentiment annotation between the automatic and manual.

We divided the gold standard annotation based on the length of the reviews. So, the AMC corpus has overall sentiment gold standard annotation and metaphor gold standard for the sentiment for the metaphor section. However, in the data collected for the corpus, some reviews had only one sentence with two words forming a metaphor, so the sentiment was judged entirely based on the metaphor. The two annotators worked separately.

I compared the gold standard annotation to Mazajak. Because there are cases where the metaphor could be interpreted as neutral using Mazajak. While manual it is positive. So, we compare to clarify those cases. For example, رواية عربية حتى النخاع means *'it is an Arabic novel to the core'*, and the expression could be positive or negative based on the context. The Mazajak tool showed a different percent of compatibility with GS than AraSAS. For example, comparing the overall sentiment annotation with the Mazajak annotation showed that the compatibility of the annotation from both was 49%, whereas the 51% of the Mazajak annotation was not compatible with the gold standard.

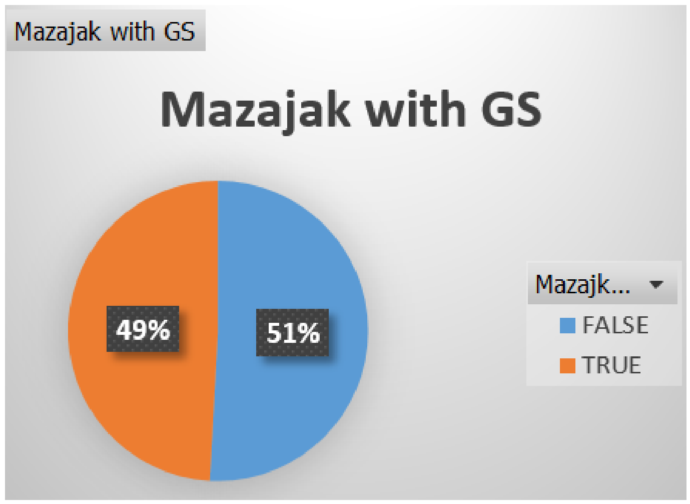


Figure 5.1: Mazajak GS annotation

The annotation started with specifying the sentiment orientation for each review by reading the whole review. Next, the sentiment for the metaphor section was specified in the context, except for short reviews. Due to the absence of additional opinionated words that are often present in longer reviews, short reviews may have fewer words that influence the sentiment decision. For some of the long reviews, the LARB Aly and Atiya (2013) rating scores of the reviews were followed to specify the sentiment accurately. In addition, some of the short reviews were found to have a level of ambiguity or were new to the Arabic language. For example, رواية عربية حتى النخاع means *'it is an Arabic novel to the core'*, and the expression could be positive or negative based on the context. However, the sentence has no indication or supporting context for negativity or positivity, so the sentiment was specified as neutral. The score from LARB, however, was 4, which indicated the positivity of the sentence. Because long sentences had mixed polarity of different book aspects even for subjects related to the book, I followed the LARB score for some of the reviews for overall sentiment. Also, I checked the LARB score for the metaphoric reviews (short) that do not have any supporting context.

For some of the lengthy reviews which discussed a book's pros and cons were annotated using the Aly and Atiya (2013) data score, but the metaphor section was annotated based on the understanding of the term's context. Contradictory sentiments for metaphors were specified based on the terms which met our criteria. For example, the term خنقت /*'suffocated'* ترسم خيالك/*'draw your imagination'* has both negative and positive contexts, and ترسم خيالك*'draw your imagination'* was the chosen term as it described the impact of the novel on the reviewer and was therefore a direct opinion about the book, which was one of our main criteria.

For some of the metaphor terms annotated as neutral, the metaphors in their contexts were aesthetic writing in order for the reviewer to express an opinion. In addition, some of the extra metaphorical expressions occurred as hyperbole and referred to subjects related to the book. As mentioned above, some of the sentences appeared to be and were understood as negative, whereas from the LARB score they were positive. Usually, such sentences describe the book content in the form of interaction. The Mazajak sentiment analyzer annotated these sentences as negative because of the absence of any aspect annotation of the sentence. Identifying the domain of the sentence can therefore affect the accuracy of the annotation.

One of the challenges for annotation we encountered was the parsing of a metaphorical expression. Undoubtedly, meaning is one of the basic requirements for accurate annotation, but even though parsing is not necessary for Arabic native speakers to understand a sentence, the online context with spelling errors can nevertheless be well understood in an Arabic text, not to mention the level of Arabic metaphor ambiguity in the online context. For example, in تركتني بعدك سكرى/*'you left me behind drunk'* in the metaphor سكرى/*'drunk'* the word سكارى/ *'drunks'* assumes that there is a spelling error. Also, سكرى could be understood as سًى by parsing, even though it is rarely written in this form in formal writing. In another example, مش طبيعى meaning *'not normal'*, which can be understood as negative, whereas it is actually an indication of a positive opinion. The previous term is a metaphorical term which often occurs as positive if it comes in a metaphorical context. The pattern of the expression can come with a positive context, for example, prefixed with عبقرية, which means *'brilliant'*.

The rating score shifts the polarity and the metaphor. For example, a sentence with a negative critique of a book's content has a positive rating because the reader/reviewer enjoyed the book which was being criticized. This concept is not, however, applicable for all sentences. Because some negative critiques have a negative rating and in a sentence in which a metaphor shifts the polarity, the sentence can still have a positive orientation. For example, a sentence can start with عمل ادبي رائع الوصف به فوق الوصف يبعث العمل على الحزن و الحقد means *'A literary work with wonderful description, beyond description, that sends sorrow and hatred'*, which is positive, and end with a negative critique from the metaphor affect Mazajak annotation Farha and Magdy (2019), and can be interpreted automatically to be negative.

Metaphor can drive the polarity. For example, during the annotation, sentences with an overall neutral sentiment were annotated as negative by Mazajak (Farha and Magdy, 2019). An example is هيدخل دماغي, which means *'will enter my brain'* and the term was annotated manually for the metaphorical section as negative, and Mazajak saw the overall review as negative. If the overall review was manually annotated as neutral, this means that the automatic annotation system saw the metaphor and predicted the sentiment based on the metaphor. Even though this possibility is not applicable to all reviews, there were nevertheless many similar cases found in our corpus. Also, this could be considered as the effectiveness of metaphor in the overall sentiment. In addition, the rating in this lengthy review is 3, which means that the reviewer has a neutral attitude towards this book. But as mentioned above, the rating is not necessarily considered a precise polarity as the reviewers could judge multiple aspects of a book. So the rating is used to give a clue about the reviewer's attitude towards the book because there are sentences with a positive attitude but are rated as neutral.

5.3 Methods

The methods were divided into two main subsections: methods that did not consider metaphor annotation and methods that did. Although metaphor information was utilized in all methods, some methods specifically addressed the metaphor annotation, and tools were designed based solely on metaphor manual annotation. Consequently, the methods were categorized accordingly.

5.3.1 Methods for sentiment annotation without metaphors

During sentiment annotation, the annotation was divided into the overall sentiment orientation of the review and the sentiment determined by the metaphorical expressions. In addition, I tested the automatic sentiment detection as sentiment classification without considering metaphor. As the automatic tools are not designed to detect metaphor prior to the sentiment classification, classification with only metaphors is one of the methods under test. For example, the sentiment annotation for the metaphor section and the metaphor classification using the sentiment score and the GS metaphor classification.

So I considered one method with metaphor and another method without metaphor for sentiment identification, even though both are considered as metaphoric because the AMC has long and short metaphorical reviews. But the short reviews are affected by too many factors to be identified as metaphoric. This means that some reviews are purely metaphorical, which are the short reviews and which comprise most of our AMC (64%). So most of our reviews from the AMC corpus were purely metaphorical. I divided them in this way as I focused on the metaphorical section in my work. So we tested methods which directly deal with metaphor for metaphor sentiment classification, which are the ones designed to predict the overall sentiment. In addition, there were those without considering metaphor and do not directly consider metaphor in sentiment classification.

5.3.1.1 Semantic tagger based sentiment detection

The Arabic semantic tagger (AraSAS) is an automatic tool for tagging Arabic text with semantic categories. This tool has a user-friendly interface with a dialog box to paste and reset the Arabic text. AraSAS has sub-categories for each tag with different polarity signs to indicate the polarity type of the sentiment. For example, the signs (+), (++) and (+++) all indicate the positive semantic type and (-), (–) and (—) are all negative semantic types. But they do not show the sentiment strength. When there is more than one polarity sign, it means that the level of sentiment is higher than when there is only one. This explains the sentiment side of the semantic sub-categories regardless of the semantic meanings. In addition, the tagger has no sentiment classification function for the tagged text. The AraSAS 0.2 El-Haj et al. (2022) is the updated version of the AraSAS.

The AraSaS was used in our methodology to tag the online AMC. Each one hundred reviews out of the one thousand total were tagged independently as the tagger cannot handle tagging more than a hundred texts at once through the web interface. We therefore put them into an Excel file to use for the classification coding. We ran the AraSAS for the AMC to tag each review in the Excel file. The AraSAS treats full stops and exclamation marks as starting a new line in a new row text, so they were replaced with English letters to avoid breaking text incorrectly.

5.3.1.2 Over all sentiment classification using AraSAS

The function written to classify the sentiment was based on the semantically tagged ACM dataset. The tagged dataset was turned into a data frame using Python. The function was used to classify sentiment by counting the number of positive and negative emotional tags (E tags only). Each review was subjected to this function to calculate the sentiment score. The sentiment score and the polarity were regarded as a dictionary and served as an argument to the count function, which were the emotional tags without any sign of polarity.

The function was designed to not calculate the neutral E-tags in order to avoid the redundancy of counting similar tags. For example, if E1 is added to the list, the counter regards E1 and E1+ as a redundancy or as the same. In addition, there were sentences in which the numbers of negative and positive tags were equal, and in some reviews, the numbers of neutral, positive, and negative tags were all equal. This means that the review had the same amount of polarities. For example, when the emotional tags counted as positive, negative, and neutral were equal, the neutral ones were removed to avoid any complications because my aim was to test the differences between the methods and to test the AMC using this method. I therefore deleted the neutral tags from the list of emotional tags so that only the negative and positive tags were defined beforehand. I will evaluate these methods further in the Conclusion chapter as suggestions for further research because, for better sentiment classification in regard to metaphor, all the polarity signs should be considered in the classification.

The calculation of the sentiment score was based on a set of conditions after linking the emotional tag with the polarity. So for a positive emotional tag (item) from the predefined list, we added 0.5 to the count, and the opposite for the negative emotional tags, by subtracting 0.5. For a double positive E++ the function adds 1 to the count, and for double negative signs (–) the function subtracts 1 from the count. For triple positive emotional tags (E+++) the function adds 1.5, and for triple negatives (E—) it subtracts 1.5. These operations were carried out for each review, and in this way, the summation of the sentiment score was calculated. Finally, sentiment was classified based on the sentiment score. If the sentiment score is greater than zero, the polarity is positive; if it is less than zero, the polarity is negative. Otherwise, the polarity is neutral.

```python
def tag_count (text, tag, weight):
    count = 0
    tag_len = len(tag)
    for i in range(len(text)):
        if text[i : i + tag_len] == tag:
            count += weight
    return count
```

Figure 5.2: Sentiment classification using AraSAS

```python
def sentiment_score(text):
    tag_list = ['E1','E2','E3','E4.1','E4.2','E5','E6']
    tag_dict = { 0.5:'+ ', -0.5:'- ', 1:'++ ', -1:'-- ', 1.5:'+++ ', -1.5:'--- '}
    score = 0

    for t in tag_list:
        for w, s in tag_dict.items():
            tag = t + s
            score += tag_count (text, tag, w)
    return score
```

Figure 5.3: Sentiment classification using AraSAS

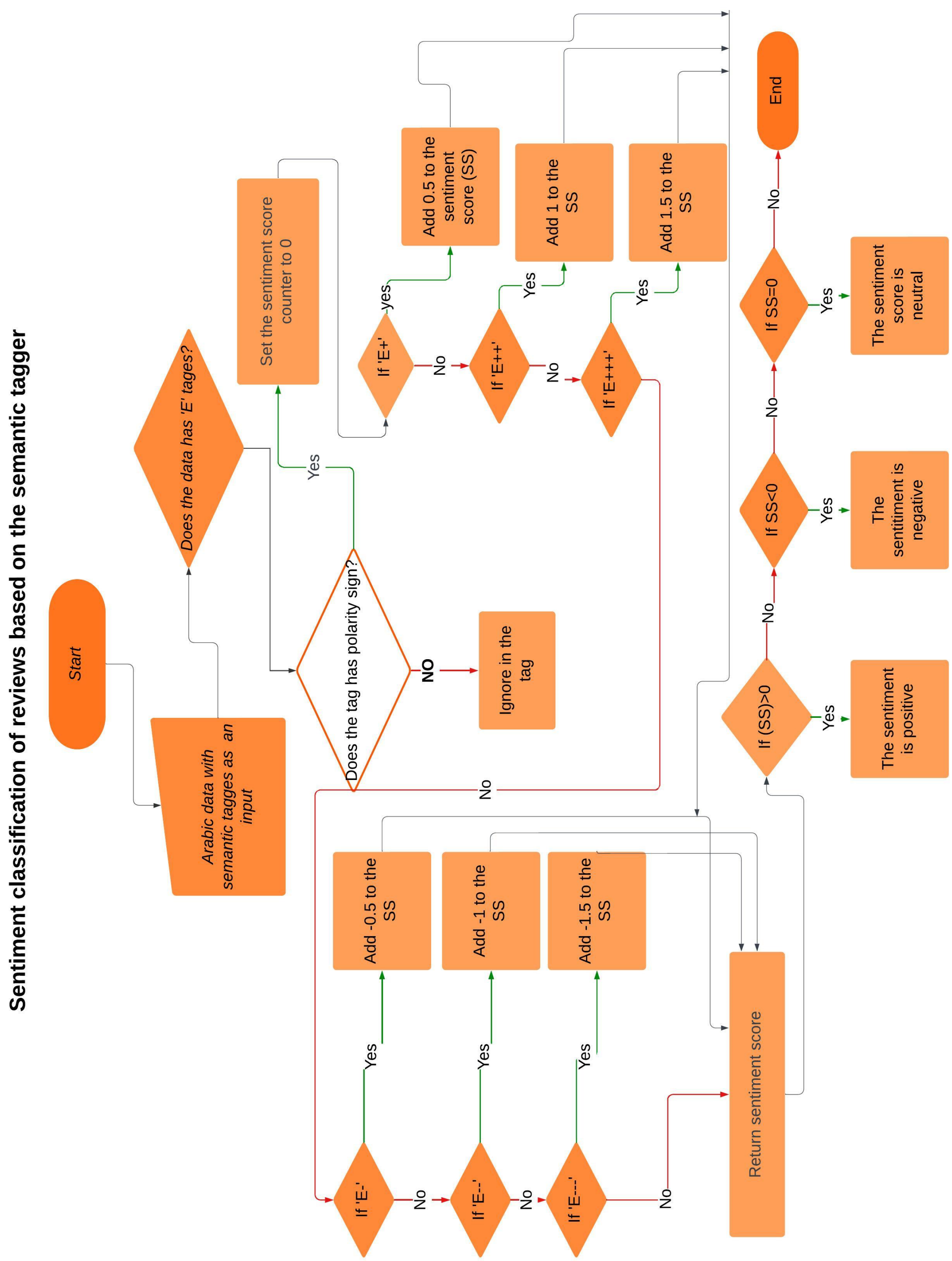


Figure 5.4: Sentiment detection tool

5.3.1.3 Mazajak Arabic sentiment analyzer

Mazajak is an Arabic sentiment analyzer with a web user interface for entering and correcting Arabic text. Mazajak is based on a neural network which can learn and adapt to new data. The user interface has a crowdsourcing annotation concept which enables the user to correct wrong sentiment predictions. Mazajak also has a feature that can upload a text file so that the data can be annotated all at once.

Using this method, the AMC was uploaded as a text file to produce an Excel file with sentiment annotation, which means that the AMC was turned into an Excel file. The Mazajak tool takes each row of the Excel file of the AMC and predicts the sentiment for each row and puts it into a new column. The output file will therefore be the AMC column with the sentiment annotation column for each row of the AMC. See the table in the appendix table.

We carried out a small similar experiment using Mazajak, which assessed the impact of a metaphor on Mazajak (Alsiyat and Piao, 2020a). In this experiment, the automatic sentiment analyzer shows the differences in the sentiment prediction on the LARB reviews. The differences are represented by testing the metaphorical and non-metaphorical reviews. The Mazajak prediction of sentiment for the Arabic metaphor reviews was evaluated, and the results showed that the achievement of a correct prediction was reduced by 40% with metaphor. However, the data used in this experiment were short reviews, whilst the AMC has reviews of different lengths with multiple metaphors. In using the AMC, as already mentioned, the Mazajak automatic method did not consider metaphor even though short sentences can contain metaphors. Moreover, the automatic sentiment analyzer has no feature to identify metaphors. The performance of the Mazajak was only comparable with the overall gold standard sentiment annotation. The percentage of the agreement between the two methods was 49%, as shown in Figure 5.1.

5.3.2 Methods for sentiment detection with metaphor

In this section, I am discussing the methods that are closely related to identifying sentiment metaphors. I divided the methods in order to determine the methods which can extract sentiment from a metaphoric section as the AMC has long reviews, which might affect the identification of sentiment. The first method was the sentiment classification from the metaphor section using the gold standard annotation and the other was automatic metaphor sentiment classification that converts the metaphor into a sentiment score.

5.3.2.1 Metaphor Gold Standard

The gold standard metaphor method relies on the manual sentiment annotation of the metaphor section of each review in the AMC. During the annotation, we specified the sentiment based on the annotated metaphor terms. We therefore chose the metaphor section to specify the sentiment based on the metaphor term annotation. We used the Gold Standard metaphor to specify the sentiment score using the Python code (see Figure 5.5).

5.3.2.2 Sentiment classification combining AraSAS and metaphors

In this method, for calculating the sentiment score for a review, a score of 2 is added to the sentiment score if the polarity is positive and subtracted 2 if it is negative. If it is neutral or null, return the sentiment score of zero. I regarded this as metaphor sentiment classification with a semantic tagger as it followed the AraSAS El-Haj et al. (2022) sentiment score tagging.

The flow chart in Figure 5.4 shows the initial framework designed to classify sentiment using the tagged data from the Arabic semantic tagger. The flow chart explains the code that performs the overall sentiment classification. As previously explained, the E-tag was the main tag for tagging emotion in the semantic tagger.

Another program was designed to detect review's sentiment based on the sentiment score given by the manual annotation for each review. The classification was based on the sentiment score. We shall now discuss the classification in more detail as a recommendation for use in future work.

As has already been discussed, the classification could be more precise if it is performed based on the numbers and types of the polarity signs for the E-tags. For example, before assigning a sentiment score to an E-tag, all E-tags with each polarity sign could be counted and the total could be assessed against a set of conditions in order to detect the overall polarity. Then the sentiment score can be assigned based on the labeled data. This means that if the polarity is negative, the sentiment will be -1, if it is positive 1 and if it is neutral 0. However, this method did not consider the degree of the positivity and negativity for each tag which we considered in our classification in the suggestion. Due to the cases which we encountered during the classification such as the equality of the E-tags and polarity signs, the work needs to be adapted to fit all cases and the metaphor detection, which would have exceeded the time frame set for this project. The sentiment classification was therefore implemented using the sentiment scores only.

```python
def add_metaphor_sentiment(text, num):
    score = num
    if text == 'positive' :
        score += 2
        return score
    elif text == 'negative' :
        score -= 2
        return score
    else:
        return score
```

Figure 5.5: Sentiment classification and sentiment score for metaphor

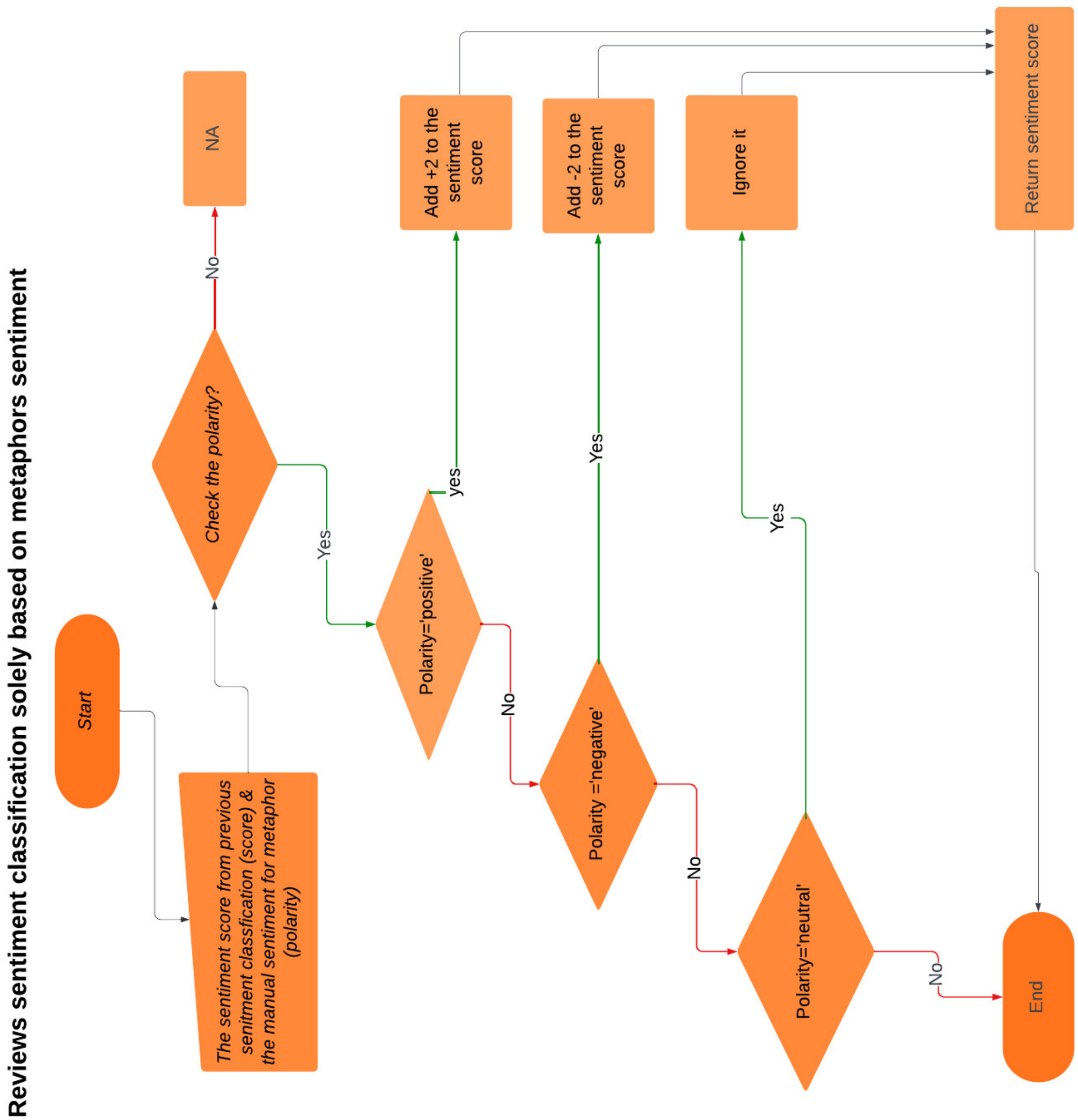


Figure 5.6: Metaphor sentiment classification tool

5.4 Evaluation

In this section, we assess the four methods for identifying the sentiment of reviews in regard to metaphor. The performances of these methods were assessed using the standard measurements of precision, recall, and F-score. The calculation of statistics was done automatically using a Python code shown in Figure 5.7. The classification results obtained by the four methods were assessed using the Python code with equations 5.3, 5.1 and 5.2. Then the F-score was used to evaluate the performances of the methods.

```
df['gold'] = df['gold'].apply(lambda x: -1 if x == "negative" else 0 if x == "neutral" else 1)
df['metaphor'] = df['metaphor'].apply(lambda x: -1 if x == "negative" else 0 if x == "neutral" else 1)
df['automatic'] = df['automatic'].apply(lambda x: -1 if x == "negative" else 0 if x == "neutral" else 1)
df['tags'] = df['tags'].apply(lambda x: -1 if x == "negative" else 0 if x == "neutral" else 1)
df['both'] = df['both'].apply(lambda x: -1 if x == "negative" else 0 if x == "neutral" else 1)
df.head(20)
```

	tokens	gold	metaphor	automatic	tags	both
0	1804	1	-1	-1	1	1
1	1762	1	0	-1	-1	-1
2	1701	1	0	-1	0	0
3	1197	1	0	-1	-1	-1
4	1034	1	0	-1	-1	-1
5	959	1	1	0	0	1
6	931	1	1	1	1	1
7	920	1	1	1	-1	-1
8	912	1	1	-1	1	1
9	864	1	-1	-1	0	-1
10	848	1	1	1	-1	-1
11	829	-1	0	0	0	0
12	763	0	1	0	1	1
13	757	1	1	1	1	1

Figure 5.7: Evaluation python code

We chose the two existing Arabic automatic tools, Mazajak Farha and Magdy (2019) and AraSAS El-Haj et al. (2022), to determine the impact of Arabic metaphors on automatic Arabic sentiment tools against the gold standard annotations by measuring the F-scores, even though not all of them are designed to identify sentiment and certainly not metaphors.

The Arabic semantic tagger is not designed to identify sentiment as the Mazajak tool. So we designed a program to classify sentiment based on basic sentiment represented by the emotional tags. The following methods are for sentiment annotation in regard to metaphors which were used to analyze the performance of each model based on the automatic tools explained above.

The following tables show the results of calculations of the standard measurement with different categories of token numbers in the AMC (Alsiyat et al., 2023). The number of tokens was considered as it can affect the sentiment decision, whereas Arabic online metaphors have no standard length of sentences.

It has already been stated that long reviews can have multiple polarities. For example, the semantic tool designed to detect and classify sentiment from the emotional tags also does this for the other tags which have other polarities. So I considered that the classification was not adequate for full sentiment classification, but it is nevertheless enough to identify sentiment in the metaphorical section. I therefore applied different codes in order to achieve an adequate classification, but it was still not optimal. This suggestion will be discussed further in the section on recommendations for future work.

Based on the statistics set for Arabic sentence length, the reviews from the AMC showed that sentences with fewer than 40 tokens are considered as short and that those with more than 40 tokens are long, even though online texts and writing do not strictly follow any of the Arabic language rules for writing a review. Long reviews formed 36% of the total, which means that the AMC contains more than 50% short reviews. The calculation starts by assessing the three columns of the automatic annotations and the gold standard annotations for overall sentiment and metaphor sentiment. The columns have to meet a set of conditions to change each polarity into a sentiment score. A polarity is then represented as a numerical value to calculate the precision, recall and F-score. For example, negative as -1, positive as 1 and 0 as neutral. The columns with sentiment scores are passed to a function to apply the precision, recall and F-score formulas. The function applies the multi-class sentiment classification for precision and recall, which means that there is a precise response to the rows of each method against the gold standard labels for each method. The recall responds to the columns of each method against the gold standard. This is depicted in the following formulas:

$$Precision = \frac{Truepositive}{Truepositive + Falsepositive} \tag{5.1}$$

$$Recall = \frac{Truepositive}{Truepositive + Falsenegative} \tag{5.2}$$

$$F - Score = \frac{2 * precision * recall}{precision + recall} \tag{5.3}$$

The formulas 5.15.2 are not applied manually to the multi-class recall and precision calculation. The annotation data cases from each model have to be extracted to calculate the recall and precision automatically. In detail, the cases of the four classes from the gold standard and one of the models had to be extracted prior to the calculation. For example, by counting the number of negative agreements between the gold standard and one of the models, and repeating it for all classes. Then the equations 5.1 and 5.2 were applied to the extracted tables for the four methods. In addition, the calculation for the precision, recall and F-score are applied for all other categories of reviews. For example, the recall and the precision will be applied to all the positive review categories for all methods.

Each method discussed below has the table containing the sentiment classification method and the method name. The sentiment classification methods have the review category and number of reviews. The review category has the reviews lengths ranges. The number of reviews column has the count number of the reviews for the corresponding category. In addition, the method name contains the calculations numbers of the standard equations 5.3, 5.1 and 5.2.

5.4.1 Mazajak method evaluation

Sentiment classification method:		**Mazajak sentence sentiment classification**		
Review categories	**Number of reviews**	**Precision**	**Recall**	**F-score**
All reviews	1000	0.7564289	0.718	0.72777
Positive reviews	702	0.9053156	0.776353	0.83589
Negative reviews	171	0.5183824	0.824561	0.63657
Neutral reviews	127	0.2539683	0.251969	0.25296
>=1000 tk	5	0	0	0
<1000 toks and >=500 toks	32	0.7705357	0.5625	0.62663
<500 toks and >=100 toks	268	0.7252957	0.608209	0.64149
<100 toks and >=90 toks	24	0.6547619	0.708333	0.67424
<80 toks and >=70 toks	30	0.7851852	0.733333	0.74222
<70 toks and >=60 toks	31	0.9205069	0.774194	0.81222
<60 toks and >=50 toks	57	0.7601726	0.649123	0.68706
<50 toks and >=40 toks	53	0.6771965	0.698113	0.68681
<40 toks and >=30 toks	70	0.6694678	0.657143	0.65778
<30 toks and >=20 toks	99	0.7703101	0.767677	0.76701
<20 toks and >=10 toks	149	0.7885742	0.812081	0.78211
<10 toks and >=5 toks	90	0.8661017	0.855556	0.83745
<5 toks and >=1 toks	63	0.9727891	0.952381	0.95825

Table 5.1: Mazajak sentiment classification result.

Before we discuss the table, the method described above in detail was applied 5.3.1.3. I shall discuss only the evaluation table and its analysis, such as the observation of the F-score in a specific category to show the impact of the metaphor (long/short review) on detecting sentiment.

The table shows the calculations of the three standard method assessments, precision, recall, and F-score. It also shows the different lengths of reviews in terms of tokens. In the column 'number of reviews', each category has the number of review counts for each range. For example, the category from 1 to 4 tokens had 63 reviews and an F-score of 0.958, which was the lowest token length with the highest F-score.

The Mazajak classification values are in the precision, recall and F-score 5.1, 5.2 and 5.3. The highest F-score is 0.958 for tokens ranging from 1 to 5. The F-score balances the classification as it has the true positive and the true negative as a binary classification. So we used the F-score to identify the highest and lowest scores of each model. The lowest F-score is 0.00 for the category of tokens in reviews less than or equal to 1000. The number of reviews in the >=1000 category had a review count of 5. The total reviews categories had an F-score of 0.728. The highest F-score from the Mazajak is 0.958 on between 1 and 4 tokens. All this information will be compared with the findings from the other methods to show the best or the highest F-score between the methods.

5.4.2 Sentiment classification based on semantic tags

Sentiment classification		**Sentence sentiment classification based on gold standard**		
Review categories	**Num. of revi.**	**Precision**	**Recall**	**F-score**
All reviews	1000	0.641	0.487	0.526
Positive reviews	702	0.831	0.520	0.640
Negative reviews	171	0.263	0.626	0.370
Neutral reviews	127	0.097	0.118	0.107
>=1000 tks	5	0.000	0.000	0.000
999 tks $\sim$500 tks	32	0.686	0.375	0.447
499 tks $\sim$100 tks	268	0.617	0.369	0.426
99 tks $\sim$90 tks	24	0.747	0.500	0.549
79 tks $\sim$70 tks	30	0.783	0.633	0.660
69 tks $\sim$60 tks	31	0.882	0.452	0.582
59 tks $\sim$50 tks	57	0.651	0.456	0.512
49 tks $\sim$40 tks	53	0.548	0.472	0.473
39 tks $\sim$30 tks	70	0.679	0.500	0.548
29 tks $\sim$20 tks	99	0.620	0.556	0.577
19 tks $\sim$10 tks	149	0.599	0.557	0.568
9 tks $\sim$5 tks	90	0.689	0.600	0.621
4 tks $\sim$1 tks	63	0.744	0.603	0.661

Table 5.2: GS annotation method measurement

This method was explained in detail in the section 5.3.1.1 above. Here, I shall explain the evaluation method and the resulting table to analyze the method's performance based on the F-score. The gold standard was used to compare and measure the performances of the other methods. A fifth column was added to the csv table, and the Python code was used to evaluate the other methods 5.7. The gold standard column acted as a sentiment score according to a set of conditions by using equations to calculate the standard measurement 5.3.

The highest F-score for this method was found to be 0.661, which was also the same as was found for the review category of between 1 and 4 tokens. The annotation for this category showed the highest scores for the two methods, which indicates that the sentiment scores for the short metaphor reviews were the same as the gold standard rather than the long reviews. The highest F-scores mean that the annotation was close to the gold standard annotation, whereas the lowest F-score was 0.0 for the less than or equal to 1000 tokens category.

5.4.3 Sentiment detection based on both semantic tags and metaphor sentiment information

Sentiment classification		Sent. senti. classification based on both semantic tags and metaphor senti. Info.		
Review categories	**Num. of rev.**	**Precision**	**Recall**	**F-score**
All reviews	1000	0.564	0.305	0.351
Positive reviews	702	0.735	0.285	0.411
Negative reviews	171	0.190	0.281	0.226
Neutral reviews	127	0.120	0.449	0.189
>=1000 tks	5	1.000	0.200	0.333
999 tks ∼500 tks	32	0.637	0.500	0.526
499 tks ∼100 tks	268	0.554	0.354	0.415
99 tks ∼90 tks	24	0.582	0.458	0.502
79 tks ∼70 tks	30	0.580	0.467	0.513
69 tks ∼60 tks	31	0.729	0.355	0.437
59 tks ∼50 tks	57	0.529	0.298	0.326
49 tks ∼40 tks	53	0.577	0.340	0.359
39 tks ∼30 tks	70	0.601	0.243	0.332
29 tks ∼20 tks	99	0.561	0.323	0.359
19 tks ∼10 tks	149	0.599	0.275	0.271
9 tks ∼5 tks	90	0.414	0.144	0.099
4 tks ∼1 tks	63	0.878	0.159	0.170

Table 5.3: Semantic tagger annotation method measurement with metaphor

This method was divided into two steps: the first to detect the sentiment based on the semantically tagged AMC and the second to convert the GS for the metaphor sentiment annotation into a sentiment score. The first step had to meet a set of conditions and functions in order to classify the sentiment (see Figure 5.2 and Figure 5.3). This method is a combination of the semantic tagger and the metaphor. The metaphor function 5.5 provided the sentiment score annotation from the AraSAS, which has sentiment scores calculated based on the semantic tags' polarity scores. The polarity scores were specified on the semantic tagger's subcategories.

For the GS of sentiment metaphor method evaluation 5.5, the precision, recall, and F-score were automatically calculated using the metaphor gold standard annotation with the sentiment classification using the Arabic semantic tagger.

In this method, the metaphor sentiment in the function is to be represented as text and the sentiment score as numerical parameters. The sentiment as a text parameter meets a set of conditions after assigning the sentiment with the sentiment score. The function's algorithm adds 2 if the sentiment is positive and subtracts 2 if it is negative. The function consists of a series of conditional statements, but it classifies metaphor sentiment based only on the metaphor gold standard annotation and not on detecting the metaphors in the text. There was no metaphor detection. After the classification, the new classification column was added to the classification table. The table below shows the values calculated from the standard measurements, which are precision, recall, and F-score for all categories 5.3. This method has lower F-scores than the other methods. The evaluation was done automatically using the Python code. The highest F-score was 0.52 for between 500 and 999 tokens, and the lowest was 0.09 for the category between 5 and 9 tokens in the reviews. These F-score results are logical as the metaphor annotations were compared with the overall sentiment gold standard during the calculations, which were completely different types of text.

5.4.4 Sentiment detection solely based on GS metaphor sentiment information

This section is to assess the gold standard metaphors. The sentiment score is compared to the gold standard annotation during calculating the equations 5.3, 5.1 and 5.2. This method is different from the automatic metaphor with semantic tagger. This method only converts automatically the gold standard manual annotation to the sentiment score, which is 0 for neutral, 1 for positive, and -1 for negative. We test this as a method, not as a tool like Mazajak tool, although I used a simple code to automatically assess the method.

The highest F-score found is 0.958 for the category 4 tokens to 1 token. Most of the methods have the highest scores for this category, which means that the sentiment predictions are better for the short reviews. The lowest F-score is for the neutral reviews category with a 0.17 F-score.

Sentiment classification		**Sent. senti. classif. only based on the gold standard metaphor senti. info.**		
Review categories	**Num. of rev.**	**Precision**	**Recall**	**F-score**
All reviews	1000	0.732	0.712	0.708
Positive reviews	702	0.883	0.788	0.833
Negative reviews	171	0.460	0.830	0.592
Neutral reviews	127	0.262	0.134	0.177
>=1000 tks	5	1.000	0.200	0.333
999 tks ∼500 tks	32	0.718	0.594	0.617
499 tks ∼100 tks	268	0.669	0.575	0.598
99 tks ∼90 tks	24	0.674	0.750	0.707
79 tks ∼70 tks	30	0.713	0.733	0.705
69 tks ∼60 tks	31	0.907	0.742	0.793
59 tks ∼50 tks	57	0.735	0.649	0.675
49 tks ∼40 tks	53	0.655	0.736	0.692
39 tks ∼30 tks	70	0.677	0.700	0.684
29 tks ∼20 tks	99	0.738	0.747	0.726
19 tks ∼10 tks	149	0.798	0.812	0.773
9 tks ∼5 tks	90	0.850	0.844	0.820
4 tks ∼1 tks	63	0.973	0.952	0.958

Table 5.4: GS automatic annotation method measurement for metaphor

5.5 F-scores Comparison

The Figure 5.8 and Table 5.5 show the values for the highest F-scores for the four models together for different review categories. The highest F-score is for the reviews with the tokens of range between 5 and 1. The category of 1 to 5 tokens produced the highest F-score because the metaphor is quite obvious in such short reviews than in longer reviews. It means the review is purely metaphoric with no other factors that may affect the sentiment prediction. Even though Mazajak has the best performance compared to the other methods, it is still inadequate. As we proved at the beginning of this research, its performance can downgrade when it comes to predicting the metaphors with sentiment (Alsiyat and Piao, 2020a).

I found Mazajak has the highest F-score between the four methods 5.8 and the gold standard for metaphor. In addition, Mazajak has the highest F-score in the most categories between all methods. The impact of the Arabic metaphor is more clearly seen using the Mazajak sentiment analyzer compared to the other methods. The reviews containing 1 to 5 tokens produced two equally high F-scores, which is 0.95 for both methods Mazajak and the automatic gold standard method. This means the Mazajak predicts sentiment well with metaphors. This observation could be confirmed further if we use big data.

These figures and the table 5.8 5.5 show the highest scores between the four methods in all categories. However, the highest F-score was achieved by the Mazajak sentiment analyzer in most of the reviews categories, while the gold standard for metaphor got the highest F-scores for four different categories. The categories with the highest F-scores for the metaphor gold standards are: for reviews of token numbers between ninety and one hundred, tokens between thirty and forty, tokens between five and one, tokens between fifty and forty, tokens of one thousand or higher, and the positive reviews. The highest F-score is 0.9 for the categories of reviews of tokens between five and one tokens. As the reviews of tokens between five and one are purely metaphoric. So, the sentiment could be more precise for that category than the others. The highest F-score was produced by Mazajak for the rest of the token ranges.

Compare the F-scores for all reviews		
Raw Label	**Highest F-score for each category**	**Method**
All reviews	0.72777423	Mazajak
Negative reviews	0.63656885	Mazajak
Neutral reviews	0.25296443	Mazajak
<10 and >=5 tk	0.83745098	Mazajak
<100 tk and >=90 tk	0.70714286	GS metaphor
<1000 tk and >=500 tk	0.6266335	Mazajak
<20 tk and >=10 tk	0.78211204	Mazajak
<30 tk and >=20 tk	0.76701156	Mazajak
<40 tk and >=30 tk	0.68374172	GS metaphor
<5 tk and >=1 tk	0.95825325	GS metaphor and Mazajak
<50 tk and >=40 tk	0.69212357	GS metaphor
<500 tk and >=100 tk	0.64149253	Mazajak
<60 tk and >=50 tk	0.68705953	Mazajak
<70 tk and>=60 tk	0.81221688	Mazajak
<80 tk and >=70 tk	0.74222222	Mazajak
>=1000 tk	0.33333333	GS metaphor
Positive reviews	0.83588957	GS metaphor

Table 5.5: F-scores comparison table

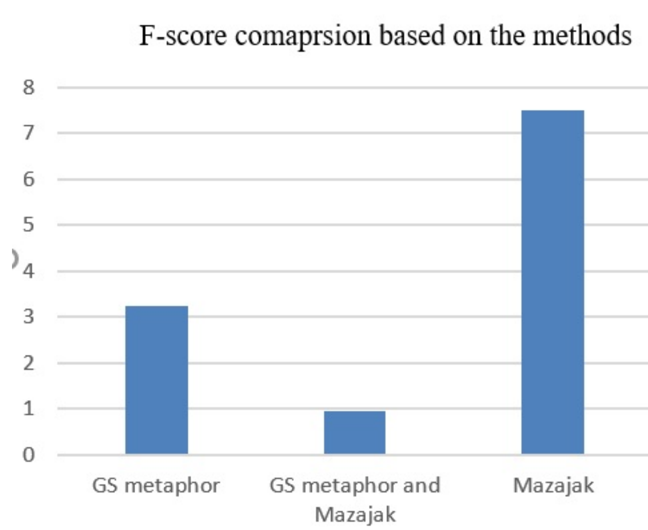


Figure 5.8: F-score comparison

5.6 Impact of Metaphors on Automatic Sentiment Detection

This research examines the impact of Arabic online metaphors, beginning with the challenges of annotation. As mentioned before in 4.8.1 that the sentiment of the same metaphor can be different in different contexts. In addition to the reviews with metaphor that shifts the sentiment of the reviews even with opinionated words after the metaphor. However, this chapter focuses on the impact using the automatic Arabic sentiment tools with the AMC. The impact reflected on the differences of the F-scores between the methods 5.5. However, the results may not be the best result even if it shows the big impact of metaphor on sentiment, due to the lack of available Arabic sentiment analyzers that can predict the sentiment accurately for any type of text. For example, Mazajak uses a deep learning method to predict the sentiment, which should have accurate prediction for the Arabic metaphors. As Mazajak was built on deep learning, which can learn from the text it analyzes. Additionally, Mazajak has a feature to correct the predicted polarity. But it was not accurate when it includes metaphors, even though it has the highest F-scores of the most review categories. There are only a few available Arabic sentiment analyzers, such as those developed by Farha and Magdy (2019) and Thelwall et al. (2010), and not all of them are built only for the Arabic language. For example, the SentiStrength is built for English language and supports the Arabic text. Therefore the result of the highest F-score for Mazajak is logical, reflecting the lack of the sufficient existed Arabic tools.

Through the experiments on the Arabic metaphor corpus AMC, I found that the Arabic online metaphor has multiple factors that can affect the sentiment, such as the meaning, length, and parsing. Also, the experiment shows the impact of metaphor on different sentiment classification methods. For example, during tagging the AMC using AraSAS, some of the parsed words are marked as Z99, which means *'unmatched'*. There were wrong tagging for some metaphoric words, too.

Not to mention the factors that affect the meaning, which are the way of the online metaphor written, the choice of words, and the form of words. Precisely, the Arabic metaphor corpus has new words from informal communication. For example, الألش/*'al-ālš'*, which is a word that has no reliable resource to find the meaning; hence, to specify the sentiment. الألش/*'al-ālš'* meaning, which is based on the annotators' annotations, is *'sarcasm'*.

For the word form, كتاب ملحد has the metaphor ملحد , which is an unusual form of word that defines the online Arabic metaphor as discussed. This means the Arabic online metaphor could be specified through the form of the word. In detail, the word ملحد, which means *'atheist'*, in the metaphor context, while the word in the normal context is إلحادي, which means *'atheistic'*. Those cases were discussed in the first chapter, but I mentioned them here again to show the factors that impact the metaphor and the sentiment respectively.

5.6.1 Why metaphorical information is important for sentiment classification

As discussed earlier, people online tend to express their opinions in a short and quick manner using metaphors. Since the metaphor takes on a new shape and meaning in the online context, the necessity to express the metaphorical text has increased to understand the opinions towards the subject. The problem of identifying the metaphorical sentiment is the scarce resource of Arabic online metaphors. During our annotation and experiment for metaphors, the metaphor demonstrates a significant impact on identifying the sentiment.

The metaphor drives the sentiment for the short reviews. It is shown by the highest F-scores for the short reviews category, which contain between 1 to 40 tokens. While the long reviews were affected by multiple factors. The impact showed in the Arabic sentiment analyzers' different performances when the metaphorical information is used. Also, the impact is shown by the differences of agreement between the GS and other annotations in the manual annotation. For example, Mazajak predictions changed when it compares to the gold standards annotations between the metaphor and overall sentiment (see Figure 5.10 and Figure 5.9). In an example for the manual annotation, the review داحمد خالد توفيق ارتدى رداء ليس برداءة . فكتب رواية أشبه بمسخ., the sentiment for the review is derived by metaphor. The metaphor ارتدى رداء ليس برداءه means *'He donned a robe that was not his own'*, which indicates the negativity of his writing style that he is not competent in. So, the sentiment is driven by metaphor.

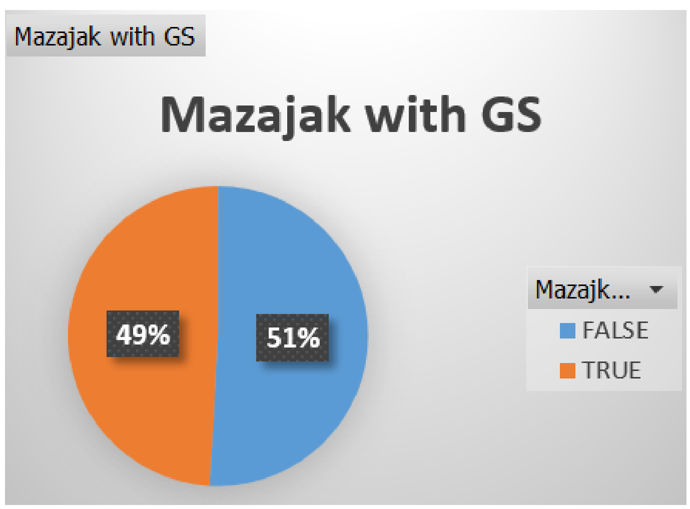


Figure 5.9: Mazajak-GS-overall comparison

The metaphor should be identified automatically to know the accurate sentiment. However, it is even more challenging for the manual annotation. In the AMC, there are multiple factors that affect finding the correct sentiment. For example, above all the known

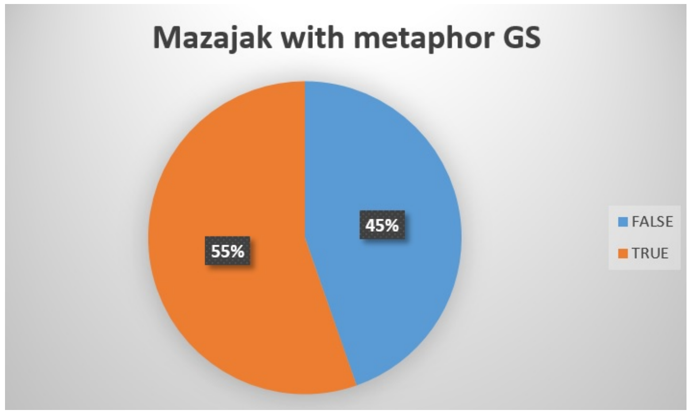


Figure 5.10: Mazajak-GS-metaphor comparison

reasons is the ambiguity of metaphor in any text. The online metaphor and the Arabic features make the identification more challenging 4.6.

The reviews were chosen to combine the metaphor and sentiment. During the annotation, some of the reviews were written in dialects. So, the metaphor's meaning was necessary to specify the sentiment. But, as there is no reliable Arabic metaphor resource, the annotators had to assume the sentiment based on the context. For example, هيلحس دماغك means *'will lick your brain'*. The annotators were uncertain if the metaphor was negative or positive. Because it can be interpreted as *'the book is good to the limit will fascinate you'* or *'the book is bad to confuse you'*. As a result, each of the annotators interprets and identifies the sentiment based on their own understanding and assumption.

Below I will discuss the methods to optimize the sentiment identification. My main method is to test the metaphor using the state of the art automatic Arabic sentiment analyzer. However, there is no available Arabic metaphor corpus to test the Arabic metaphor. So, I built the Arabic metaphor corpus. I used methods including the manual annotation of the Arabic metaphor corpus and the automatic Arabic sentiment analyzer.

For the manual annotation of the sentiment for metaphor and the overall, the statistics proved fair agreement between the annotators. However, as mentioned in the discussion, the metaphor sentiment can be affected by multiple factors depending on the Arabic online metaphor context. So, multiple annotators may be needed to improve the quality of the annotation.

For the automatic annotation of sentiment, the sentiment was identified using two state-of-the-art Arabic automatic sentiment analyzers, which are Mazajak and Arabic semantic tagger. As discussed above, the data was passed as a text file for the Mazajak tagger and returned as tagged sentences in an Excel file. Mazajak's prediction is forty-nine percent compared to the gold standard sentiment annotation, which has about 0.8 precision, 0.7 recall, and 0.72 f-score.

During the Mazajak annotation, it considered metaphor information for the short review as the sentiment is driven by the metaphors. However, for the lengthy sentences containing metaphor, the sentiment could be driven by other words as well. So, I suggest a solution for better annotation, which is identifying the sentiment after splitting the word rather than sentence. Splitting the words will raise the capacity of identifying the other factors that affect the sentiment such as metaphor and semantics.

To specify the polarity of each word in the text, the sentiment is specified based on the biggest number from the counted polarity. The previous solution could be sufficient in detecting the overall sentiment at the sentence level with the neural network, but not if other factors are considered such as the semantics.

The other method applies a similar method to detect the sentiment of the Arabic metaphor corpus, but based on the Arabic semantic taggerEl-Haj et al. (2022) using the emotional tags. The emotional tags can be considered as the main sentiment tags to identify the sentence polarity. However, all the tags could be considered in the sentiment classification after splitting the reviews into words. The classification can be carried out based on a set of conditions.

The idea of detecting the sentiment of the Arabic metaphor corpus was explained in the previous section. The Arabic semantic tagger assigns each word with a semantic tag. So, the association of tags can be used as a solution for sentiment classification and the semantic, which could detect metaphor. In other words, instead of using an Arabic sentiment analyzer, which is based on a neural network or machine learning, the Arabic semantic tagger can be used. Because we consider the tags as a feature for classifying the semantic, we can classify metaphors. The classification of metaphor will be based on the assigned semantic tags. As we classify the metaphor terms to be on drugs, mental health, food, etc. So, the semantic tagger could be a good solution to classify the metaphors. However, we don't know the accuracy of this method as the metaphor could be assigned with regular words, not categorized as semantic words. For example, the formal metaphor was usually assigned with a regular form of words: verb, noun, and adjective. So, the Arabic semantic tagger does not consider the previous words as semantic.

In order to apply the above suggestion, there are multiple things that need to be considered in the Arabic semantic tagger. The Arabic semantic tagger is affected by the dialects and the data pre-processing. For example, in the review جننتني حراااااااام عليك , the elongation annotates as unmatched. Similarly, for the dialect example, هيلحس دماغك, the word

هيلحس was tagged as *'unmatched'*. While the metaphor means *'will lick your brain'*, *'lick'* is a verb and it is an indication of confusion. The pre-processing in our automatic annotation was not applied as the punctuation affects the metaphor meaning. For example, تحفة رائعة . لم أتمتع بعمل مثل هذا من قبل. Also, during the annotation, the metaphor expression and the overall annotation perform based on the concept where the metaphor expression stops. So, the Arabic semantic tagger could not specify the semantic of some Arabic metaphors. For example, مخدرات, which means *'drugs'*, was tagged as Z99 (means unmatched). Another example, the word التحشيش/ *'al-tahšīš'*, which is a verb derived from *'weed'*, is tagged as 'Z99'. Those words should be tagged as F3, which means smoke and non-medical drugs. The mis-tagging of those words can indicate two things. First, the Arabic semantic tagger cannot recognize the Arabic dialect. Second, it cannot recognize the Arabic verb form, which is an online metaphorical term. For example, التحشيش which is a recognized noun in the Arabic language, it is a new metaphor term. Another example of the dialectal verbs regarding the Arabic semantic tagger is the verb هيلحس, which was tagged as 'Z99', maybe because the verb هيلحس is written in dialectal form. And it means *'will lick'* in literal sense, metaphorically means *'confusing'*

5.7 Comparison

This section compares the agreement between manual and automatic annotations in a nontraditional way, different from the traditional Inter-Annotator Agreement (IAA) and standard calculation methods 5.5. Specifically, it focuses on observing identical agreements between the annotators in both approaches, referred to as observed agreement or raw agreement (Artstein, 2017). The automatic annotation systems compared were Mazajak and AraSAS. The manual annotations from both annotators were also compared using Excel features. The Excel sheet columns were trimmed using the trim function to remove any extra spaces in the inserted annotations. Additionally, unique values were used to identify any typos in the annotations. Comparison figures were extracted by checking the value equality (using the Equal function in Excel) of each cell in both columns. If the cells in both columns were equal, the result was TRUE; otherwise, it was FALSE. The resulting columns with binary results indicated the agreement and disagreement between the automatic and gold standard annotations. Based on these binary results, figures were extracted to analyze the annotations. The pie charts illustrate the percentage of observed agreements in the annotations. This section analyzes each figure, considering the performance and factors discussed above that affect the annotation.

In addition, this comparison highlights the challenges of identifying the sentiment of Arabic metaphors using different annotation methods, both automatic and manual. The agreements were compared to show the differences in annotation approaches for identifying sentiment, despite their reliability. The figures below show the comparisons between the annotators' annotations and the gold standard annotation. Additionally, they compare the automatic annotations of Mazajak and the semantic tagger AraSAS El-Haj et al. (2022) with the gold standard annotation.

In the first group of annotations 5.11 5.12, the manual annotations show high agreement with the gold standard annotation, with over 50% raw agreement. In contrast, the second group of annotations 5.13 5.14 shows lower raw agreement, below 50%. This comparison reflects the impact of Arabic online metaphors on sentiment using automatic annotation methods. Even with the high agreement on metaphor annotation, the two annotators achieved 70% agreement on metaphor term choice.Means the two annotators have 70% agreement of the metaphor term choice.

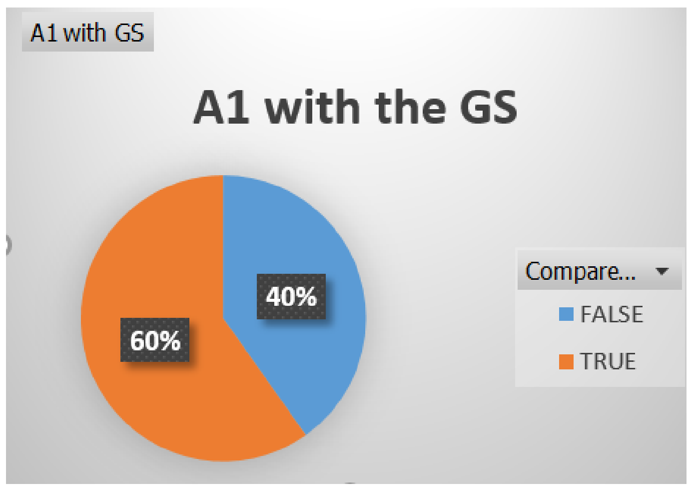


Figure 5.11: Compare the first annotator with GS

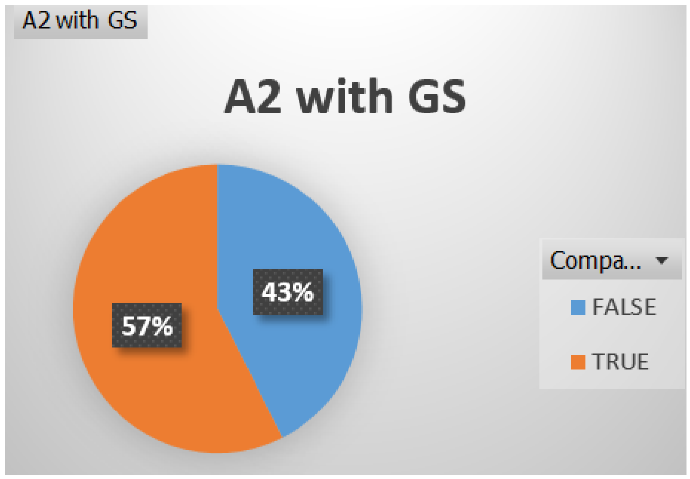


Figure 5.12: Compare the second annotator with GS

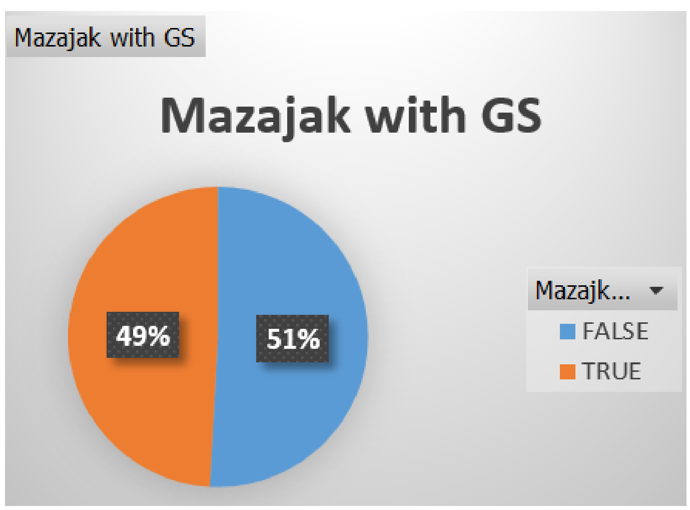


Figure 5.13: Compare Mazajak with GS

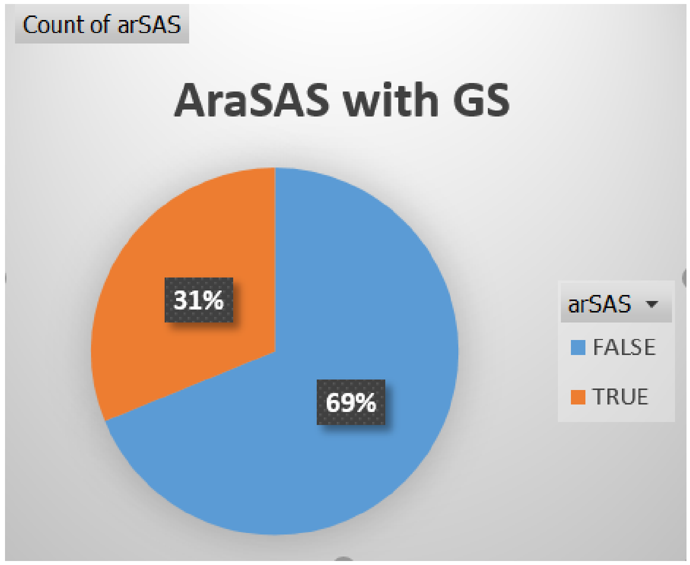


Figure 5.14: Compare automatic tagger AraSAS with GS

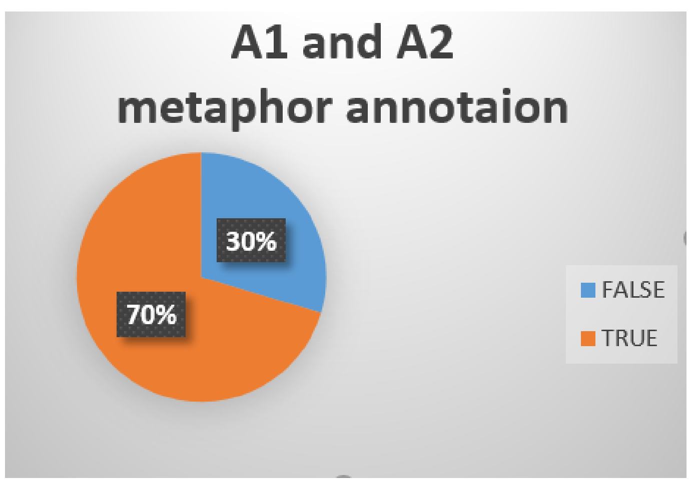


Figure 5.15: annotators raw agreement on the metaphor term

Annotations	**False**	**True**	**Total**
A2 and GS	57%×1000=570	43%×1000=430	1000
A1 and GS	40%×1000=400	60%×1000=600	1000
Mazajak and GS	51%×1000=510	49%×1000=490	1000
AraSAS and GS	69%×1000=690	31%×1000=310	1000
Total	2170	1830	4000

5.7.1 Statistical significance calculations

The p-value $4.38x10^{-36}$ is extremely small (much smaller than $\alpha = 0.05$). The calculation above is a confirmation of the extremely significant statistics of the two groups of categorical data.

5.8 Chapter summary

This chapter demonstrates the impact of Arabic online metaphors using three automatic methods. The manual annotation was converted to an automatic process using Python code for the evaluation. The Gold Standard (GS) manual annotations were changed to sentiment scores using Python code and then used to find sentiment in the designed tool and in the evaluation. Although these are not considered automatic tools for showing the impact of metaphors, they were included in the methodology as they were used in the tool design.

Overall, all the outcomes are approximate results, as the tools are still inadequate for accurately finding sentiment in relation to metaphors. As a consequence, we evaluate the exited and designed tool for accurate results. In addition, the annotation results are observed to analyze and suggest a solution. Evidence has been gathered from the suggestion, but it is still undergoing. The tools need rounds of development to achieve the best results. However, the tool evaluation could reveal the most effective technique for finding sentiment in relation to metaphors. Based on the F-scores results, the best performance to find the sentiment using AMC was Mazajak. Also, comparisons were made to show the impact of using the raw agreement of the automatic and manual annotations.

To illustrate, this chapter demonstrates the impact of Arabic online metaphors on the current state of automatic Arabic sentiment analysis. The statistical results and tools evaluation (F-Scores) highlight this impact. However, they also indicate that Arabic sentiment analysis still needs to advance to provide accurate predictions for Arabic sentiment using Arabic online metaphors. So, the designed tools and the results were designed and evaluated under unevolved Arabic sentiment analysis tools compared to sentiment for the English language.

Therefore, a suggestion discussed in the conclusion chapter 6.3 6.4 is to improve sentiment predictions and identify metaphors. These tools are not to show the impact; rather, they aim to optimize sentiment identification with metaphor. However, those suggestions are only subject to the AMC corpus.

Chapter 6

Conclusions

6.1 Summary of Research

This study investigated the impact of Arabic online metaphor using state-of-the-art automatic sentiment analyzers for the Arabic language. However, there is no available annotated resource for Arabic online metaphor regarding identifying sentiment. In addition, the available automatic Arabic sentiment analyzers do not identify metaphor to specify sentiment accurately, especially in the new Arabic online metaphor text. Therefore, the identification of sentiment in Arabic metaphors has not yet been possible. Building a reliable Arabic metaphor resource was, therefore, a crucial step towards showing the impact of the Arabic online metaphor on sentiment.

The findings of the current study therefore contribute to knowledge by building an Arabic online metaphor corpus (AMC) with sentiment, semantic structure, meaning, context, theme and metaphor types. Building the AMC was an essential and promising step for this research and for other researchers seeking to carry out further investigation into Arabic online metaphors and sentiment. This research shows the impact of the Arabic online metaphor on sentiment by using different methods despite the limited resources for Arabic metaphor research and Arabic sentiment analyzers. For example, there are only three automatic Arabic sentiment analyzers: Farha and Magdy (2019), Thelwall et al. (2010) and El-Masri et al. (2017). These sentiment analyzers are not all currently available and are not all designed to identify sentiment in Arabic text. For example, Thelwall et al. (2010) identifies the strength of sentiment rather than the sentiment itself, as discussed in the literature review chapter 2.

The experiments of this research do not produce the desirable results as a consequence of the inadequacy of the automatic Arabic sentiment analyzers, in addition to the high levels of ambiguity in the new Arabic online metaphor. The IAA was able to test the AMC's reliability, which we had expected to be high, but this is normal for an immature and new type of data. In addition, the impact was more shown by the Mazajak F-score than by the other Arabic sentiment analyzer methods, which we expected to be shown in all methods. A high F-score signified high compatibility between the GS (human annotation) and the Mazajak (automatic annotation). However, different results had been expected as the automatic tools were not all built based on adapting to the new Arabic metaphor data to detect sentiment. Even so, the outcomes and analyses of this study are a fruitful resource for future researchers seeking to identify Arabic metaphor and to study the structure of Arabic online metaphor. In addition, the AMC corpus is a fundamental step in identifying Arabic metaphor automatically. It will open another perspective on identifying Arabic metaphor regarding sentiment and semantics. The AMC was annotated manually by Arabic native speakers, semantically using AraSAS El-Haj et al. (2022) and automatically for sentiment using Farha and Magdy (2019) and the designed tools were used based on the output of the AraSAS to identify and classify sentiment.

As explained above, the AMC is a foundation for the automatic detection of sentiment in Arabic metaphor as the corpus contains different categories including context, meaning, theme and metaphor type. It was constructed as the first Arabic metaphor resource in regard to sentiment and semantic structure because no previous resource was available for identifying sentiment in Arabic metaphor. Throughout the process of building the corpus, the structure of the Arabic metaphor was investigated and the findings showed that sentiment is affected by metaphor and many other factors such as the length, meaning, and context of a review. We therefore annotated the AMC with context and meaning to clarify those terms' which can have different semantic polarities.

During the annotation, we observed changes in the structure of Arabic metaphors when used in the online context. As part of the annotation process, the structure of Arabic metaphor was examined to understand the sentiment and the meaning. The AMC showed the impact of Arabic metaphor through the different annotation choices made by different annotation methods. With regards to the low agreement between the annotations made by different annotators, the low IAA score is a reflection of the high level of ambiguity in Arabic online metaphors. The analysis of the Arabic online metaphor structure enabled us to extract the features from which we can identify sentiment in the metaphors. The Arabic metaphor schema used was based on the practical aim of identifying Arabic online metaphor, but it is not clear yet how effective those features are for identifying sentiment in metaphor using the AMC corpus as the features need to be applied to a large amount of data in order to train the tool to detect metaphor accurately. So meaning and context, in addition to the other categories mentioned above, were annotated in the AMC. The methods used to annotate the AMC were compared to the gold standard annotation using the standard measurements for evaluation and agreement. The annotation methods showed the closest annotation to the gold standard in the F-scores calculation. However, the results show that there is no best method to annotate Arabic online metaphor because there are multiple factors that affect the various annotation methods, the most significant of which are the unstable structure and meaning of online Arabic metaphor. These reasons are specific to this type of data and annotation.

Generally, however, the automatic tools are still inadequate for identifying sentiment in metaphors. Even so, Mazajak showed relatively better performance. This raised the need to build an Arabic sentiment analyzer to identify sentiment in metaphor or to construct a sentiment tool which can detect metaphor before the identification of sentiment.

6.2 Achievements and findings of my research

The main two aims of this research are to build an Arabic metaphor corpus with sentiment as a consequence of no Arabic metaphor ready to use to identify Arabic metaphor in regards to sentiment. In addition to test the Arabic metaphor corpus on a web-based Arabic sentiment analyzer to show the impact of metaphor on predicting the sentiment. However, the Arabic metaphor changes in an online context. So, the annotation scheme should fit the purpose of identifying the metaphor with sentiment and facilitate the automatic identification in the future. During our experiments, the Arabic metaphor in online contexts takes different meanings in similar contexts. So, the meaning was found to be very necessary to be annotated. In addition, being specific about the context before and after the metaphorical expressions to accurately define the Arabic metaphor term.

The impact was shown by using state-of-the-art Arabic sentiment analyzers. However, there are no more than three automatic Arabic sentiment tools, and not all of them are available to use. So, our impact was shown by only using the Arabic sentiment analyzer that fits to show the impact. The Arabic sentiment analyzers have been evaluated using the AMC to show the impact and to reflect the limitations of the available Arabic sentiment tools. The Arabic semantic tagger was used to tag the AMC as the Arabic online metaphor uses semantic words to express opinion. The Arabic semantic tagger called AraSAS El-Haj et al. (2022) was used to fill the gap of the available Arabic sentiment analyzers as well. The tools were designed and analyzed in assistance with the El-Haj et al. (2022), which should be optimized as we believe the sentiment classification should be done on the review as a whole. As a limitation of the AraSAS, it is not designed to identify the sentiment. So, as a suggestion to identify the metaphor in regards to sentiment, one should link the semantic category to see the possibility of having a metaphor in the review. This means the output should be only YES or NO if the review has a metaphor or not by setting up a list of conditions.However, as discussed, this is only for AMC, as we do not know about other data that could be used for this purpose.

Corpus building is an extensive task that requires time and effort to annotate the data. Our approach focuses on corpus building with text analysis of Arabic metaphors in an online context. Although we aimed to construct the AMC corpus for practical purposes, we could not ignore the differences and the new structure of Arabic metaphors in online contexts for annotation purposes. It is crucial to understand the text before the annotation. The practical approach in our method involves using state-of-the-art existing Arabic tools, assisted by code design for sentiment classification using semantic information of AMC with AraSAS El-Haj et al. (2022). Additionally, we evaluate the state-of-the-art Arabic tools used to predict sentiment for the AMC. Thus, the findings are beyond the aim and tasks of this research. They represent what we discovered and observed during the application of the experiments. Future works mentioned are mostly about improving and creating a new Arabic sentiment tool that can identify metaphor and sentiment. In addition to extending the amount of the corpus to fit any advanced tools to predict the metaphor and sentiment. Therefore, we divide the findings and different future works based on the chapters' results and findings of this research as follows:

6.2.1 Building the AMC

This task was one of the main aims of this study. We wanted to build a reliable resource as a first and reliable reference for any researchers investigating Arabic metaphors. The AMC is the first resource and reference for Arabic online metaphor for automatic detection. The corpus contains several sub-tasks with multiple findings at each stage, starting from the structure, context, meaning, and type of Arabic metaphor used online and ending with the annotation criteria for multiple categories, challenges, and the annotation evaluation.

During the building of the AMC, we investigated the structure and meaning of the Arabic metaphor. We sought to identify the same Arabic metaphor terms with different sentiments in different contexts, and we found that the meaning of the same metaphor term and sentiment did change in different contexts. The new structure of online Arabic metaphor showed the pattern of Arabic metaphors for feature extraction 6.2.2. However, we did not know the efficiency of the features on a larger amount of data. The features were not applied due to the time limitations and the specific aim of this research. In addition, those features were beneficial for the AMC corpus as it was designed in the same way as the features. Also, it could be used for retrieving information from the corpus. Similar work has been done using an English lexicon, which is the BNC Leech (1992), but the BNC has a huge amount of data, whereas our corpus had a different aim. We did not go through that work as it does not show the impact of metaphor on sentiment and we were constrained by the time limitation on this research and because building a corpus is expensive. We wanted to create a tool for retrieving information and finding occurrences to distinguish the metaphoric from the literal context.

6.2.2 Extracted features

The features extracted are one of the AMC findings. These features could serve as conditions in functions linked to our AMC corpus. The AMC was analyzed to identify sentiment based on our annotation of metaphorical words and their contexts. Although these patterns are useful for extracting sentiment and identifying metaphors to be tested in machine learning in case of the corpus expansion,the AMC is restricted to trigram words of context words based on the metaphor position. This means that the trained algorithm will look only three words before or after the metaphor term following the AMC annotation. However, this may not have high accuracy for the metaphor terms that come as one term only or the context that clarifies the metaphor words after the three words.

- The metaphor can be identified using the polarity, and the algorithm will learn from the annotated AMC with sentiment. The terms associated with contradictory polarity (positive and negative words) ممتعه لدرجة مرعبه, which means *'joyful till terrifying'* in a literal sense; metaphorically, it is so joyful, it is more likely to be positive and metaphor. Because the term (for example, مرعبة/*'scary'* is negative in the literal sense, when it's associated with a positive word, it is more likely to be expressed as negative and metaphor. Another example, رائعة حد الثمالة means *'fabulous till intoxication'*. So, the metaphor comes in the negative word associated with the positive word. In a literal context, the same term comes with positive words (ex رائعة و مرعبة/*'terrifying and amazing'*). It is more likely to be positive and literal. We put this assumption as the word مرعبة *'terrifying'* usually comes in a literal sense to describe a negative incident. But when it is associated with positive words, it is more likely to be metaphoric. In addition, the previous example has *'till'* word, which comes often with metaphorical words in the AMC. However, this discusses the Arabic metaphor in the online context, where the text is unpredictable and changeable, as discussed in the previous chapters. So, it could not apply to all social media/online text.

- In addition, the metaphor can be identified using the target words and the contexts following the feature below. The AMC is annotated with the context of the metaphorical words. The contexts were specified as a definition of the metaphorical words. So, the contexts specified after or before the metaphorical terms. The contexts can be compared with the same term in different contexts. However, this won't detect the semantic meaning of the sentences. For example, morphology affects the similarity even when the two words are similar in meaning.

  - Where the target words same with different contexts.
  - Where the target words (different form but same meaning) with different contexts.

In this condition, the meaning annotation could be used to spot the term with the same meaning.

– Where the target words similar (different morphology) with similar contexts.
– Where the target words same with the same context. This condition produce same classification for the polarity and the if it is metaphor or not metaphor.

The conditions mentioned are based only on the AMC corpus as it has no literal context to distinguish the literal/metaphoric context. However, this could be improved later in future works by adding the literal context and extending the AMC.

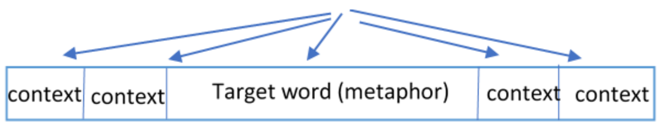


Figure 6.1: Features code suggestion

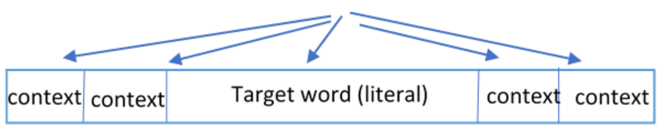


Figure 6.2: Features code suggestion-literal

6.2.3 Analysis of Impact of metaphors on sentiment analysis

I carried out a series of experiments to test the impact of metaphors on sentiment detection. I tested some methods by including the metaphors' sentiment information in the sentiment detection process, based on an existing Arabic sentiment analysis tool and semantic tagging tool. Although the metaphor information did not always improve the results, the experiment results provide deep insight into how metaphors can be integrated into the automatic Arabic sentiment detection algorithms and framework.

6.2.4 The aims achieved

1. For the RO1, we collected one thousand Arabic reviews containing metaphors in an online context from a large-scale Arabic lexicon annotated with sentiment, called LARB Aly and Atiya (2013). So, we produce an Arabic metaphor data set containing online Arabic metaphors.

2. In regard to RO2, The Arabic metaphor dataset collected was manually annotated with metaphor, sentiment, theme, metaphor type, metaphor meaning, context and part of speech. So, we produced the first Arabic Metaphor Corpus (AMC).

3. With regards to RO3, there are a few Arabic metaphor tools, but not all of them are suitable for this purpose. So, we chose the ones that are based on predicting the sentiment as polarity more than the ones that produce the sentiment as a score.

4. With regards to RO4, I evaluated the Arabic sentiment analyzers by applying the standard measurements to assess the tools. F-score is used to evaluate the impact on different tools.

5. In regard to RO5 and RO6, I annotated the AMC with semantics as another potential Arabic sentiment analyzer in order to cover the lack of available Arabic sentiment analyzer. So, we designed and tested programs to identify the sentiment of the semantically annotated reviews of the AMC.

6.2.5 Research questions revisited

The answers to the research questions have been discussed throughout the chapters. The research questions are answered based on what each chapter proves and produces. The answers are summarized below:

1. The first research question (RQ1) addressed the analysis of the Arabic online metaphor structure, revealing the unpredictable nature of these metaphors. While the most frequent online metaphor structures were defined by their context, some still required interpretation (meaning annotation) prior to sentiment annotation. The varied metaphorical structures that define online Arabic metaphors may consist of a single word to convey an opinion. For example, the word كحولية literally means *'alcoholic'*. However, in a metaphorical sense, it could be interpreted as *'romantic'* or "calming" based on the annotators' annotations. Consequently, such words need to be interpreted before sentiment annotation to accurately determine polarity. Furthermore, the unpredictable structures of online metaphors became evident during the design of the annotation scheme to meet the data analysis requirements. In some instances, there was no supportive text to clearly define the metaphor. For example,

the context was not always three words before or after the metaphorical term, but sometimes only a single word or even just punctuation following the metaphorical word. So, the schema was restricted to those cases.

2. The second question (RQ2) was answered in new Arabic metaphor terms types in terms of the new structure and new terms and whether they have a source to know the metaphoric meaning and the literal meaning as well. The new Arabic metaphor terms contain new dialectal words which do not have a reliable source to know even their literal meaning. In addition to the new mean of describing metaphor in online context using semantic voice to express opinion.

3. This question correspond to RQ3, was explained by analysing the meaning and the structure of the standard and the new Arabic metaphor with evidence from the collected Arabic metaphor data. The schema designing was done after the data analysis for the Arabic metaphor data that meets the previous questions, which is correspond to the fact that Arabic metaphor has unpredictable structure. In addition, the schema was designed to meet the practical purpose of metaphor identification. For example, annotate the context to understand the sentiment and define the metaphorical expressions before or after the term.

4. In the RQ4, The schema designed as mentioned before based on the practical purpose. However, it is still influence by other annotation schema designed for English metaphor such as Krennmayr and Steen (2017). Even though, they followed the standard annotation concept which is MIPVU (Metaphor Identification Procedure VU University Amsterdam), which is not the case in this research. Our aim is to identify online metaphor terms as novel, identify the metaphor in online context, and annotate the sentiment of the AMC. The corpus annotation chapter discuss the comparison between the previous studies that influence our schema. The VU Amsterdam identify metaphor linguistically annotating only the existing of metaphor rather identify the metaphorical term. In addition, they annotate the sentences as all and binary annotation for metaphor. Means yes/no annotation of the metaphor existence. our schema follow only some of the XML annotation structure of the VU Amsterdam and understand the contextual meaning of the metaphor context to identify the metaphor. Hence, identify the sentiment. The vu Amsterdam is not all metaphoric, it has literal sentences. While the AMC is all metaphoric. Our schema designed based on the practical purpose, which used to show the impact of metaphor using the AMC. However, the AMC consider as a base knowledge for many research ideas as future work.

5. RQ6 was addressed by designing a sentiment identification tool in association with the Arabic Semantic Tagger (AraSAS) and the semantically tagged Arabic Metaphor Corpus (AMC). The tools were designed to determine the overall sentiment of the

semantically tagged AMC. In addition, another tool designed to identify sentiment based on manual metaphor annotation with the semantically tagged AMC.

6. RQ7 was answered through statistical analysis by comparing the F-scores of the tools to find the best-performing method with the highest F-scores. Additionally, the impact was shown by comparing the raw agreement between the automatic and manual annotations.

6.3 Limitations of this research

This section discusses the limitations of this project.

6.3.1 Need for further exploring tools and techniques

During the application of showing the impact of metaphor on sentiment, I found that Arabic sentiment analysis needs improvement in automatically identifying Arabic sentiment. This is especially apparent due to the limited availability of automatic Arabic sentiment analyzers compared to English ones. The attempt to optimize the code as part of demonstrating the impact of metaphor on sentiment was observed in the overall classification process. For example, the classification should encompass all polarity signs of different aspects within reviews.

We utilized the Arabic semantic tagger El-Haj et al. (2022) to demonstrate the impact of metaphor. Although it was not initially designed for sentiment analysis, the two tools were created to detect sentiment. The AraSAS has the potential to identify metaphor through semantic tags, but this has not been fully explored yet.

The designed tools exhibited different performances on the AMC, indicating the need for further improvement. This underscores the ongoing process of optimizing sentiment classification using the AMC. For example, the AMC occasionally showed equal polarity for some reviews, making it challenging to classify sentiment when the tool cannot decide the sentiment in cases of equal polarity. Additionally, AraSAS sometimes provided incorrect semantic tagging for certain metaphorical words, such as assigning 'unmatched' for a metaphorical word, which is incorrect 5.6. Thus, the tool itself still requires improvement to handle different types of Arabic text. Furthermore, this evaluation is only applicable to the AMC data and not to any other dataset.

6.3.2 Need for further exploring annotation information for Sentiment Classification

To specify a program for classifying sentiment in the AMC, we used only the emotional tags of the AraSAS. For better sentiment classification, the program needs to apply all the tags

```
def word_wise_score_sentiments(data2):
    try:
        split_sentence =  data2.split(' ')
        final_list = list()

        for new_word in split_sentence:
            word = new_word.strip()
            if len(word)>0:
                if '+' in word:
                    final_list.append('+')
                elif '-' in word:
                    final_list.append('-')
                else:
                    final_list.append('N')
            else:
                pass

        final_d = dict(Counter(final_list))

        return final_d if final_d else dict()

    except Exception as e:
        print(str(e))
        return 'NA'
```

Figure 6.3: code optimization

of the Arabic sentiment analyzer. The classification could then be performed based on the polarity of the most frequent polarity sign in each review. This method is yet to be explored.

The code splits each review into words and processes them through a set of conditions to count the highest polarity signs in each review and determine sentiment. The result is presented as a dictionary for each review containing the number of polarity signs. However, this method faced challenges, particularly in cases where polarity signs were equal. Consequently, adjustments had to be made to the code to address such cases.

Even though these tools do not directly detect Arabic metaphor, they still utilize the AMC as part of demonstrating the impact of Arabic metaphors on sentiment. Furthermore, the code could be further optimized to detect semantic possibilities of metaphoric reviews, but this remains to be explored.

The analysis of AMC structure cannot be generalized for all online contexts unless the data contains a considerable amount of online metaphor occurrences. So advanced techniques such as machine learning can learn from the pattern of the big amount of new Arabic metaphor terms.

For the Arabic sentiment analyzers, those analyzers lack recognizing the metaphorical text in advance of identifying the sentiment. So, the results were not the desirable ones. We expect dramatic changes in the Mazajak performance, but it appears to produce unstable predictions with different amounts of data. For example, when we tested using Mazajak, Mazajak's performance was downgraded, although in this research Mazajak had the best performance among the other tools.

With the assistance of the AMC, this problem could be solved by identifying the context of the Arabic online metaphorical text. In addition, this could be solved by tokenising each text to a word assigned with the sentiment. Then identifying the metaphor depends on the metaphor context using the AMC. However, this process could be effective after multiple rounds of improvements to the tool that was designed. In addition, the tool that will be designed should be compatible with the automatic Arabic sentiment analyzer.

```
def word_wise_score(data2):
    try:
        list_sub=['E1','E1+','E1-','E2','E2+','E2-','E3', 'E3+', 'E3-', 'E4.1', 'E4.1-',
        found_emotion_list = list()
        split_sentence =  data2.split(' ')
        final_list = list()
        for word in split_sentence:
            if '+' in word:
                final_list.append('+')
            elif '-' in word:
                final_list.append('-')
            else:
                final_list.append('N')

        final_d = dict(Counter(final_list))

        if final_d:
            pos_count = final_d ('+',0)
            neg_count = final_d.get('-',0)
            neutral_count = final_d.get('N',0)

            if (pos_count > neg_count) and (pos_count > neutral_count):
                return 'Positive'
            elif (neg_count > pos_count) and (neg_count > neutral_count):
                return 'Negative'
            elif (neutral_count > pos_count) and (neutral_count > neg_count):
                return 'Netural'
            elif (neutral_count == pos_count):
                return 'Positive'
            elif (neutral_count == neg_count):
                return 'Negtaive'
            elif (pos_count == neg_count):
                return 'Neutral'
        else:
            return 'Neutral'

    except Exception as e:
        print(str(e))
        return 'NA'
```

Figure 6.4: Equality polarity

6.3.3 Limitation of data collection

As already explained, the new online Arabic metaphor dataset was collected manually by Arabic native speakers from publicly available online data (Aly and Atiya, 2013). They were asked to select frequently used metaphors on social media and search through the LABR data Aly and Atiya (2013). The size of the dataset was about 51,000 words, which might be considered small. However, the data could be extended in future work by adding the same online metaphor terms found in online sources such as Twitter. Also, the data could be expanded by more than a thousand reviews with different semantic categories. As mentioned before, the collection was limited to LABR, which means that the occurrence of metaphors was limited compared with the social media feeds. The data were collected from a specific source, namely LARB, which limited the analysis to the LABR (Aly and Atiya, 2013) dataset only. Although the metaphorical terms gathered were found to be the same as the metaphors found in social media feeds, the analysis could be widened if data were to be gathered mainly from social media because social media have a similar data structure. In future research, social media data could be added and compared with the current data to show similarities and differences. The data gathered from a specific dataset and the analysis of the new Arabic metaphor corpus cannot apply to all online contexts.

6.3.4 Limitation of corpus building

The first limitation is the number of book reviews, even though we consider the corpus has a good amount of words to be built as lexicon. But the practical aspect was the main aim of this research to train any machine learning algorithm accurately. for which a big amount of data is crucial. So, extending the number of reviews is a solution for this limitation. The second limitation of this corpus is that it is not following the standard methods. This is controversial as the scheme was designed only for this type of data, which means that it will not fit every online Arabic metaphor data. For example, the online text discussed in the introduction chapter could come in Arabizi Duwairi and Qarqaz (2014), and metaphor could come in this form as well. So, this may not fit all metaphor text types in the online context.

Nonetheless, the scheme is ideal for practical use as it has all the necessary aspects to identify the online metaphor starting from the context. Because most of the previous studies in the English language relied on the lexicon to identify metaphors. And the metaphor identification based on the lexicon follows the verb-noun violation. So, the AMC has all aspects of the lexicon to identify the metaphor. However, this again is still only for the AMC, not all types of metaphor text in the online context.

6.3.5 Lack of accurate automatic Arabic sentiment analyzers

The lack of Arabic sentiment analyzers is a limitation of this project, too. The lack of available Arabic sentiment analyzers was an obstacle to achieving the desirable results. I found that all the available ones are inadequate to use. For example, the Mazajak tool is the only one that predicts the polarity, while the other ones predict using the sentiment score, which is unreliable. In addition, none of the Arabic sentiment analyzers were built to identify metaphors and use them for sentiment classification. So, we suggest a solution to only find the existence of metaphor based on the AMC corpus and ARASAS semantic tagger.

6.4 Future direction of research

This section discusses the future direction of this research, building on the arguments and solutions presented previously. I have already achieved the key objectives of this project and identified future directions for this research based on the findings of this project. Although I have already implemented some of these potential suggestions, there is still much room for further exploration. The findings were discussed in each chapter, and here I focus on the discussion of how to carry this research forward in the future.

One of the future directions is to expand the AMC by adding Arabic metaphors from social media, in order to facilitate a comparative study of the AMC data with social media data to find their similarities and differences. The new data could be gathered from one of the social media platforms such as X, Facebook, and Instagram. Even though we believe the AMC is similar to the social media context, I suggest a deeper comparison and analysis could lead to interesting findings.

In further detail, the data suggested will be gathered on the concept of finding similar Arabic metaphor terms in one of the social media platforms. If the AMC is expanded with a similar annotation category, the AMC could be considered as a newly created Arabic online metaphor lexicon if it is assigned a friendly user interface to find concordances and retrieve metaphorical information. For example, the corpus could be converted to XML tags and linked to a web page with categories, which is one of the key features for publication. I considered this idea at the beginning of this research. However, this did not meet the aim of this project and the corpus building with this specification would take extensive time.

Another suggestion to go further with this research is to advance the sentiment classification of the AMC. To advance the classification, we need to consider conditions to fit all sentiment polarities, in the cases where the sentiment of the metaphor cannot be recognized by the AraSAS El-Haj et al. (2022). In classifying the sentiment using El-Haj et al. (2022), the emotional tags were not assigned to all metaphorical terms, but for all words without specifications. There are reviews with only two words, which contain metaphors. Such short reviews in our corpus make up forty-six percent of the total reviews. Considering the 'E' tags (indicating emotion) only could be a better solution for the data that has only metaphors combined with emotional tags, while other semantic tags can indicate the sentiment such as '-L1' (meaning *'dead'*. So, the suggestion is to consider all tags to detect the sentiment of a text.

The AMC is a fruitful resource for many research applications. Linking the AMC to a friendly user interface will be a good start for many applications:

- It could be built as a lexicon. So, the AMC could have linked to a Friendly user interface to retrieve information.

- In addition, finding the metaphor based on the noun-adjective violation concept with sentiment, which can be achieved using the context of each metaphorical term and sentiment annotation.
- Using the AMC as a resource to be linked to an automatic Arabic metaphor detection in regards to sentiment.
- It could be used for linguistic purposes as well as to extract statistical information to find the occurrences of an Arabic metaphor term.

Appendix A

Figure

The figure describes a sample of the Arabic metaphor authentication result for the meaning and dialect. The figure shows the highest number of responses for certain meaning and dialect for one online Arabic metaphor. The figure added just to show the online Arabic metaphor authentication process mentioned in the data collection chapter 3.2. Here is the full version of the questionnaire Questionnaire full version

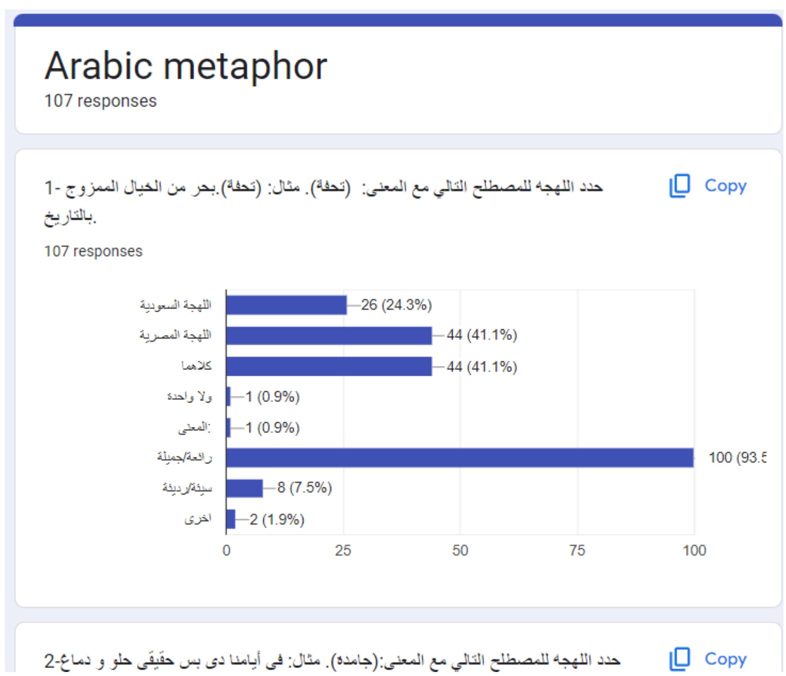


Figure A.1: A showcase of the Arabic metaphor dialect authentication

Appendix B

The annotation table sample

Here is a sample of the AMC corpus with metaphor, sentiment and meaning only. The sample has metaphor annotation, which labeled by one of the Arabic native speaker. The sentiment is the gold standard sentiment for metaphor. The meaning column is for one of the Arabic native speaker meaning annotation. The full annotation table AMC is available online in AMC corpus.

Table B.1: Arabic metaphor terms Table

Metaphor	**Sentiment**	**Meaning**	**Transliteration**
هيلحس	negative	يفاجئك	*hayilḥis*
غفلنا	negative	محى	*ghafalnā*
العتمة	negative	الضياع	*al- 'utmah*
سوداء	negative	حزينة	*sawdā'*
تتقلص	negative	تنفي	*tattaqallaṣ*
ملحد	negative	جحد	*mulḥid*
برداءة	negative	أسلوبه	*biradā'ah*
مريضة	negative	سيئة	*marīḍah*
تشدك	positive	تجذبك	*tashudduka*
خرابة	negative	فساد	*kharābah*
جامده	positive	جميلة	*jāmidah*
خطفت	positive	استولت	*khaṭafat*
يحلق	positive	يسرح	*yuḥalliq*

حرقت	negative	أغاظتني	*ḥaraqat*
تتفجر	positive	مليئة	*tatafajjar*
الثمالة	negative	أقصاه	*al-thumālah*
الثمالة	positive	جداً	*al-thumālah*
متدفقة	positive	ممتلئة	*mutadaffiqah*
عبقرية	positive	ذكية	*abqarīyah*
وروده	positive	جماله	*wurūdih*
محمل	negative	ممتلئ	*muḥammal*
الثوب	positive	المناسب	*al-thawb*
الأم	positive	الاكتفاء	*al-umm*
تحفة	positive	إبداعية	*tuḥfah*
مخيف	positive	مهيب	*mukhīf*
بالقلم	negative	تقليدية	*bi-l-qalam*
تجديف	negative	سيئة	*tajdīf*
سحر	positive	جاذبية	*siḥr*
أبحر	positive	تسافر	*'abḥara*
تلتهم	positive	تهي	*taltahim*
لصراع	negative	تناقض	*li-ṣirā'*
يسلط	positive	يركز	*yusalliṭ*
يحطم	positive	يقتل	*yuḥaṭṭim*
أتنفس	positive	أعمق	*'atanaffas*
تغوص	positive	تتعمق	*taghūṣ*
السم	negative	الخبث	*al-summ*
سلاح	positive	علاج	*silāḥ*
علم	positive	متداول	*' ilm*
صديق	positive	مواسي	*ṣadīq*
يخفى	positive	الجمال	*yukhfā*
بيطبطب	positive	يطمن	*bi-yaṭbaṭib*
تصنع	positive	حنون	*taṣanna '*

قتلت	positive	أصبت	*qatalt*
تعيش	positive	تبهرك	*ta' īsh*
ليخونك	negative	الغضب	*li-yakhūnuka*
أكلت	positive	أثرت	*'akalt*
بئر	positive	عميق	*bi'r*
ترسمين	positive	تبدعين	*tarsumīn*
جرح	negative	ألم	*jarḥ*
أدمت	negative	أحزنت	*'admat*
الدافنين	negative	الحزينين	*al-dāfinīn*
الانزلاق	negative	الضياع	*al-inzilāq*
القبعة	positive	الاحترام	*al-qubba'ah*
البياض	positive	الصفاء	*al-bayāḍ*
نسمة	positive	خفيف	*nasmah*
يشع	positive	يملأ	*yasha '*
تشم	positive	تتخزن	*tashumm*
تأخذ	positive	الخفة	*ta'khudh*
مغرقة	negative	مليانة	*mughraqah*
مكمل	positive	علاج	*mukammil*
أنيس	positive	مواسي	*'anīs*
يبحر	positive	يعلمك	*yubḥir*
رفيقي	positive	مواسي	*rafīqī*
يخاطب	positive	يوعي	*yukhāṭib*
كنز	positive	ثمين	*kanz*
نقطة	positive	يغيرك	*nuqṭah*
ارتدى	negative	التقليد	*irtadā*
بتطبطب	positive	يحن	*bi-taṭbaṭib*
ثوب	negative	الطفش	*thawb*
اليد	positive	الحنان	*al-yad*
خشبية	negative	تقليدية	*khashabiyyah*

تهزك	positive	تتأثر	*tuhizzuka*
محموم	negative	متعمق	*maḥmūm*
تسحرني	positive	تعجبني	*tusḥirnī*
مبحر	positive	متوسع	*mubḥir*
قطعة	negative	ابداعي	*qiṭ' ah*
قمة	positive	إبداعية	*qimmah*
ستأكلك	negative	ستتهي	*sata'kuluka*
بحر	positive	مجموعة	*baḥr*
يحفر	negative	يبحث	*yaḥfir*
رائحة	positive	الحنين	*rā'ḥah*
أبواب	negative	تؤلمك	*'abwāb*

Table B.2: Arabic Informal and Transliterations

Arabic	**Transliteration**
تحفه	tuḥfah
جامد	jāmid
جامده	jāmidah
الألش	al-‘lsh
بيض	bayḍ
قشطه	qishṭah
اشطا	ishṭā
لذيذة	ladhīdhah
لذيذ	ladhīd ā
عسل	’ asal
مسكر	musakkir
سكر	sukkar
شهي	shahī
نكهه	nakhah
طزاجة	ṭazājah
وجبة	wajbah
بصل	baṣal
مذاقاُ	madhāqan
مقبلات	muqabbilāt
دسم	dasim
البهارات	al-bahārāt
الطبخة	al-ṭabkhah
بطيخ	baṭṭīkh
خلطة	khalṭah
طعم	ṭa ’ m

Arabic	Transliteration
فظيعة	faẓī' ah
فظيع	faẓī'
مخيف	mukhīf
قاتلة	qātilah
خيالي	khayālī
خورافي	khurāfī
مجرم	mujrim
اجرام	ijrām
فقيع	faqī'
فشخ	fashkh
افشخ	afshakh
فشيخة	fashīkhah
مرعبة	mur ' ibah
فتاك	fattāk
موت	mawt
جنان	junūn
خطير	khaṭīr
قصة	qiṣṣah
جنونية	junūnīyah
خوقاقي	khawqāqī
خوقاق	khawqāq
علاج	' ilāj
جرعة	jur ' ah
مضاد	muḍādd
كبسولة	kapsūlah
تخدّر	takhaddur
يجيبلك مغص	yījīb-lak maghaṣ

Arabic	**Transliteration**
مريضة	marīḥah
صرع	ṣara ’
تخلف عقلي	takhalluf ’ aqlī
ع الجرح	’ al-jurḥ
بدوخة	bidūkhah
ملعون ابو الصدق	mal ’ ūn abūal-ṣidq
اكتئاب	ikti‘āb
للجرح	lil-jurḥ
مجنونة	majnūnah
هلس	halas
سلاح	silāḥ
قنبلة	qunbulah
دماغ	dimāgh
حد النخاع	ḥadd al-nukhā’
بوشين	bushayn
تسكر	taskar
سكرى	sukrā
تحشيش	taḥshīsh
مخدرات	mukhaddarāt
كحولية	kuḥūliyyah
الثمالة	al-thumālah
خرا	kharā
زبالة	zibālah
زفت	zift
خرابة	kharābah
أكلت	‘akalt
يفطس	yifṭis

Arabic	Transliteration
تسبح	tasbaḥ
القاتلة	al-qātilah
تلتهم	taltahim
يجيبلك	yījīb-lak
خضني	khuḍnī
زغزغني	zaghzaghnī
بتطبطب	bi-ṭaṭbṭib
بتسحب	bi-tasḥab
نحفر	naḥfur
يخرب بيت	yikhrib bīt
هيلحس	hayilḥis
بتطلعني	biṭṭalla ' nī
تشدك	tashuddak
سوداء	sawdā'
خفيف	khafīf
مضروبة	maḍrūbah
نسمة	nasmah
خنيقة	khanīqah
ممطوطه	mamṭūṭah
غسيل مخ !	ghasīl mukh
خبط لزق	khabaṭlazq
الرغي الرغي	al-raghīal-raghī
غسيل مخ	ghasīl mukh
دماغه عالية	dimāghuh ' ālīyyah
خفيف دم	khafīf damm
شهادتى مجروحة	shahādatīmajrūḥah
ثرثرة مجانية	thartharah majjānīyyah

Arabic	**Transliteration**
ضحك السنين!!!!	ḍaḥik al-sinīn
سهلة الهضم	sahlah al-haḥm
حد الموت	ḥadd al-mawt
كتاب ملحد	kitāb mulḥid

Appendix C

Corpus annotation in XML format

This is the full annotation format of the AMC, the samples added to show the annotation format as some was not included on the extracted Excel files. The XML annotation contains the metaphor types and part of speech, which is not included in the AMC as Excel table.


```xml
<?xml version="1.0" encoding="UTF-8"?>
<sentences>
    <body>
        <s no="1"> ميلحس دماغك <genreT Type="general">ميلحس</genreT>
        <metaphor Type="Verb">
            <Informal>يلحس</Informal>
            <Tword Type="MD">ه</Tword>
            <Tword Type="VBP">يلحس</Tword>
        </metaphor>
        <literal>
            <Tword Type="VBP">يحيرك</Tword>
        </literal>
        <context Type="before">
            <Tword Type="NN">دماغ</Tword>
            <Tword Type="PRP">ك</Tword>
        </context>
    </s>
    <s no="2"> تاريخ غفلنا عنه ودوّن في كتب التاريخ باختصار قتل وجعه وجماله <genreT Type="general"
            >كتب</genreT>
        <metaphor Type="Verb">
            <Formal>قتل</Formal>
            <Tword Type="VB">قتل</Tword>
        </metaphor>
        <literal>
            <Tword Type="VB">مسح</Tword>
        </literal>
        <context Type="after">
            <Tword Type="JJ">وجع</Tword>
            <Tword Type="PRP">ه</Tword>
            <Tword Type="CC">و</Tword>
            <Tword Type="JJ">جمال</Tword>
            <Tword Type="PRP">ه</Tword>
        </context>
    </s>
    <s no="3"> يبقى في العتمة <genreT Type="general">الرواية</genreT>
        <metaphor Type="Noun">
            <Formal>العتمة</Formal>
            <Tword Type="NNO">العتمة</Tword>
        </metaphor>
        <literal>
            <Tword Type="NNO">الجهل</Tword>
        </literal>
```


Figure C.1: Annotation XML sample

```xml
        </context>
    </s>
    <s no="4">الرواية سوداء وربما تسأل نفسك لم لم يكن هناك ولا لحظة سعادة في هذه الرواية إطلاقاً <genreT Type="general">
      الرواية</genreT>
        <metaphor Type="Adjective">
            <formal>سوداء</formal>
            <Tword Type="JJ">سوداء</Tword>
        </metaphor>
        <literal>
            <Tword Type="JJ">متشائمة</Tword>
        </literal>
        <context Type="before">
            <Tword Type="CC">و</Tword>
            <Tword Type="RB">ربما</Tword>
            <Tword Type="VBP">تسأل</Tword>
        </context>
    </s>
    <s no="5"> تتقلص مفرداتها كل يوم <genreT Type="general">الرواية</genreT>
        <metaphor Type="Verb">
            <Formal>تتقلص</Formal>
            <Tword Type="VB">تتقلص</Tword>
        </metaphor>
        <literal>
            <Tword Type="VB">تقل</Tword>
        </literal>
        <context Type="after">
            <Tword Type="NN">مفردات</Tword>
            <Tword Type="PRP">ها</Tword>
            <Tword Type="CC">كل</Tword>
            <Tword Type="NN">يوم</Tword>
        </context>
    </s>
    <s no="6"> كتاب ملحد <genreT Type="general">كتاب</genreT>
        <metaphor Type="Adjective">
            <Formal>ملحد</Formal>
            <Tword Type="JJ">ملحد</Tword>
        </metaphor>
        <literal>
```


Figure C.2: Annotation XML sample

```xml
        </context>
    </s>
    <s no="5"> تتقلص مفرداتها كل يوم <genreT Type="general">الرواية</genreT>
        <metaphor Type="Verb">
            <Formal>تتقلص</Formal>
            <Tword Type="VB">تتقلص</Tword>
        </metaphor>
        <literal>
            <Tword Type="VB">تقل</Tword>
        </literal>
        <context Type="after">
            <Tword Type="NN">مفردات</Tword>
            <Tword Type="PRP">ها</Tword>
            <Tword Type="CC">كل</Tword>
            <Tword Type="NN">يوم</Tword>
        </context>
    </s>
    <s no="6"> كتاب ملحد <genreT Type="general">كتاب</genreT>
        <metaphor Type="Adjective">
            <Formal>ملحد</Formal>
            <Tword Type="JJ">ملحد</Tword>
        </metaphor>
        <literal>
            <Tword Type="VB">بعيد عن الدين</Tword>
        </literal>
        <context Type="before">
            <Tword Type="NN">كتاب</Tword>
        </context>
    </s>
    <s no="7"> د.احمد خالد توفيق ارتدي رداء ليس برداءة . فكتب رواية أشبه بمسخ <genreT Type="author"
            >الكاتب</genreT>
        <metaphor Type="Noun">
            <Formal>برداءة</Formal>
            <Tword Type="NN">رداء</Tword>
        </metaphor>
        <literal>
            <Tword Type="NN">تغير</Tword>
        </literal>
        <context Type="before">
            <Tword Type="VB">ارتدى</Tword>
            <Tword Type="NN">رداء</Tword>
```

Figure C.3: Annotation XML sample

Appendix D

AMC file as an input file for Mazajak annotation

The figure below is for the AMC with no annotations inserted to Mazajak as input file.

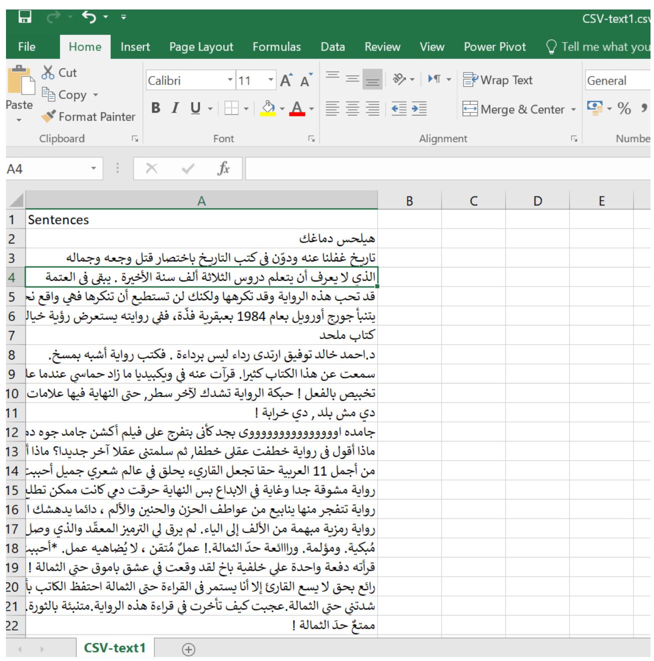


Figure D.1: AMC file as input file for Mazajak annotation

Appendix E

Full AMC

Here is the AMC with full annotation categories. The table contains the human annotations of the two Arabic native speakers and the automatic annotations of the Mazajak and AraSAS. Although the AMC here have no reviews tagged with semantic tags, but it could be found in the link AMC corpus.

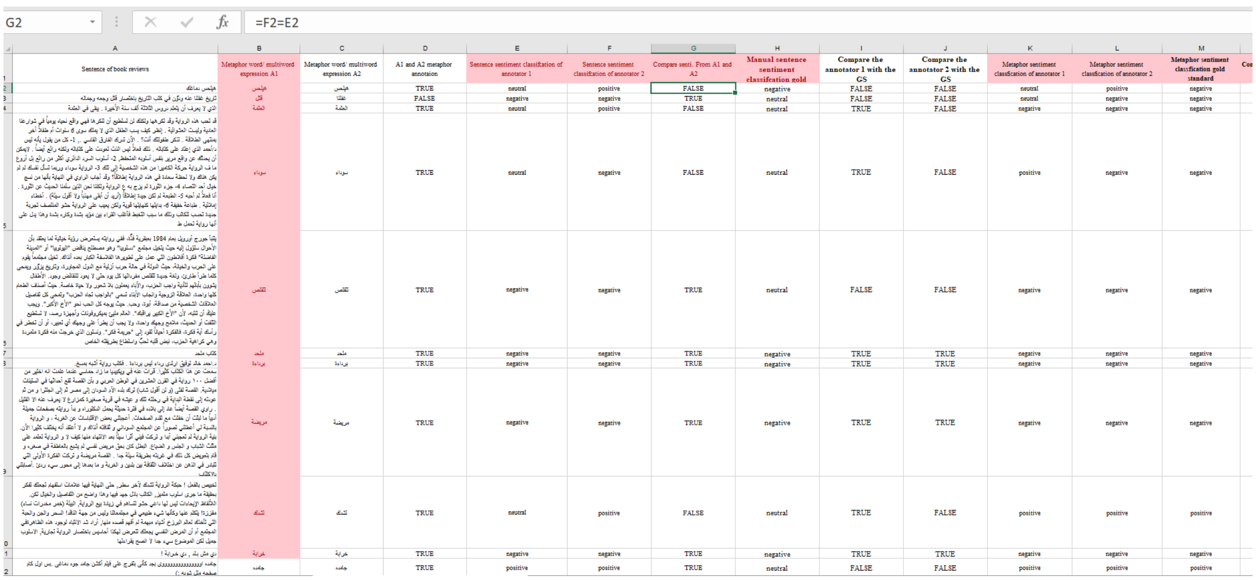
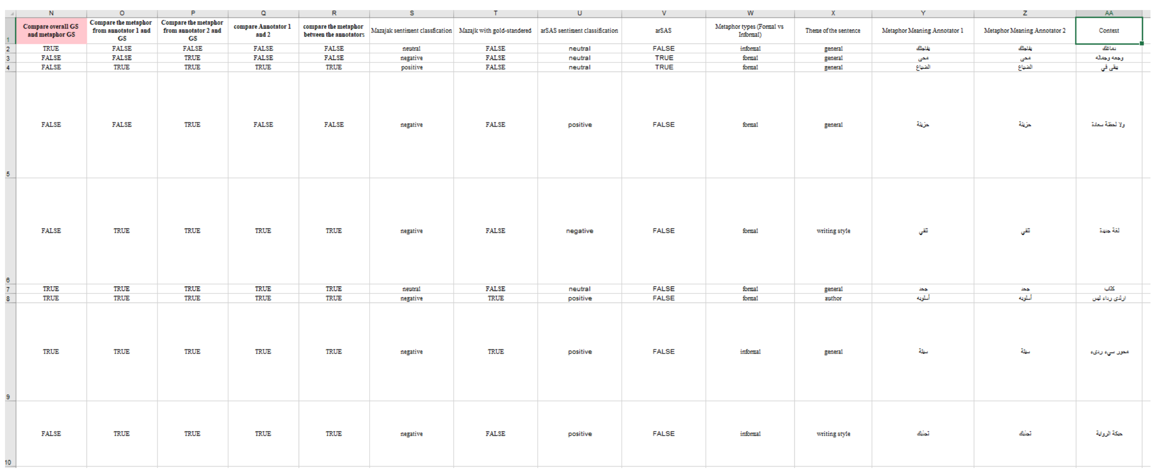


Figure E.1: AMC with all annotation categories



	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1	Compare overall GS and metaphor GS	Compare the metaphor from annotator 1 and GS	Compare the metaphor from annotator 2 and GS	compare Annotator 1 and 2	compare the metaphor between the annotators	Mazajak sentiment classfication	Mazajk with gold-standered	arSAS sentiment classification	arSAS	Metaphor types (Formal vs Informal)	Theme of the sentence	Metaphor Meaning Annotator 1	Metaphor Meaning Annotator 2	Context
2	TRUE	FALSE	FALSE	FALSE	FALSE	neutral	FALSE	neutral	FALSE	informal	general	يفاجئك	يفاجئك	دماغك
3	FALSE	FALSE	TRUE	FALSE	FALSE	negative	FALSE	neutral	TRUE	formal	general	محى	محى	وجهه وجماله
4	FALSE	TRUE	TRUE	TRUE	TRUE	positive	FALSE	neutral	TRUE	formal	general	الضياع	الضياع	يبقى في
5	FALSE	FALSE	TRUE	FALSE	FALSE	negative	FALSE	positive	FALSE	formal	general	حزينة	حزينة	ولا لحظة سعادة
6	FALSE	TRUE	TRUE	TRUE	TRUE	negative	FALSE	negative	FALSE	formal	writing style	تلقي	تلقي	لغة جديدة
7	TRUE	TRUE	TRUE	TRUE	TRUE	neutral	FALSE	neutral	FALSE	formal	general	جحد	جحد	كتاب
8	TRUE	TRUE	TRUE	TRUE	TRUE	negative	TRUE	positive	FALSE	formal	author	أسلوبه	أسلوبه	ارتدى رداء ليس
9	TRUE	TRUE	TRUE	TRUE	TRUE	negative	TRUE	positive	FALSE	informal	general	سيئة	سيئة	محور سيء رديء
10	FALSE	TRUE	TRUE	TRUE	TRUE	negative	FALSE	positive	FALSE	informal	writing style	تجذبك	تجذبك	حبكة الرواية

Figure E.2: AMC with all annotation categories

Bibliography

Abdul-Mageed, Muhammad, Mona Diab, and Sandra Kübler (Jan. 2014). “SAMAR: Subjectivity and sentiment analysis for Arabic social media”. en. In: *Computer Speech & Language* 28.1, pp. 20–37. ISSN: 08852308. DOI: 10.1016/j.csl.2013.03.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S088523081300017X (visited on 05/20/2019).

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

— (2014). “Sana: A large scale multi-genre, multi-dialect lexicon for arabic subjectivity and sentiment analysis.” In: *LREC*, pp. 1162–1169.

Abugharsa, Azza (2022). “Metaphor Detection in Poems in Misurata Arabic Sub-Dialect: An LSTM Model”. In.

Aksan, Yesim, Mustafa Aksan, Ahmet Koltuksuz, Taner Sezer, Ümit Mersinli, Umut Ufuk Demirhan, Hakan Yilmazer, Gülsüm Atasoy, Seda Öz, Ipek Yildiz, et al. (2012). “Construction of the Turkish National Corpus (TNC).” In: *LREC*, pp. 3223–3227.

Alhazmi, Samah, William Black, and John McNaught (2013). “Arabic SentiWordNet in relation to SentiWordNet 3.0”. In: *2180* 1266.4, p. 1.

Alhumoud, Sarah O, Mawaheb I Altuwaijri, Tarfa M Albuhairi, and Wejdan M Alohaideb (2015). “Survey on arabic sentiment analysis in twitter”. In: *International Science Index* 9.1, pp. 364–368.

Alkhatib, Manar and Khaled Shaalan (2017). “Natural Language Processing for Arabic Metaphors: A Conceptual Approach”. en. In: *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016*. Ed. by Aboul Ella Hassanien, Khaled Shaalan, Tarek Gaber, Ahmad Taher Azar, and M. F. Tolba. Vol. 533. Cham: Springer International Publishing, pp. 170–181. ISBN: 978-3-319-48307-8. DOI: 10.1007/978-3-319-48308-5_17. URL: http://link.springer.com/10.1007/978-3-319-48308-5_17 (visited on 05/20/2019).

Almuqren, Latifah, Arwa Alzammam, Shahad Alotaibi, Alexandra Cristea, and Sarah Alhumoud (2017). “A review on corpus annotation for Arabic sentiment analysis”. In: *International Conference on Social Computing and Social Media*. Springer, pp. 215–225.

Alowisheq, Areeb, Sarah Alhumoud, Nora Altwairesh, and Tarfa Albuhairi (2016). "Arabic sentiment analysis resources: a survey". In: *Social Computing and Social Media: 8th International Conference, SCSM 2016, Held as Part of HCI International 2016, Toronto, ON, Canada, July 17–22, 2016. Proceedings 8*. Springer, pp. 267–278.

alOwisheq, Areeb, Sarah alHumoud, Nora alTwairesh, and Tarfa alBuhairi (2016). "Arabic Sentiment Analysis Resources: A Survey". en. In: *Social Computing and Social Media*. Ed. by Gabriele Meiselwitz. Vol. 9742. Cham: Springer International Publishing, pp. 267–278. DOI: `10.1007/978-3-319-39910-2_25`. URL: `http://link.springer.com/10.1007/978-3-319-39910-2_25` (visited on 05/20/2019).

Alrefai, Mo'ath, Hossam Faris, and Ibrahim Aljarah (Sept. 2018). "Sentiment analysis for Arabic language: A brief survey of approaches and techniques". en. In: *arXiv:1809.02782 [cs]*. arXiv: 1809.02782. URL: `http://arxiv.org/abs/1809.02782` (visited on 05/20/2019).

Alsayat, Ahmed and Nouh Elmitwally (2020). "A comprehensive study for Arabic sentiment analysis (challenges and applications)". In: *Egyptian Informatics Journal* 21.1, pp. 7–12.

Alshutayri, Areej and Eric Atwell (Apr. 2018). "Arabic dialects annotation using an online game". en. In: *2018 2nd International Conference on Natural Language and Speech Processing (ICNLSP)*. Algiers: IEEE, pp. 1–5. ISBN: 978-1-5386-4543-7. DOI: `10.1109/ICNLSP.2018.8374371`. URL: `https://ieeexplore.ieee.org/document/8374371/` (visited on 05/20/2019).

Alsiyat, Israa (Apr. 2025). "Arabic Metaphor Sentiment Classification Using Semantic Information". In: *International Journal on Cybernetics and Informatics (IJCI)*. URL: `https://ijcionline.com/volume/v14n2`.

Alsiyat, Israa and Scott Piao (May 2020a). "Metaphorical Expressions in Automatic Arabic Sentiment Analysis". English. In: *Proceedings of LREC2020 Conference*. The 12th Edition of the Language Resources and Evaluation Conference, LREC2020 ; Conference date: 11-05-2020 Through 16-05-2020. European Language Resources Association (ELRA), pp. 4911–4916. URL: `https://lrec2020.lrec-conf.org/en/`.

— (May 2020b). "Metaphorical Expressions in Automatic Arabic Sentiment Analysis". English. In: *Proceedings of LREC2020 Conference*. The 12th Edition of the Language Resources and Evaluation Conference, LREC2020 ; Conference date: 11-05-2020 Through 16-05-2020. European Language Resources Association (ELRA), pp. 4911–4916. URL: `https://lrec2020.lrec-conf.org/en/`.

Alsiyat, Israa, Scott Piao, and Mansour Almansour (July 2023). "Arabic Metaphor Corpus (AMC) with Semantic and Sentiment Annotation". English. In: The twelfth International Corpus Linguistics Conference, CL2023 ; Conference date: 03-07-2023 Through 06-07-2023. URL: `https://wp.lancs.ac.uk/cl2023/`.

Aly, Mohamed and Amir Atiya (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.

Ana, Otto Santa (1999). "Like an animal I was treated': Anti-immigrant metaphor in US public discourse". In: *Discourse & society* 10.2, pp. 191–224.

Antoun, Wissam, Fady Baly, and Hazem Hajj (2020). "Arabert: Transformer-based model for arabic language understanding". In: *arXiv preprint arXiv:2003.00104*.

Artstein, Ron (2017). "Inter-annotator agreement". In: *Handbook of linguistic annotation*. Springer, pp. 297–313.

Ayed, Raja, Amina Chouigui, and Bilel Elayeb (Oct. 2018). "A New Morphological Annotation Tool for Arabic Texts". en. In: *2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA)*. Aqaba: IEEE, pp. 1–6. ISBN: 978-1-5386-9120-5. DOI: 10.1109/AICCSA.2018.8612798. URL: https://ieeexplore.ieee.org/document/8612798/ (visited on 05/20/2019).

Al-Ayyoub, Mahmoud, Amal Gigieh, Areej Al-Qwaqenah, Mohammed N Al-Kabi, Bashar Talafhah, and Izzat Alsmadi (2017). "Aspect-Based Sentiment Analysis of Arabic Laptop Reviews". en. In: p. 8.

Al-Ayyoub, Mahmoud, Abed Allah Khamaiseh, Yaser Jararweh, and Mohammed N Al-Kabi (2019). "A comprehensive survey of arabic sentiment analysis". In: *Information Processing & Management* 56.2, pp. 320–342.

Ayyoub, Mahmoud Al, Safa Bani Essa, and Izzat Alsmadi (2015). "Lexicon-based sentiment analysis of Arabic tweets". en. In: *International Journal of Social Network Mining* 2.2, p. 101. ISSN: 1757-8485, 1757-8493. DOI: 10.1504/IJSNM.2015.072280. URL: http://www.inderscience.com/link.php?id=72280 (visited on 05/20/2019).

Badaro, Gilbert, Ramy Baly, Hazem Hajj, Wassim El-Hajj, Khaled Bashir Shaban, Nizar Habash, Ahmad Al-Sallab, and Ali Hamdi (May 2019). "A Survey of Opinion Mining in Arabic: A Comprehensive System Perspective Covering Challenges and Advances in Tools, Resources, Models, Applications, and Visualizations". en. In: *ACM Transactions on Asian and Low-Resource Language Information Processing* 18.3, pp. 1–52. ISSN: 23754699. DOI: 10.1145/3295662. URL: http://dl.acm.org/citation.cfm?doid=3305347.3295662 (visited on 05/20/2019).

Baly, Ramy, Gilbert Badaro, Georges El-Khoury, Rawan Moukalled, Rita Aoun, Hazem Hajj, Wassim El-Hajj, Nizar Habash, and Khaled Shaban (2017). "A characterization study of arabic twitter data with a benchmarking for state-of-the-art opinion mining models". In: *Proceedings of the third Arabic natural language processing workshop*, pp. 110–118.

Bauer, Martin W and Bas Aarts (2000). "Corpus construction: A principle for qualitative data collection". In: *Qualitative researching with text, image and sound: A practical handbook*, pp. 19–37.

Beust, Pierre, Stéphane Ferrari, Vincent Perlerin, et al. (2003). "NLP model and tools for detecting and interpreting metaphors in domain-specific corpora". In: *Proceedings of the Corpus Linguistics 2003 conference*. Vol. 16. Citeseer, pp. 114–123.

Biber, Douglas (1993). "Representativeness in corpus design". In: *Literary and linguistic computing* 8.4, pp. 243–257.

Boudad, Naaima, Rdouan Faizi, Rachid Oulad Haj Thami, and Raddouane Chiheb (Dec. 2018). "Sentiment analysis in Arabic: A review of the literature". en. In: *Ain Shams Engineering Journal* 9.4, pp. 2479–2490. ISSN: 20904479. DOI: 10.1016/j.asej.2017.04.007. URL: https://linkinghub.elsevier.com/retrieve/pii/S2090447917300862 (visited on 05/20/2019).

Castilho, Richard Eckart de, Eva Mujdricza-Maydt, Seid Muhie Yimam, Silvana Hartmann, Iryna Gurevych, Anette Frank, and Chris Biemann (2016). "A web-based tool for the integrated annotation of semantic and syntactic structures". In: *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)*, pp. 76–84.

Chamlertwat, Wilas, Pattarasinee Bhattarakosol, Tippakorn Rungkasiri, and Choochart Haruechaiyasak (2012). "Discovering Consumer Insight from Twitter via Sentiment Analysis." In: *J. Univers. Comput. Sci.* 18.8, pp. 973–992.

Cignarella, Alessandra Teresa, Cristina Bosco, Viviana Patti, and Mirko Lai (May 2018). "Application and Analysis of a Multi-layered Scheme for Irony on the Italian Twitter Corpus TWITTIRÒ". In: *Proceedings of the 11th Language Resources and Evaluation Conference*. Miyazaki, Japan: European Language Resource Association. URL: https://www.aclweb.org/anthology/L18-1664 (visited on 05/20/2019).

Cohen, Jacob (1960). "A Coefficient of Agreement for Nominal Scales". In: *Educational and Psychological Measurement* 20.1, pp. 37–46. DOI: 10.1177/001316446002000104. eprint: https://doi.org/10.1177/001316446002000104. URL: https://doi.org/10.1177/001316446002000104.

Dahou, Abdelghani, Shengwu Xiong, Junwei Zhou, Mohamed Houcine Haddoud, and Pengfei Duan (2016). "Word embeddings and convolutional neural network for arabic sentiment classification". In: *Proceedings of coling 2016, the 26th international conference on computational linguistics: Technical papers*, pp. 2418–2427.

Do, Hai Ha, Pwc Prasad, Angelika Maag, and Abeer Alsadoon (Mar. 2019). "Deep Learning for Aspect-Based Sentiment Analysis: A Comparative Review". en. In: *Expert Systems with Applications* 118, pp. 272–299. ISSN: 09574174. DOI: 10.1016/j.eswa.2018.10.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S0957417418306456 (visited on 05/20/2019).

Do Dinh, Erik-Lân and Iryna Gurevych (2016). "Token-Level Metaphor Detection using Neural Networks". en. In: *Proceedings of the Fourth Workshop on Metaphor in NLP*. San Diego, California: Association for Computational Linguistics, pp. 28–33. DOI: 10.18653/v1/W16-1104. URL: http://aclweb.org/anthology/W16-1104 (visited on 05/20/2019).

Dunn, Jonathan (2013). "What metaphor identification systems can tell us about metaphor-in-language". In: *Proceedings of the First Workshop on Metaphor in NLP*, pp. 1–10.

Duwairi, Rehab M and Islam Qarqaz (2014). "Arabic sentiment analysis using supervised classification". In: *2014 International Conference on Future Internet of Things and Cloud*. IEEE, pp. 579–583.

Elarnaoty, Mohamed, Samir AbdelRahman, and Aly Fahmy (2012). "A machine learning approach for opinion holder extraction in Arabic language". In: *arXiv preprint arXiv:1206.1011*.

Eldin, Ahmad Abdel Tawwab Sharaf (2015). "A Cognitive Metaphorical Analysis of Selected Verses in the Glorius Qu'ran". In: *International Journal of Applied Linguistics and English Literature* 4.2, pp. 193–198.

Elmitwally, Nouh Sabri and Saad Alanazi (2020). "Arabic Corpus for Figurative Sentiment Analysis". In: *International Journal of Advanced Science and Technology* 29.3, p. 14.

Elnagar, Ashraf and Omar Einea (2016). "Brad 1.0: Book reviews in arabic dataset". In: *2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA)*. IEEE, pp. 1–8.

Elnagar, Ashraf, Yasmin S Khalifa, and Anas Einea (2018). "Hotel Arabic-reviews dataset construction for sentiment analysis applications". In: *Intelligent Natural Language Processing: Trends and Applications*. Springer, pp. 35–52.

Eskander, Ramy (2018). "SentiArabic: A Sentiment Analyzer for Standard Arabic". In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*.

Farha, Ibrahim Abu and Walid Magdy (2019). "Mazajak: An Online Arabic Sentiment Analyser". In: *Proceedings of the Fourth Arabic Natural Language Processing Workshop*, pp. 192–198.

Farra, Noura, Elie Challita, Rawad Abou Assi, and Hazem Hajj (2010). "Sentence-level and document-level sentiment mining for arabic texts". In: *2010 IEEE international conference on data mining workshops*. IEEE, pp. 1114–1119.

Faycel, Dakhlaoui (2012). "Food metaphors in Tunisian Arabic proverbs". In: *Rice Working Papers in Linguistics* 3.

Garside, Roger, Geoffrey N Leech, and Anthony Mark Mcenery (1997). *Corpus annotation: linguistic information from computer text corpora*. Taylor and Francis.

Gholami, Samaneh, Iraj Montashery, and Morteza Khorrami (2016). "Metaphors in Arabic and English texts: A case study of translation of metaphors in the English versions of Al-Sahifah Al-Sajjadiyyah". In: *Journal of Literature, Languages and Linguistics* 21, pp. 56–67.

Hadni, Meryeme, Saıd El Alaoui Ouatik, and Abdelmonaime Lachkar (2016). "Word sense disambiguation for Arabic text categorization." In: *Int. Arab J. Inf. Technol.* 13.1A, pp. 215–222.

El-Haj, Mahmoud, Elvis de Souza, Nouran Khallaf, Paul Rayson, and Nizar Habash (2022). "AraSAS: The open source Arabic semantic tagger". In: pp. 23–31.

Al-Harrasi, Abdulla Nasser Khalifa (2001). "Metaphor in (Arabic-into-English) translation with specific reference to metaphorical concepts and expressions in political discourse". PhD thesis. Citeseer.

Hovy, Dirk, Shashank Srivastava, Sujay Kumar Jauhar, Mrinmaya Sachan, Kartik Goyal, Huying Li, Whitney Sanders, and Eduard Hovy (2013). "Identifying metaphorical word use with tree kernels". In: *Proceedings of the First Workshop on Metaphor in NLP*, pp. 52–57.

Al-Huri, Ibrahim et al. (2015). "Arabic Language: Historic and Sociolinguistic Characteristics". In: *English Literature and Language Review* 1.4, pp. 28–36.

Hussein, Doaa Mohey El-Din Mohamed (Oct. 2018). "A survey on sentiment analysis challenges". en. In: *Journal of King Saud University - Engineering Sciences* 30.4, pp. 330–338. ISSN: 10183639. DOI: 10.1016/j.jksues.2016.04.002. URL: https://linkinghub.elsevier.com/retrieve/pii/S1018363916300071 (visited on 05/20/2019).

Itani, Maher (2018). "Sentiment analysis and resources for informal Arabic text on social media". en. PhD. Sheffield Hallam University. DOI: 10.7190/shu-thesis-00118. URL: http://shura.shu.ac.uk/23402/ (visited on 05/20/2019).

Izwaini, Sattar (2003). "A corpus-based study of metaphor in information technology". In: *Corpus Linguistics*.

Jakobson, Roman (1956). "Halle". In: *M.: Fundamentals of Language. S-Gravenhage, Mouton*.

Klebanov, Beata Beigman, Chee Wee Leong, and Michael Flor (2015). "Supervised word-level metaphor detection: Experiments with concreteness and reweighting of examples". In: *Proceedings of the Third Workshop on Metaphor in NLP*, pp. 11–20.

Korayem, Mohammed, David Crandall, and Muhammad Abdul-Mageed (2012). "Subjectivity and Sentiment Analysis of Arabic: A Survey". en. In: *Advanced Machine Learning Technologies and Applications*. Ed. by Aboul Ella Hassanien, Abdel-Badeeh M. Salem, Rabie Ramadan, and Tai-hoon Kim. Vol. 322. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 128–139. DOI: 10.1007/978-3-642-35326-0_14. URL: http://link.springer.com/10.1007/978-3-642-35326-0_14 (visited on 05/20/2019).

Krause, Thomas and Amir Zeldes (2014). "ANNIS3: A new architecture for generic corpus query and visualization". In: *Digital Scholarship in the Humanities* 31.1, pp. 118–139.

Krennmayr, Tina and Gerard Steen (2017). "Vu amsterdam metaphor corpus". In: *Handbook of Linguistic Annotation*. Springer, pp. 1053–1071.

Lakoff, George and Mark Johnson (1980). "Conceptual metaphor in everyday language". In: *The journal of Philosophy* 77.8, pp. 453–486.

Leech, Geoffrey Neil (1992). "100 Million Words of English: The British National Corpus (BNC)". In.

Liu, Bing (2012). "Sentiment analysis and opinion mining". In: *Synthesis lectures on human language technologies* 5.1, pp. 1–167.

Maamouri, Mohamed, Ann Bies, Tim Buckwalter, and Wigdan Mekki (2004). "The penn arabic treebank: Building a large-scale annotated arabic corpus". In: *NEMLAR conference on Arabic language resources and tools*. Vol. 27. Cairo, pp. 466–467.

El-Makky, Nagwa, Khaled Nagi, Alaa El-Ebshihy, Esraa Apady, Omneya Hafez, Samar Mostafa, and Shimaa Ibrahim (2014). "Sentiment analysis of colloquial Arabic tweets". In: *ASE BigData/SocialInformatics/PASSAT/BioMedCom 2014 Conference, Harvard University*, pp. 1–9.

Manning, Christopher D, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky (2014). "The Stanford CoreNLP natural language processing toolkit". In: *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pp. 55–60.

Mao, Rui, Chenghua Lin, and Frank Guerin (2018). "Word embedding and wordnet based metaphor identification and interpretation". In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1222–1231.

Marcus, Mitch, Beatrice Santorini, and Mary Ann Marcinkiewicz (1993). "Building a large annotated corpus of English: The Penn Treebank". In: *Computational linguistics* 19.2, pp. 313–330.

El-Masri, Mazen, Nabeela Altrabsheh, Hanady Mansour, and Allan Ramsay (2017). "A web-based tool for Arabic sentiment analysis". In: *Procedia Computer Science* 117, pp. 38–45.

Medhat, Walaa, Ahmed Hassan, and Hoda Korashy (Dec. 2014). "Sentiment analysis algorithms and applications: A survey". en. In: *Ain Shams Engineering Journal* 5.4, pp. 1093–1113. ISSN: 20904479. DOI: 10.1016/j.asej.2014.04.011. URL: https://linkinghub.elsevier.com/retrieve/pii/S2090447914000550 (visited on 05/20/2019).

Meskaldji, Khouloud, Salim Chikhi, and Imene Bensalem (Oct. 2018). "A New Multi Varied Arabic Corpus". en. In: *2018 3rd International Conference on Pattern Analysis and Intelligent Systems (PAIS)*. Tebessa: IEEE, pp. 1–5. ISBN: 978-1-5386-4238-2. DOI: 10.1109/PAIS.2018.8598524. URL: https://ieeexplore.ieee.org/document/8598524/ (visited on 05/20/2019).

Miller, George A (1995). "WordNet: a lexical database for English". In: *Communications of the ACM* 38.11, pp. 39–41.

Mohamed, Ghada, Amanda Potts, and Andrew Hardie (2013). "AraSAS: a semantic tagger for Arabic". In.

Mohammad, Saif, Ekaterina Shutova, and Peter Turney (2016). "Metaphor as a medium for emotion: An empirical study". In: *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*, pp. 23–33.

Obaidat, Islam, Rami Mohawesh, Mahmoud Al-Ayyoub, Mohammad AL-Smadi, and Yaser Jararweh (Nov. 2015). "Enhancing the determination of aspect categories and their polarities in Arabic reviews using lexicon-based approaches". en. In: *2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*. Amman, Jordan: IEEE, pp. 1–6. ISBN: 978-1-4799-7442-9. DOI: 10.1109/AEECT.2015.7360595. URL: http://ieeexplore.ieee.org/document/7360595/ (visited on 05/20/2019).

Özbal, Gözde, Carlo Strapparava, and Serra Sinem Tekiroglu (2016). "PROMETHEUS: A Corpus of Proverbs Annotated with Metaphors". In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pp. 3787–3793.

Özbal, Gözde, Carlo Strapparava, Serra Sinem Tekiroglu, and Daniele Pighin (2016). "Learning to identify metaphors from a corpus of proverbs". In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 2060–2065.

Peng, Ying, Chang Su, and Yijiang Chen (2018). "Chinese metaphor sentiment analysis based on attention-based LSTM". In: *2018 Tenth International Conference on Advanced Computational Intelligence (ICACI)*. IEEE, pp. 478–483.

Pontiki, Maria, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, AL-Smadi Mohammad, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, et al. (2016). "Semeval-2016 task 5: Aspect based sentiment analysis". In: *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, pp. 19–30.

Rai, Sunny, Shampa Chakraverty, and Devendra K Tayal (2016a). "Supervised metaphor detection using conditional random fields". In: *Proceedings of the Fourth Workshop on Metaphor in NLP*, pp. 18–27.

— (2016b). "Supervised Metaphor Detection using Conditional Random Fields". en. In: *Proceedings of the Fourth Workshop on Metaphor in NLP*. San Diego, California: Association for Computational Linguistics, pp. 18–27. DOI: 10.18653/v1/W16-1103. URL: http://aclweb.org/anthology/W16-1103 (visited on 05/20/2019).

Raii, Jalal (2009). "Metaphor in day-to-day Arabic speech: a conceptual approach". In: *Tishreen University Journal for Research and Scientific Studies-Arts and Humanities Series* 31.1, pp. 175–193.

Refaee, Eshrag and Verena Rieser (2014). "An arabic twitter corpus for subjectivity and sentiment analysis." In: *LREC*, pp. 2268–2273.

Rentoumi, Vassiliki (2012). "Sentiment Analysis of Metaphorical Language". PhD thesis. Institute of Informatics.

Rentoumi, Vassiliki, George Giannakopoulos, Vangelis Karkaletsis, and George A Vouros (2009). "Sentiment analysis of figurative language using a word sense disambiguation approach". In: *Proceedings of the International Conference RANLP-2009*, pp. 370–375.

Reyes, Antonio and Paolo Rosso (2012). "Building corpora for figurative language processing: The case of irony detection". In: *Proceedings of the 4th International*

Workshop on Corpora for Research on Emotion Sentiment & Social Signals (in conjunction with LREC 2012). Citeseer, pp. 94–98.

Ruder, Sebastian, Parsa Ghaffari, and John G Breslin (2016). "Insight-1 at semeval-2016 task 5: Deep learning for multilingual aspect-based sentiment analysis". In: *arXiv preprint arXiv:1609.02748*.

Rushdi-Saleh, Mohammed, M Teresa Martín-Valdivia, L Alfonso Ureña-López, and José M Perea-Ortega (2011a). "Bilingual experiments with an arabic-english corpus for opinion mining". In: *Proceedings of the International Conference Recent Advances in Natural Language Processing 2011*, pp. 740–745.

— (2011b). "OCA: Opinion corpus for Arabic". In: *Journal of the American Society for Information Science and Technology* 62.10, pp. 2045–2054.

Schmidt, Klaus M (1988). "Der Beitrag der begriffsorientierten Lexicographie zue systematischen Erfassung von Sprachwandel und das Begriffwörterbuch zur wdh. Epik". In: *Mittelhochdeutches Wörterbuch in der Diskussion. Tubingen*, pp. 25–49.

Schuler, Karin Kipper (2005). *VerbNet: A broad-coverage, comprehensive verb lexicon*. University of Pennsylvania.

Shiha, Mohammed and Serkan Ayvaz (2017). "The effects of emoji in sentiment analysis". In: *Int. J. Comput. Electr. Eng.(IJCEE.)* 9.1, pp. 360–369.

Shutova, Ekaterina (2010). "Models of metaphor in NLP". In: *Proceedings of the 48th annual meeting of the association for computational linguistics*, pp. 688–697.

Shutova, Ekaterina, Douwe Kiela, and Jean Maillard (2016). "Black holes and white rabbits: Metaphor identification with visual features". In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 160–170.

AL-Smadi, M., M. Al-Ayyoub, H. Al-Sarhan, and Y. Jararweh (Dec. 2015). "Using Aspect-Based Sentiment Analysis to Evaluate Arabic News Affect on Readers". In: *2015 IEEE/ACM 8th International Conference on Utility and Cloud Computing (UCC)*, pp. 436–441. DOI: 10.1109/UCC.2015.78.

Al-Smadi, Mohammad, Omar Qawasmeh, Mahmoud Al-Ayyoub, Yaser Jararweh, and Brij Gupta (July 2018). "Deep Recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews". en. In: *Journal of Computational Science* 27, pp. 386–393. ISSN: 18777503. DOI: 10.1016/j.jocs.2017.11.006. URL: https://linkinghub.elsevier.com/retrieve/pii/S1877750317305252 (visited on 05/20/2019).

Al-Smadi, Mohammad, Omar Qawasmeh, Bashar Talafha, and Muhannad Quwaider (Aug. 2015). "Human Annotated Arabic Dataset of Book Reviews for Aspect Based Sentiment Analysis". en. In: *2015 3rd International Conference on Future Internet of Things and Cloud*. Rome, Italy: IEEE, pp. 726–730. ISBN: 978-1-4673-8103-1. DOI: 10.1109/FiCloud.2015.62. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7300895 (visited on 05/20/2019).

Al-Smadi, Mohammad, Bashar Talafha, Mahmoud Al-Ayyoub, and Yaser Jararweh (Mar. 2018). "Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews". en. In: *International Journal of Machine Learning and Cybernetics*. ISSN: 1868-8071, 1868-808X. DOI: 10.1007/s13042-018-0799-4. URL: http://link.springer.com/10.1007/s13042-018-0799-4 (visited on 05/20/2019).

Soliman, Abu Bakr, Kareem Eissa, and Samhaa R El-Beltagy (2017). "Aravec: A set of arabic word embedding models for use in arabic nlp". In: *Procedia Computer Science* 117, pp. 256–265.

Stenetorp, Pontus, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii (2012). "BRAT: a web-based tool for NLP-assisted text annotation". In: *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, pp. 102–107.

Strapparava, Carlo and Rada Mihalcea (2007). "Semeval-2007 task 14: Affective text". In: *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pp. 70–74.

Al-Subaihin, Afnan A, Hend S Al-Khalifa, and AbdulMalik S Al-Salman (2011). "A proposed sentiment analysis tool for modern arabic using human-based computing". In: *Proceedings of the 13th international conference on information integration and web-based applications and services*, pp. 543–546.

Thelwall, Mike, Kevan Buckley, Georgios Paltoglou, Di Cai, and Arvid Kappas (2010). "Sentiment strength detection in short informal text". In: *Journal of the American society for information science and technology* 61.12, pp. 2544–2558.

Thet, Tun Thura, Jin-Cheon Na, and Christopher SG Khoo (2010). "Aspect-based sentiment analysis of movie reviews on discussion boards". In: *Journal of information science* 36.6, pp. 823–848.

Tsvetkov, Yulia, Leonid Boytsov, Anatole Gershman, Eric Nyberg, and Chris Dyer (2014). "Metaphor detection with cross-lingual model transfer". In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 248–258.

Al-Twairesh, Nora, Abeer Al-Dayel, Hend Al-Khalifa, Maha Al-Yahya, Sinaa Alageel, Nora Abanmy, and Nouf Al-Shenaifi (2016). "MADAD: a readability annotation tool for Arabic text". In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pp. 4093–4097.

Al-Twairesh, Nora, Hend Al-Khalifa, and AbdulMalik Al-Salman (Nov. 2014). "Subjectivity and sentiment analysis of Arabic: Trends and challenges". en. In: *2014 IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA)*. Doha, Qatar: IEEE, pp. 148–155. ISBN: 978-1-4799-7100-8. DOI: 10.1109/AICCSA.2014.

7073192. URL: http://ieeexplore.ieee.org/document/7073192/ (visited on 05/20/2019).

Veale, Tony (2016). "Round Up The Usual Suspects: Knowledge-Based Metaphor Generation". en. In: *Proceedings of the Fourth Workshop on Metaphor in NLP*. San Diego, California: Association for Computational Linguistics, pp. 34–41. DOI: 10.18653/v1/W16-1105. URL: http://aclweb.org/anthology/W16-1105 (visited on 05/20/2019).

Wawer, Aleksander, Agnieszka Mykowiecka, and Jana Kazimierza (2017). "Detecting Metaphorical Phrases in the Polish Language." In: *RANLP*, pp. 772–777.

Wilks, Yorick (1978). "Making preferences more active". In: *Artificial intelligence* 11.3, pp. 197–223.

Wilks, Yorick, Adam Dalton, James Allen, and Lucian Galescu (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.

Wissler, Lars, Mohammed Almashraee, Dagmar Monett Díaz, and Adrian Paschke (2014). "The Gold Standard in Corpus Annotation." In: *IEEE GSC* 21.

Zaghouani, Wajdi, Nizar Habash, Houda Bouamor, Alla Rozovskaya, Behrang Mohit, Abeer Heider, and Kemal Oflazer (2015). "Correction Annotation for Non-Native Arabic Texts: Guidelines and Corpus". en. In: *Proceedings of The 9th Linguistic Annotation Workshop*. Denver, Colorado, USA: Association for Computational Linguistics, pp. 129–139. DOI: 10.3115/v1/W15-1614. URL: http://aclweb.org/anthology/W15-1614 (visited on 05/20/2019).

Zeroual, Imad and Abdelhak Lakhouaja (June 2016). "A new Quranic Corpus rich in morphosyntactical information". en. In: *International Journal of Speech Technology* 19.2, pp. 339–346. ISSN: 1381-2416, 1572-8110. DOI: 10.1007/s10772-016-9335-7. URL: http://link.springer.com/10.1007/s10772-016-9335-7 (visited on 05/20/2019).

— (2018). "Arabic Corpus Linguistics: Major Progress, but Still a Long Way to Go". en. In: *Intelligent Natural Language Processing: Trends and Applications*. Ed. by Khaled Shaalan, Aboul Ella Hassanien, and Fahmy Tolba. Vol. 740. Cham: Springer International Publishing, pp. 613–636. ISBN: 978-3-319-67055-3. DOI: 10.1007/978-3-319-67056-0_29. URL: http://link.springer.com/10.1007/978-3-319-67056-0_29 (visited on 05/20/2019).