

DAICT: A Dialectal Arabic Irony Corpus Extracted from Twitter

Ines Abbes, Wajdi Zaghouni, Oaima El-Hardlo, Faten Ashour

College of Humanities and Social Sciences, Hamad Bin Khalifa University

Doha, Qatar

{iabbesepmnasri, wzaghouni, oelhardlo}@hbku.edu.qa, fashour@mail.hbku.edu.qa

Abstract

Identifying irony in user-generated social media content has a wide range of applications; however to date Arabic content has received limited attention. To bridge this gap, this study builds a new open domain Arabic corpus annotated for irony detection. We query Twitter using irony-related hashtags to collect ironic messages, which are then manually annotated by two linguists according to our working definition of irony. Challenges which we have encountered during the annotation process reflect the inherent limitations of Twitter messages interpretation, as well as the complexity of Arabic and its dialects. Once published, our corpus will be a valuable free resource for developing open domain systems for automatic irony recognition in Arabic language and its dialects in social media text.

Keywords: Arabic Dialects, Irony, Twitter, Corpus Generation

1. Introduction

Irony detection has a broad range of applications in various research areas, including sentiment analysis, text mining and author profiling (Van Hee, 2017). Nowadays, social-media platforms allow users to express and share their opinions, attitudes and beliefs instantaneously. Users tend to employ figurative language devices, such as irony and sarcasm, to achieve different communication purposes (Hernández Farías, 2017). This user-generated content represents a rich source of valuable data. Handling this data is a challenging task in Natural Language Processing, particularly when dealing with ironic statements, with users intending to convey a message that goes beyond the literal meaning.

Current scholarship is yet to reach an agreement on a universal definition of the concept of irony (Van Hee, 2017; Hernández Farías, 2017). Various definitions have been proposed (Grice et al., 1975; Wilson and Sperber, 1992; Kumon-Nakamura et al., 1995; Attardo, 2000; Giora et al., 2013) and irony is broadly defined as a rhetorical device used to express the opposite of what is literally meant (Grice et al., 1975). This study investigates the use of irony in Arabic Twitter messages. For the purpose of our analysis, we do not distinguish between irony and sarcasm, thereby using irony as an umbrella term.

Annotated corpora for irony detection are now available in many languages; however, to the best of our knowledge, there is no publicly and freely available irony corpus for Arabic language or its dialects (Rosso et al., 2018; Zaghouni, 2014). To bridge this gap, this research introduces a new irony corpus of Dialectal Arabic extracted from Twitter: DAICT¹.

This paper is structured as follows. The following section gives an overview of the existing work on irony corpora generation. Section 3 presents the data collection process as well as annotation guidelines and challenges. Section 4 elaborates the corpus analysis. We conclude this paper in

Section 5 with some suggestions for future research.

2. Related Work

Recent years have witnessed a surge in the availability of corpora for the Arabic language with focus on dialectal Arabic such as the corpora created by (Bouamor et al., 2018; Zaghouni and Charfi, 2018; Maamouri et al., 2010; Zaghouni et al., 2014; Bouamor et al., 2015). Moreover, we observed a growing interest in collecting and processing Arabic user-generated content from social media sources as in the projects discussed in (Rangel et al., 2019a; Rangel et al., 2019b; Atanasova et al., 2018; Barrón-Cedeño et al., 2018). Similarly, irony and sarcasm detection has recently drawn a significant attention in computational linguistics as a standard text classification problem (Joshi et al., 2017). Multiple annotated corpora and their discussions have been published for the English language (Davidov et al., 2010; Filatova, 2012; Rajadesingan et al., 2015). Building on the approaches proposed in these studies, corpora have been created for other languages, for example, Italian (TWIT-TIRÒ (Cignarella et al., 2017), ironITA (Cignarella et al., 2018)); French (Karoui et al., 2017a); and Chinese (Tang and Chen, 2014; Lin and Hsieh, 2016). In comparison, fewer studies considered in detail irony detection in Arabic. Although Arabic is the third most spoken language in the world and is widely used online in social media, the only corpus on irony-detection – SOUKHRIA corpus (Karoui et al., 2017b) – remains available for a limited group of researchers and has not been released to public yet.

In order to build corpora of ironic text, previous studies have used a variety of sources: Twitter, Facebook, Amazon.com, blogs, newspaper sites, etc. Some studies have defined their source by way of focusing on the sites, dedicated to ironic or sarcastic discourse (Barbieri et al., 2014). Other researchers experiment with different methods of identifying ironic content within a non-specific context. To take a typical case, Davidov et al. (2010), González-Ibáñez et al. (2011), and Rajadesingan et al. (2015) collect only the tweets that include hashtags expressing sarcasm and consider these hashtags as ground-truth labels. This ap-

¹The dataset can be downloaded at <https://www.hbku.edu.qa/en/DAICT>

proach can be applied to Arabic Twitter, where hashtags are very commonly used; for this reason, hashtags expressing irony in Arabic have been used to collect data for the SOUKHRIA corpus (Karoui et al., 2017b). Indeed, some tweets in the corpus include a combination of hashtags for irony, sarcasm or satire: it appears that tweeters do not always distinguish between these notions. For instance, in tweet (1) the author employs two hashtags: *#سخرية* (“#irony”) and *#تهكم* (“#sarcasm”):

- (1) قررت أعيش عند الهنود الحمر كل ليلة مشويات و طقطقة وسعة صدر مالهم ب الهم والمشاكل ، و على رأسهم ريشه بعد .. *#تهكم* *#سخرية*
I decided to live with the Red Indians: barbecue and drumming every night ! They have no worries or concerns; and on top of that, they have feathers on the top of their caps like kings do! #sarcasm #irony²

While hashtags offer a convenient indication of an author’s stance, researchers highlight an important limitation: some tweets containing #sarcasm are *about* sarcasm (e.g. “I love #sarcasm”), while the tweets themselves are not sarcastic (Davidov et al., 2010). For this reason, in addition to digital methods of identifying ironic content, researchers have been experimenting with manual annotation of the data. Most published work has relied on manual annotation by crowd workers or linguists. Reyes and Rosso (2011), Filatova (2012), and Walker et al. (2012) employ crowdsourcing platforms and decide whether a text is ironic or not by considering inter-annotator agreement. However, crowd workers are not linguistic specialists and receive no specialized training; as a result, the annotation could be inaccurate. To address this problem, Karoui et al. (2017a) employ professional linguists and researchers in computational linguistics to label a multilingual irony corpus. This approach, while advancing the linguistic accuracy of annotating, brings to the fore the need for a clear definition of an elusive concept irony.

The complexity of the concept of irony, as well as the need for a nuanced and careful manual annotation, is amply demonstrated by the existing Arabic corpus, SOUKHRIA (Karoui et al., 2017b). In this corpus, tweets were collected using a set of politicians’ names as keywords. Then, tweets including one of the hashtags *#مسخرية* / *#سخرية* (“#irony”) or *#استهزاء* / *#تهكم* (“#sarcasm”) were automatically labeled as ironic, and the rest as non ironic. It is not clear whether Karoui et al. (2017b) used any manual annotation after the publication of the research paper. The authors recognize the limitations of their corpus and accept that “wrong pre-annotations” need to be manually corrected. They assume that the error will be “around 3%” (Karoui et al., 2017b). In order to test the accuracy of this assumption for our project, we manually annotated a random selection of the tweets used in SOUKHRIA. The initial examination suggests that at least 8% of the tweets were misclassified.

SOUKHRIA corpus offers researchers a unique opportu-

nity to analyse in detail the results of applying methods – which have been devised for irony corpora-building in other languages – to Arabic. In several ways, the findings agree with previous research for other languages. Firstly, the corpus confirms that irony-related hashtags do not always correspond to an ironic content. Secondly, irony proves to be an elusive and subjective concept in Arabic, too.

Typically, tweet (2)³, which was chosen by Karoui et al. (2017b) as an example of a non-ironic message tagged with the hashtag *#سخرية* (“#irony”), exemplifies the subjectivity of manual annotation.

- (2) رسالة للاخوان انتم مش هترحمونا من مظاهراتكم
الحاشدة بتاعت كل جمعة دي مرسي مش راجع
والله *#سخرية*

A message for extremist Muslims. You are not going to exempted us from your gatherings every Friday. Morsi will not come back I swear #irony (Karoui et al. (2017b) translation)

A message for Muslim Brotherhood supporters: Stop giving us a headache with your mass protests every Friday. Morsi is gone forever; I swear to God! #irony (Our translation)

Our version of translation suggests that the tweet was ironic. This example highlights the importance of knowing the dialect used in the text (in this case, Egyptian) and the political context of the message.

Context is essential for an accurate interpretation of the stance of the author; and on Twitter, irony hashtags often provide the only clue. Notably, once Karoui et al. (2017b) removed the hashtags which signaled irony, some messages immediately became ambiguous (e.g. 3).

- (3) معلومة هامة:كثير منا مش عارف ان *#مرسي* لسه
محطوط على قوائم المنوعين من السفر
Important piece of information: many of us are unaware of the fact that #Morsi is still on travel-ban lists!

In this case, without the hashtag for irony it is not clear whether the author states a fact or not. Another example (4) further demonstrates the importance of a hashtag as an interpretative clue:

- (4) أالجالية النوبية في مصر قامت بدور عظيم في الحرب
The Nubian community in Egypt played a great role in the war.

Without a hashtag for irony, the intention of the tweeter is impossible to ascertain. These examples strongly suggest that removing irony hashtags in the final dataset could subvert subsequent manual annotation. Our study therefore proposes to keep the hashtags as an important context marker for further annotation.

The existing irony-detection corpora provide an important foundation for our research into irony recognition in Arabic. This study aims to produce a more nuanced dataset, which will embrace the dialects of the region as well as

²All translations in this paper have been revised by accredited translators.

³referred to as tweet (12) in (Karoui et al., 2017b)

the Modern Standard Arabic (MSA), and navigate the challenges and limitations identified in previous studies.

3. Corpus Generation

3.1. Data Collection

To collect the maximum number of ironic tweets, we chose to use irony hashtags. We did not use any keywords since we wanted our corpus to cover as many domains as possible. We manually gathered 100 hashtags, in Arabic, English and French, which are generally used by tweeters writing in different Arabic dialects to express irony and sarcasm. Using these hashtags and “Arabic” as a language parameter, we could collect around 1.47 M tweets over the period spanning from March 31st, 2012 to March 1st, 2019. Given the large volume of data, we opted to work on the top four most used hashtags; #سخرية (“#irony”), #مسخرة (“#ironical”), #استهزاء (“#mockery”), and #sarcasm and which could be placed anywhere in the tweet. Then, we applied automatic filtering to remove retweets, duplicates, username mentions, tweets with URLs and pictures, and tweets shorter than 70 characters. Additionally, we removed tweets containing more than four hashtags. The resultant corpus consists of a set of 5,358 tweets written in MSA, dialectal Arabic and a mix of both.

When analyzing the collected tweets, we noticed that some users embed hashtags within their messages although their intention was not necessarily ironic. For this reason manual annotation was deemed mandatory to judge whether the text was ironic or not. This approach allows us to collect both ironic and non ironic tweets and helps us provide an estimation of the credibility of irony hashtags in Arabic tweets.

3.2. Annotation Guidelines

In SOUKHRIA corpus, Karoui et al. (2017b) rely on hashtags as ground-truth for labeling ironic tweets and they confirm that manual verification remains indispensable. Our team includes two language specialists in Arabic who come from two different regions and represent different dialects of Arabic. To provide our annotators with a clue to the author’s intention, we did not remove hashtags. As the first step, the annotators were given simple instructions to label a random sample of 100 tweets as “*Ironic*” or “*Not Ironic*”. Using Fleiss’ kappa (Fleiss, 1971), we obtained an inter-annotator agreement of 0.37 between our two annotators, which indicates a “fair agreement” (Landis and Koch, 1977). This relatively low rate reflects the subjectivity of human annotation.

During the discussion of obtained results, it emerged that in general the disagreement could be explained by the lack of immediate context or complete structure. To take a typical case, tweet (5) cannot be conclusively interpreted as ironic or not. To address this problem, we have added “*Ambiguous*” label as an umbrella reference to all the cases where the annotators cannot individually decide with certainty whether a tweet is ironic or not. Besides, in order to reach a better inter-annotator agreement and to move on to the next stage of annotation, the annotators had to consider developing a working definition of irony. At that stage, for

- (5) اي مثل ما وصلني عربي انجليزي واسلامية
واجتماعيات وعلوم فقط الرسم والبدنيه وخدمة
اجتماع #سخرية
Yes, as I have heard; Arabic, English, Islamic Studies, Social Sciences, and Science; only Fine Arts, PE, and Community Services #irony

the purpose of our annotation, we agreed to assume the definition of irony as an “*evaluative expression whose polarity (i.e., positive, negative) is inverted between the literal and the intended evaluation, resulting in an incongruence between the literal evaluation and its context*” (Van Hee et al., 2016).

With this definition in mind, the full dataset has been individually annotated by two linguists. The inter-annotator agreement between them increased to 0.92. The remaining disagreement was resolved by giving the disputed tweets to a third annotator for further analysis and a group discussion. For instance, tweets labeled as “*Ambiguous*” by any of the annotators were re-examined and discussed by the team of annotators. If consensus can be reached, these tweets are assigned to one of the categories, “*Ironic*” or “*Not Ironic*”. If the three annotators could not agree on how to label a tweet, it was kept in the final corpus as “*Ambiguous*”. These tweets provide an invaluable starting point for an in-depth analysis of annotation challenges (see Section 3.3). At this stage the corpus includes only the tweets for which the agreement was established.

To recap, preliminary annotation and subsequent agreement on the definition of the concept of irony allowed our annotators to improve the inter-annotator agreement.

3.3. Annotation Challenges

Along with the elusiveness of the concept of irony, the process of manual annotation revealed the following additional challenges:

- *Dialects*: Most, if not all, of the Arabic speakers understand MSA. However some dialects are not mutually intelligible. For instance in the Gulf region, Arabic speakers typically report difficulties in understanding the Maghrebi dialect. To take a typical case, tweet (6) is written in Moroccan dialect, and both of our annotators were not able to understand its meaning. Some words like قشعو (“understand”), ديال (“of”), ولايني (“become”), حرامي (“devious”), الدرري (“child”), and ديرها (“trick”) are specific to the Moroccan dialect. In order to understand the tweet, we asked a Moroccan native speaker to interpret it for us.

- (6) اه اه اه غير الادكياء اللي قشعو الخطه ديال ايهاب
ولايني حرامي هاد الدرري ديرها بيا انا هاي هاي
#sarcasm
Ah Ah Ah ... Only clever people will understand Ehab’s plan. This child is devious, but he can not trick me! Hay Hay #sarcasm

- *Confusion between irony and other figurative language devices*: Irony is often confused with other figurative language devices, such as humor or metaphor. Most people do not distinguish between the use of humor in ironic statements and jokes. In tweet (7), the writer uses a joke to make an ironic statement about the social problem of the reluctance of young men to marry. However in example (8) and despite the use of the hashtag #سخرية (“#irony”), the tweet is just a simple joke about a girl mosquito.

(7) #سخرية نكتة أطفال تحاكي الواقع المرير...
ديناصور طلب من ديناصوره (صورتها) قالت هذا
اللي خلانا نقرض...(!?)

#Irony, A children’s joke reflects our bitter reality: A male dinosaur asked for a female dinosaur’s personal (photo), and she replied: “This is what made us go extinct ...(!?)”

(8) ناموسه اول مرت (ن) تطير يوم رجعت قال لها ابوها
وشو شعورك؟ قالت والله حماس كل الناس يصفقون(ن)
لي #سحقاً #سخرية #نكت ...

A girl mosquito went out on her first flight. When she came back, her father asked her: “How did it feel?” She said: “I swear it was exciting as everyone around was clapping.” #hell #irony #jokes ...

- *Transliteration of English*: Using the Arabic alphabet to write English words is common in social media platforms and is occasionally used for humor or irony purposes. Arabic and English have different sounds and with the variety of dialects used in tweets, it may be difficult to identify transliterated words. Tweet (9) includes two transliterated words: كوتش *kuwtiṣ*⁴ (“coach”) and نقتف *niqatif* (“negative”). The word كوتش is relatively easy to be recognized as a transliteration of “coach” because the English sounds in this word generally correspond to Arabic letters. However the other transliterated word نقتف is more obscure, since both sounds /g/ and /v/ do not have corresponding letters in Arabic. So, the tweeter replaced them with *q* and *f* respectively. As a result the transliterated word becomes open to misinterpretation.

(9) مساعد الكوتش يوجه الكوتش وهو بدوره سيوجه
اللاعيبين ، الكوتش طلع « نقتف » يا جماعة :...!
#الأهلي_الهلال #سخرية

The assistant coach instructs the coach, who in turn will instruct players, the coach turned to be “negative”, guys :) ...! #Al_Ahli_Al_Hilal #irony

- *User’s cultural background*: The user’s culture defines his perception and the way his opinions are expressed.

In the context of irony, this is more complicated, since the annotator needs to understand the cultural background behind each statement. For example in tweet (10), the reader needs to know that in Saudi Arabia women are forbidden from marrying men outside their tribe.

(10) #المجتمع_السعودي_متناقض الرجال مايتزوج
حبيبته بس عادي يتزوج حبيبة غيره يعني أهم
شيء موب هي اللي حباها #سخرية

#Saudi_society_is_self-contradictory: A man does not marry his beloved, but it is normal that he marries another man’s beloved. So, the most important thing: she is not the one he loved ! #irony

- *Implicit irony*: Situational irony can sometimes be difficult to identify. Detecting the incongruence between two situations is not always straightforward, especially in the absence of “evaluative remarks” (Hernández Farías, 2017), facial expression, and voice tone in online texts. This is illustrated by the the following example:

(11) أخبرت القوم يوماً أن لا أحد عبد لأحد، وأن كلاً
منا حر.. فقالوا صح القول، وتوجوني ملكاً عليهم.
#سخرية

I told the people one day that no one is a slave to anyone, and that each of us is free ... They said that is true, and crowned me their king. #irony

- *Absence of Arabic diacritics*: Arabic diacritical marks are used to facilitate the pronunciation, to disambiguate the meaning of some words, and for إعراب (“parsing”). They are not frequently used in online text and their absence can tremendously affect and change the sentence’s meaning. For instance in tweet (12), خيار *xayaAr* (“option”) and خيار *xiyaAr* (“cucumber”) are written the same without diacritics, although their meaning is completely different.

(12) مصيبة كبرى : عندما لا يفرق السياسي بين الخيار
الاستراتيجي والخيار باللبن ! عندها لا ترى أي فرق
بين السلطة والسلطة. #سخرية

A major calamity: When a politician is unable to tell the difference between the strategic option and the dish of cucumber with yogurt! It made me feel that power and salad-making are the same. #irony

- *Lack of conversational context*: In some tweets, the writer’s intention is not clear because either the the conversational context is missing or the sentence is not complete. In the following tweet (13), our annotators were not able to establish whether it is ironic or not, because the message seems to belong to a chain of interactions as the author was referring to something that was previously mentioned by using the word هيك *hyk* (“this way”).

⁴Arabic transliteration is presented in the Habash-Soudi-Buckwalter scheme (Habash et al., 2007).

- (13) مسم يعني ما حدا كان حيحكي هيك بس هسا ممكن
اي حدا يفكر هيك ##سخرية ##فكاهة
Mmmm... So no one could have said it this way, but
now any one can think this way ! #irony #humor

To recap, our annotators faced two kinds of challenges, those related to Arabic (e.g. absence of diacritics) and those generally occurring in social media texts regardless of the language (e.g. lack of context).

4. Results and Discussion

Our corpus covers multiple domains and is not restricted to specific topics. In our dataset, the majority of ironic Arabic tweets refers to international affairs, football and social issues. According to the annotators' notes, most of the tweets (over 70%) are written in the Gulf and Egyptian dialects. This reflects the fact that the top four countries ranked by number of tweets, according to the Arab Social Media Report, are Saudi Arabia, Kuwait, Egypt and the UAE respectively⁵. Further analysis of the corpus will allow us to elaborate the distribution of the dialects in more detail.

The tweeters use either verbal or situational irony. When verbal irony is used, authors sometimes employ very specific dialectal terms to express it; furthermore, the sentences are not always complete, which presents an additional challenge for our annotators. In cases of situational irony, where there is no overt verbal contrast, irony is often expressed through an incongruence between two situations (Lucariello, 1994). This type of implicit irony, combined with limited context or insufficient information about users' cultural background, is harder to detect and requires further investigation in order to develop a digital irony-detection tool.

Unlike Karoui et al. (2017b), we kept the hashtags for the annotators' reference. We observed that tweeters tend to insert hashtags to clarify their message. Tweet (14) below ponders this need for an explicit statement of intention.

- (14) هو الواحد يا جدعان لازم يكتب في آخر التويته
##فكاهة ##سخرية ##تمت عشان الناس تفهم إنه
بيتبريق؟ :

Hey folks, does one have to end a tweet with #irony, #humor, #ends to let readers know one is being sarcastic? :-)

Using the following irony-related hashtags: ##سخرية (“#irony”), ##مسخرة (“#ironical”), ##استهزاء (“#mockery”), and #sarcasm, we built a dataset of 5,358 Arabic messages posted on Twitter. As a result of individual analysis and follow-up discussions, our annotators have labeled 4,809 (89.75%) as “Ironic”, 435 (8.12%) as “Not Ironic”, and 114 (2.13%) as “Ambiguous”. According to the annotation results, almost 90% of the collected tweets were confirmed to be ironic. This suggests that relying on hashtags as ground-truth labels introduces around 10% of noise to

the dataset (Kunneman et al., 2015). These outcomes support our hypothesis about the relative reliability of irony-related hashtags, and the importance of manual annotation for obtaining a high-quality dataset. Even after group discussions, our annotators could not reach consensus on 2% of the collected tweets. These messages were open to different interpretations of whether irony is present or not, and remained classified as ambiguous.

At this stage, our focus has been on identifying the features of non ironic tweets tagged with an irony-related hashtag. In some cases, as mentioned above, irony hashtags were not used accurately because users confused irony with other figurative language devices, such as humor or metaphor (see tweet (8) above). In other cases, users commented on the *topic* of irony, and used irony-related hashtags to tag the topic for other users, not to clarify their messages or intention. For instance in tweet (15), it is clear that the hashtags ##سخرية (“#irony”) and ##الاستهزاء (“#mockery”) do not signal any irony.

- (15) أحياناً: قد تكره إنسان، بل لا تطيق رؤيته، ولربما لا
تتعامل معه نهائياً، لكن كل هذا لا يميز لك السخرية
منه والاستهزاء به. ##سخرية ##الاستهزاء
Sometimes: You may hate someone; you may even
hate seeing them; and you may never deal with
them; but all of this does not allow you to mock or
make fun of them. #irony #mockery

Besides, our research suggests that linguistic features of a message and the position of a hashtag within the tweet can clarify the stance of the author. For example, the annotators doubt that tweet (16) below can be labeled as ironic:

- (16) كثير من ##الناس .. لو حدثتهم عن ألامك سخروا منك ..
تعليقاً و ##سخرية ..! لذلك .. احتفظ بها في داخلك ..
ولا تُبدها لأحد مهما كان ..
A lot of #people ... If you tell them about your pains,
they will make fun of you – by commenting with
#irony [ironically] ..! So ... Keep it to yourself, and
don't share it with anyone whomsoever ...

In this case, the tweeter uses ##سخرية (“#irony”), but the content does not seem ironic. The text is written in the MSA, so there is no reason to suspect a misunderstanding here. Arguably, the writer's intention cannot be ascertained without a prior conversational context. We notice, however, that an irony-related hashtag is used here in the first part of the message and as part of a sentence. This rule seems to apply to a substantial number of tweets: hashtags in these cases are employed as part of an opening clause which is not ironic or sarcastic. In comparison, if irony-related hashtags are embedded in the punchline, concluding clauses, they tend to correspond to the sentiments they describe, as in tweet (17). These examples point to the possibility of refining our model by way of assessing the position of hashtags within messages and analyzing the length and grammatical structure of tweets.

The initial analysis of the manual annotation data in this paper confirms our hypothesis about the relative credibil-

⁵<http://www.arabsocialmediareport.com/Twitter/LineChart.aspx>

(17) أكيد رح يكون هذا جوابك.لانه لما يصارحكم واحد
 بالحقيقة وبحقيقتكم دوروها #تهكم و #سخرية.
 لكن وش يبيك من #صغار_العقول

Of course, this would be your answer; because when someone tells you the truth, your truth, you turn that into #sarcasm and #irony. Anyway, what would you expect from #trivial_ people !

ity of hashtags for labeling irony. While hashtags can be used as a convenient ground rule for collecting data in general, manual annotation is indispensable for assembling a nuanced irony dataset.

5. Conclusion and Future Work

This paper outlined our research methodology and discussed the results of building a new irony-detection corpus in Arabic (MSA and dialects). The collection of Twitter messages was based on specific hashtags for irony and sarcasm. These tags were not removed for the subsequent manual annotation by professional linguists and served as an interpretative clue. The initial analysis of the findings confirms the relative reliability of hashtags for labeling and illuminates the importance of the position of hashtags and their grammatical role within sentences for the development of the model. Most importantly, critical examination of previous research and our data demonstrate the importance of professional manual annotation for providing refined results, which will reflect the linguistic richness and cultural diversity of the region.

Further annotation and analysis of the corpus, with special attention to the frequency of challenging cases and the tweets which were tagged as “Ambiguous”, will allow us to build an Arabic irony-detection model attuned to the wide range of irony expression employed in Arabic tweets.

Acknowledgements

We are grateful to our translators (Mr. Nabeel Rashid and Mr. Sayed Mohamed) and the AWARE Centre (Mr. Owen Connor and Dr. Julia Hudson) for their support. This publication was made possible by NPRP grant 9-175-1-033 from the Qatar National Research Fund (a member of Qatar Foundation). The findings achieved herein are solely the responsibility of the authors.

6. Bibliographical References

Atanasova, P., Barron-Cedeno, A., Elsayed, T., Suwaileh, R., Zaghouni, W., Kyuchukov, S., Martino, G. D. S., and Nakov, P. (2018). Overview of the clef-2018 checkthat! lab on automatic identification and verification of political claims. task 1: Check-worthiness. *arXiv preprint arXiv:1808.05542*.

Attardo, S. (2000). Irony as relevant inappropriateness. *Journal of Pragmatics*, 32(6):793–826.

Barbieri, F., Ronzano, F., and Saggion, H. (2014). Italian Irony Detection in Twitter: a First Approach. In *Proceedings of the First Italian Conference on Computational Linguistics (CLiC-it'14)*, volume 28, Pisa, Italy, Dec.

Barrón-Cedeño, A., Elsayed, T., Suwaileh, R., Márquez, L., Atanasova, P., Zaghouni, W., Kyuchukov, S., Da San Martino, G., and Nakov, P. (2018). Overview of the clef-2018 checkthat! lab on automatic identification and verification of political claims. task 2: Factuality. In *CLEF (Working Notes)*.

Bouamor, H., Zaghouni, W., Diab, M., Obeid, O., Oflazer, K., Ghoneim, M., and Hawwari, A. (2015). A pilot study on arabic multi-genre corpus diacritization. In *Proceedings of the Second Workshop on Arabic Natural Language Processing*, pages 80–88.

Bouamor, H., Habash, N., Salameh, M., Zaghouni, W., Rambow, O., Abdulrahim, D., Obeid, O., Khalifa, S., Eryani, F., Erdmann, A., et al. (2018). The madar arabic dialect corpus and lexicon. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Cignarella, A. T., Bosco, C., and Patti, V. (2017). TWIT-TIRÒ: a Social Media Corpus with a Multi-layered Annotation for Irony. In *Proceedings of the Four Italian Conference on Computational Linguistics (CLiC-it'17)*, volume 2006, pages 1–6, Rome, Italy, Dec. CEUR.

Cignarella, A. T., Frenda, S., Basile, V., Bosco, C., Patti, V., Rosso, P., et al. (2018). Overview of the EVALITA 2018 Task on Irony Detection in Italian Tweets (IronITA). In *Proceedings of the Six Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA'18)*, volume 2263, pages 1–6, Turin, Italy, Dec. CEUR-WS.

Davidov, D., Tsur, O., and Rappoport, A. (2010). Semi-Supervised Recognition of Sarcastic Sentences in Twitter and Amazon. In *Proceedings of the Fourteen Conference on Computational Natural Language Learning (CoNLL'10)*, pages 107–116, Uppsala, Sweden, Jul. Association for Computational Linguistics.

Filatova, E. (2012). Irony and Sarcasm: Corpus Generation and Analysis Using Crowdsourcing. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pages 392–398, Istanbul, Turkey, May. European Language Resources Association (ELRA).

Fleiss, J. L. (1971). Measuring Nominal Scale Agreement Among Many Raters. *Psychological bulletin*, 76(5):378.

Giora, R., Livnat, E., Fein, O., Barnea, A., Zeiman, R., and Berger, I. (2013). Negation Generates Nonliteral Interpretations by Default. *Metaphor and Symbol*, 28(2):89–115.

González-Ibáñez, R., Muresan, S., and Wacholder, N. (2011). Identifying Sarcasm in Twitter: A Closer Look. In *Proceedings of the Forty-nine Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (HLT'11)*, pages 581–586, Portland, Oregon, USA, Jun. Association for Computational Linguistics.

Grice, H. P., Cole, P., Morgan, J., et al. (1975). Logic and conversation. 1975, pages 41–58.

Habash, N., Soudi, A., and Buckwalter, T., (2007). *On Arabic Transliteration*, pages 15–22. Springer, Dordrecht.

Hernández Farías, D. I. (2017). *Irony and Sarcasm De-*

- tection in Twitter: The Role of Affective Content*. Ph.D. thesis.
- Joshi, A., Bhattacharyya, P., and Carman, M. J. (2017). Automatic Sarcasm Detection: A Survey. *ACM Computing Surveys (CSUR)*, 50(5):73:1–73:22.
- Karoui, J., Benamara, F., Moriceau, V., Patti, V., Bosco, C., and Aussenac-Gilles, N. (2017a). Exploring the Impact of Pragmatic Phenomena on Irony Detection in Tweets: A Multilingual Corpus Study. In *Proceedings of the Fifteen Conference of the European Chapter of the Association for Computational Linguistics (EACL'17)*, pages 262–272, Valencia, Spain, Apr. Association for Computational Linguistics.
- Karoui, J., Zitoune, F. B., and Moriceau, V. (2017b). SOUKHRIA: Towards an Irony Detection System for Arabic in Social Media. *Procedia Computer Science*, 117:161–168.
- Kumon-Nakamura, S., Glucksberg, S., and Brown, M. (1995). How About Another Piece of Pie: The Allusional Pretense Theory of Discourse Irony. *Journal of Experimental Psychology: General*, 124(1):3.
- Kunneman, F., Liebrecht, C., Van Mulken, M., and Van den Bosch, A. (2015). Signaling sarcasm: From hyperbole to hashtag. *Information Processing & Management*, 51(4):500–509.
- Landis, J. R. and Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, pages 159–174.
- Lin, S.-K. and Hsieh, S.-K. (2016). Sarcasm Detection in Chinese Using a Crowdsourced Corpus. In *Proceedings of the Twenty-Eight Conference on Computational Linguistics and Speech Processing (ROCLING'16)*, pages 299–310, Tainan, Taiwan, Oct. The Association for Computational Linguistics and Chinese Language Processing (ACLCLP).
- Lucariello, J. (1994). Situational irony: A concept of events gone awry. *Journal of Experimental Psychology: General*, 123(2):129–145, Jun.
- Maamouri, M., Bies, A., Kulick, S., Zaghouni, W., Graff, D., and Ciul, M. (2010). From speech to trees: Applying treebank annotation to arabic broadcast news. In *LREC*.
- Rajadesingan, A., Zafarani, R., and Liu, H. (2015). Sarcasm Detection on Twitter: A Behavioral Modeling Approach. In *Proceedings of the Eight ACM International Conference on Web Search and Data Mining (WSDM'15)*, pages 97–106, Shanghai, China, Feb. ACM.
- Rangel, F., Rosso, P., Charfi, A., and Zaghouni, W. (2019a). Detecting deceptive tweets in arabic for cybersecurity. In *2019 IEEE International Conference on Intelligence and Security Informatics (ISI)*, pages 86–91. IEEE.
- Rangel, F., Rosso, P., Charfi, A., Zaghouni, W., Ghanem, B., and Sanchez-Junquera, J. (2019b). Overview of the Track on Author Profiling and Deception Detection in Arabic. In *Working Notes of the Forum for Information Retrieval Evaluation (FIRE'19)*. *CEUR Workshop Proceedings*. In: *CEUR-WS.org, Kolkata, India*.
- Reyes, A. and Rosso, P. (2011). Mining Subjective Knowledge from Customer Reviews: A Specific Case of Irony Detection. In *Proceedings of the Second Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA'11)*, pages 118–124, Portland, Oregon, Jun. Association for Computational Linguistics.
- Rosso, P., Rangel, F., Farías, I. H., Cagnina, L., Zaghouni, W., and Charfi, A. (2018). A survey on author profiling, deception, and irony detection for the Arabic language. *Language and Linguistics Compass*, 12(4):e12275, Apr.
- Tang, Y.-j. and Chen, H.-H. (2014). Chinese Irony Corpus Construction and Ironic Structure Analysis. In *Proceedings of Twenty-five International Conference on Computational Linguistics: Technical Papers (COLING'14)*, pages 1269–1278, Dublin, Ireland, Aug. Dublin City University and Association for Computational Linguistics.
- Van Hee, C., Lefever, E., and Hoste, V. (2016). Guidelines for Annotating Irony in Social Media Text. Technical report.
- Van Hee, C. (2017). *Can machines sense irony? Exploring automatic irony detection on social media*. Ph.D. thesis, Ghent University.
- Walker, M. A., Tree, J. E. F., Anand, P., Abbott, R., and King, J. (2012). A Corpus for Research on Deliberation and Debate. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pages 812–817, Istanbul, Turkey, May. European Language Resources Association (ELRA).
- Wilson, D. and Sperber, D. (1992). On Verbal Irony. *Lingua*, 87(1):53–76.
- Zaghouni, W. and Charfi, A. (2018). Arap-tweet: A large multi-dialect twitter corpus for gender, age and language variety identification. *arXiv preprint arXiv:1808.07674*.
- Zaghouni, W., Mohit, B., Habash, N., Obeid, O., Tomeh, N., Rozovskaya, A., Farra, N., Alkuhlani, S., and Oflazer, K. (2014). Large scale arabic error annotation: Guidelines and framework.
- Zaghouni, W. (2014). Critical Survey of the Freely Available Arabic Corpora. In *Proceedings of the Nine International Conference on Language Resources and Evaluation (LREC'14)*, *OSACT Workshop*, Reykjavik, Iceland, May. European Language Resources Association (ELRA).