

A Fine-Grained Annotated Multi-Dialectal Arabic Corpus

Anis Charfi

Information Systems
Carnegie Mellon University in Qatar
acharfi@andrew.cmu.edu

Wajdi Zaghouni

College of Humanities and Social Sciences
Hamad Bin Khalifa University Qatar
wzaghouni@hbku.edu.qa

Syed Hassan Mehdi

Information Systems
Carnegie Mellon University in Qatar
smehdi@andrew.cmu.edu

Esraa Mohamed

Information Systems
Carnegie Mellon University in Qatar
emohamad@andrew.cmu.edu

Abstract

We present ARAP-Tweet 2.0, a corpus of 5 million dialectal Arabic tweets and 50 million words of about 3000 Twitter users from 17 Arab countries. Compared to the first version, the new corpus has significant improvements in terms of the data volume and the annotation quality. It is fully balanced with respect to dialect, gender, and three age groups: under 25 years, between 25 and 34, and 35 years and above. This paper describes the process of creating the corpus starting from gathering the dialectal phrases to find the users, to annotating their accounts and retrieving their tweets. We also report on the evaluation of the annotation quality using the inter-annotator agreement measures which were applied to the whole corpus and not just a subset. The obtained results were substantial with average Cohens Kappa values of 0.99, 0.92, and 0.88 for the annotation of gender, dialect, and age respectively. We also discuss some challenges encountered when developing this corpus.

1 Introduction

As the popularity of natural language processing (NLP) based tools increases, there is a rising need for language resources such as annotated corpora to develop NLP tools. In this regard, Arabic lags behind other languages due to several reasons. First, Arabic is complex at various levels of linguistic representation (phonology, orthography, morphology, and syntax). Even though native Arabic speakers prefer using Modern Standard Arabic (MSA) for official communications, we are witnessing with the popularity of social

media a rise in the use of dialectal Arabic for informal online interactions such as those found in blogs, forums, chats, tweets, posts, etc. (Mubarak and Darwish, 2014; Bouamor et al., 2018).

The content written by the users in social media sites can reveal some characteristics and attributes about them, which is the main focus of author profiling research. However, the lack of Arabic language resources limits that research on author profiling for the Arabic language in particular dialectal Arabic.

We present in this paper ARAP-Tweet 2.0, which is a large-scale manually-annotated multi-dialectal Arabic corpus. A first version of this corpus was presented in (Zaghouni and Charfi, 2018a,b) and it was extended significantly in terms of data volume, number of users, and annotation quality. ARAP-Tweet 2.0 covers dialects from 17 Arab countries and 15 regions. The number of users per region is 198, including a total of about 3000 users and approximately 5 million tweets. All users' accounts were manually annotated with respect to the dialect, gender, and age. Thereby, we distinguished three age groups: under 25 years, between 25 years and 34 years, and 35 years and above. Moreover, significant effort was put in checking and improving the annotation quality.

The rest of this paper is organized as follows. Section 2 presents related work and Section 3 describes the methodology used in creating and annotating our corpus. Section 4 reports on the verification of the annotation as well as the evaluation. Section 5 discusses some challenges that we encountered when developing this corpus. Finally, Section 6 concludes the paper and outlines possible directions for future work.

2 Related Work

Most research on Arabic NLP has focused on Modern Standard Arabic (MSA) (Habash, 2010). There are many parallel and monolingual annotated data collections with syntactic and semantic information such as the different iterations of Penn Arabic Probanks (Palmer et al., 2008; Zaghouani et al., 2010, 2012) and treebanks (Maamouri et al., 2010). Based on such resources, various tools were developed for syntactic parsing and morphological analysis (Habash, 2010).

Even though there are relatively many resources for MSA, Dialectal Arabic (DA) lags behind in terms of available resources. There have been some limited efforts toward creating resources for the most popular dialects such as the Egyptian and Levantine dialects which were presented in (Habash et al., 2013; Diab and Habash, 2007; Pasha et al., 2014). For example, Habash et al. created resources for morphological analysis of Egyptian dialect (Habash et al., 2013). For their work on machine translation, Zbib et al. (2012) created Levantine-English and Egyptian-English parallel corpora using crowd sourcing. Khalifa et al. (2018) created a morphologically annotated data corpus of Emirati Dialect.

Khalifa et al. (2016) created a corpus of 100M words covering various Arabic dialects. Other related projects worth to be mentioned are: the Egyptian Arabic Treebank (Maamouri et al., 2014); the Levantine Arabic Treebank (Maamouri et al., 2006), The Curras Palestinian Arabic annotated corpus with more than 70,000 words of various genres (Jarrar et al., 2014).

Furthermore, AlShargi et al. (2016) created a Yemeni (Sanaa Dialect) dataset and also a Moroccan Arabic corpus, while Al-Twairish et al. (2018) built SUAR, a Najdi Arabic corpus annotated with the morphological analyzer MADAMIRA (Pasha et al., 2014). Finally, Voss et al. (2014) presented a Moroccan Arabic corpus annotated for code-switching (French, Berber and Moroccan Arabic).

Moreover, there have been some efforts towards creating Dialectal Arabic corpora by either translating existing corpora to dialects or by crowd sourcing annotation for data collected through various sources such as microblogs (e.g., Twitter). Along these lines, DART was developed as a Twitter based data set of dialectal Arabic covering five Arabic dialects: Egyptian, Levantine, Gulf,

and Iraqi (Alsarsour et al., 2018). The annotation of this corpus was done through crowd sourcing. Bouamor et al. (2014) presented a multi-dialectal parallel corpus with 2,000 sentences translated to MSA, Tunisian, Jordanian, Palestinian and Syrian Arabic. Later on, MADAR was developed as a Multi-dialectal large scale corpus that provides parallel translation for 25 Arabic city dialects (Bouamor et al., 2018). All these efforts on creating Dialectal Arabic corpora either targeted some specific dialects only or did not provide the necessary annotation for author profiling such as annotation about age and gender.

3 Methodology

In the following, we report on the methodology and process followed for the creation, annotation, and validation of our corpus.

3.1 Corpus Overview

Twitter is an increasingly popular social media platform among the Arabic speaking people who tend to frequently use their Arabic dialects when sharing their stories and opinions. Therefore, we focused on Twitter to collect the data for our corpus. ARAP-Tweet 2.0 was developed in the context of the ARAP research project¹ and it includes about 5.3 million Arabic tweets of approximately 3000 users from the following 15 Arabic speaking regions: Qatar, Kuwait, United Arab Emirate (UAE), Kingdom of Saudi Arabia (KSA), Oman, Yemen, Iraq, Tunisia, Algeria, Libya, Morocco, Lebanon-Syria, Palestine-Jordan, Egypt, and Sudan. To the best of our knowledge, there is no other corpus that covers as many Arab regions as ours. For every region, we collected the tweets of 198 users that were equally-balanced over gender and three age groups: under 25, 25 until 34, and 35 and up. Compared to the first version of our corpus (Zaghouani and Charfi, 2018a) the number of users per region doubled in this second version.

3.2 Users

We used Twitter as our source for finding the users. In the beginning, we got a number of accounts by the geographical location of tweets and by searching on Twitter using specific words. We collected the users who posted tweets in a specific geographical location as defined in the tweet's object. As this information was not available for all

¹arap.qatar.cmu.edu

tweets, we searched for users who had tweets that include some unique words, which are specific to a certain Arab region. Table 1 shows some of these unique words from every region covered by our corpus along with a tweet example for each word.

After that, we found further accounts using the followers of the initially identified users. Once the users were found, we hired experienced annotators to manually annotate their age, gender, and dialect. The annotators followed well-defined annotation guidelines, which are based on an extended version of the guidelines published in (Zaghouani and Charfi, 2018b). We continuously monitored the annotators' work and updated the guidelines when needed. For each region, we retrieved the tweets of the users from their Twitter profiles. The selected users had to have at least 100 Arabic tweets and at most 3200 tweets. Moreover, the tweets had to be the original tweets, which means that we did not include retweets in our corpus. The average number of words per tweet is 10.

3.3 Annotation

In the following, we report on the annotation of the users with respect to gender, age and dialect.

3.3.1 Gender Annotation

The gender was manually annotated for users and the number of users is balanced with respect to the gender across all regions, which means that we have 99 male users and 99 female users for each region. The annotation was done based on guidelines and criteria that were used to distinguish male and female users. For every user, the annotator determined the gender by analyzing the username, profile picture, and indicative words. The annotator checked the username if it is denoting a male or female name. Some users put their real photos in their Twitter profile, which also helped in determining their gender. In some cases, the username and the profile picture did not give sufficient information to identify their gender. Therefore, the annotators were looking for some indicative words in the user's tweets. For example, the adjectives that describe feminine in Arabic usually end with *ة* (*Taa' Marbootah*). Therefore, finding words that end with the Arabic letter *Taa' Marbootah* such as *بردانة* (*Feeling cold*), *نعسانة* (*Feeling sleepy*), or *جعانة* (*Feeling hungry*) in the tweets of a user indicates that the writer is a female. However, if the writer is a male, the adjectives

used in his tweets would not have that letter. The annotators used this Arabic rule to determine the user's gender when it was not possible to do so using the username or the profile photo. It is noteworthy that users for which the gender could not be determined were not included in the corpus and were replaced by other users for which the gender could be determined as explained above.

3.3.2 Age Annotation

We annotated the age of the selected Twitter users using three age groups: under 25, 25 until 34, and 35 and above. In many cases, we were able to find the exact ages whereas in some other cases we were able to determine the correct age group without the exact age. The annotators went through the following steps to determine the age of a user:

Getting the exact age: Several Twitter users include their birth year as a part of their usernames. For example, in the username *Omda1981m* the birth year is most likely 1981, which indicates that this user is 38 years old. Some other users put their date of birth in their Twitter biography or in their other social media accounts such as Facebook, Instagram, or LinkedIn. Other users put their exact age in some of their tweets and we were able to search for those tweets using some relevant keywords in different languages as shown in Table 2.

Getting the estimated age: There were several cases in which we were not able to determine the exact age even after following the above-mentioned steps. In these cases, we opted to determine the approximate ages, so that we could annotate the user with the correct age group. This was done by searching for the user on other social media networks such as LinkedIn and Facebook. We often found either their exact ages or other age related hints such as the year of their graduation from university, which helped us in making an educated guess towards the user's age. Furthermore, as a last resort for some few users, two annotators guessed the age separately using the Twitter photo. Then, we used Microsoft's online tool² that predicts the age by analyzing the facial features of a user through machine learning. Only users were included for which the age group guessed by Microsoft's age determination tool.

3.3.3 Dialect Annotation

As mentioned above, we selected the users for our corpus by searching for tweets with dialect spe-

²www.how-old.net

Region	Unique Word	User's Tweet
Lebanon-Syria	هيدا	عملت غوغل و هيدا يلي طلع معي
Tunisia	برشا	ذقت أنواع كسكسي متاع برشا دول يبقى الألد والافضل التونسي
Iraq	اهوايه	اي والله صحيح اهوايه اكو هيحي نماذج مع الاسف
KSA	حفله	صدقوني الموضوع فيه إن وفيه حفله
Palestine-Jordan	هاظ	طبعاً هاظ اشي مفروغ منه
Qatar	احاتي	انا لعبتي احاتي الدراسة بس ما ادرس
Kuwait	طنطل	عاد انتي دلو مفروض تكونين طنطل شفيكم يا جماعه
Sudan	زول	تاني ما هعمل ريتويت لي اي زول
UAE	مايوب	تسلمين يالوافيه يا راعيه الذوق يالراقيه دوج راعيه مايوب
Yemen	معادبش	يعني معقوله معادبش ولا شي كملوها السرق
Morocco	دابا	كنت دهما كنتسائل أش كاين فالكتاب دابا لقيت الجواب الشافي
Algeria	زاوش	كنا انا وخويا وولد عمي نصيدو زاوش ونسلخوه ونديروه مشوي
Libya	شهدره	عليش شهدره اسمي في رييلي
Egypt	دلوقتي	سيبوني دلوقتي انا كويس
Oman	شقبه	انزين عادي م خسرانه بطلعي شقبه صدقيني

Table 1: Examples of unique words from each dialectal region with tweets that include these words.

Indicative word	Language	Tweet Example
(My birthday) ميلادي	Arabic	اكتوبر ١٩٩٠ عيد ميلادي
(My age) عمري	Arabic	الليله اخر يوم لي وأنا عمري ٢٢
(Year) سنة	Arabic	بكون ٢٣ سنة ان شاء الله في مونديال ٢٠٢٢
Turn	English	Hey Sosoo I will turn 20 in few days
(Years) Ans	French	16 ans?? J'ai 35 ans ma chérie

Table 2: Examples of keywords from different languages used to determine the users age.

cific words and then retrieving the users of those tweets and their followers. In addition, once we retrieved the user's tweets, the annotator(s) manually went through them and checked that at least half of the tweets are in the user's respective dialect. When the users' tweets were not mostly in their originally identified dialect, the annotators were instructed to replace them with other users from the same dialect, the same gender, and the same age group.

3.4 Tweets Retrieval

Twitter offers an Application Programming Interface (API) that allows developers to interact with Twitter's web services. We used the Python pro-

gramming language with the *tweepy.py*³ library to interact with the Twitter API and retrieve each user's timeline (tweets). Due to restrictions by Twitter, we were only allowed to retrieve at most 3,200 of the users most recent Tweets. The number of tweets of every user in our corpus ranges from a required minimum of 100 tweets to a maximum of 3,200 tweets. After collecting the maximum possible number of tweets for each user, we filtered them. Specifically, we removed the non-Arabic tweets, retweets, and short tweets that contain less than 3 words. Eventually, we kept only the Twitter accounts who had at least 100 original users tweets that are in Arabic and that have

³<http://docs.tweepy.org/en/v3.5.0/index.html>

three words at least. At the end of this process, we were able to have a well-balanced multi-dialect corpus of 198 users for 15 regions spanning 17 Arab countries. The corpus is equally-balanced with respect to gender and age.

4 Verification and Evaluation

Our main objective was to produce a multi-dialectal corpus of tweets with a very good quality of annotation so that it can be used to promote Arabic NLP research including research on author profiling. This is a major improvement compared to the previous version of the corpus. To ensure a good quality annotation, we performed two rounds of annotation for each Arab region, which means that each region was annotated by two different annotators. Next, we report on the verification and evaluation process for the annotation of dialect, age, and gender. Moreover, we present the evaluation results for each annotation category.

4.1 Verification of Dialect Annotation

The annotation was done by experienced Arab annotators who were asked to go through each file of retrieved users tweets and check the dialect used in these tweets. This step was necessary to verify the overall dialect of all tweets for every user, and not only the last tweets in a users timeline. Based on this dialect verification, we did the following:

Removing the users who post tweets in multiple dialects: many Arab users do not live in their original countries, and consequently, their dialects are affected by the dialects of the countries they live in. We opted to drop such users from our corpus and we replaced them by users from the same gender and age group and which used one dialect only.

Removing the users who have many tweets in MSA: some Arab users post few tweets in their own dialects, and they write their tweets mostly in MSA. After checking all retrieved tweets of a given user, we included in our corpus only those who have more than half of their tweets written in their dialect.

4.2 Verification of Age and Gender Annotation

Although the age was determined manually, the corpus included initially some users for which neither the exact age nor the age group could be determined with high confidence. These cases oc-

curred for example when the age of the user was determined mainly by using their Twitter profile photo.

Moreover, the annotators found initially some difficulties with determining the gender for some users for instance when the profile did not include a real name nor a profile photo. For the latter cases, the accounts were replaced by accounts from the same gender and age group for which the age could be determined with higher confidence. For all accounts, a second round of annotation was performed by a different annotator. At the end of this phase, we were able to have two annotations from different contributors for each user account. Eventually, variations were checked and resolved by a member of the research team together with the annotators by reviewing and discussing the provided justifications for a certain annotation of age, gender, or dialect.

4.3 Annotation Evaluation

In order to evaluate the quality of the annotation, we used the inter-annotator agreement measures. We were able to have two rounds of annotation by different annotators for each user and this applies for the gender, dialect and age group for the whole corpus unlike in the previous version (Zaghouani and Charfi, 2018a) in which the inter-annotator agreement was based on a 10 % subset of the corpus data.

We measured the inter-annotator agreement using Cohens Kappa, which is a statistical measure of the degree of agreement for a data point (gender, dialect, and age in our case), that was labeled by two annotators, over what would be expected by chance. We obtained substantial results for the agreement with average Kappa values of **0.99**, **0.92**, and **0.88** respectively for the annotation of gender, dialect, and age. The exact Kappa values per region are shown in Table 3.

5 Challenges

In the following, we report on some challenges that we faced when developing this corpus:

First, we encountered some difficulties in finding user accounts for some age groups for certain regions. This was the case for Sudani, Iraqi, and Gulf females whose age is 35 or above. For some Arab regions, female Twitter users tend to hide their real name and also avoid putting their photos on social media because of the local cul-

Region	Gender	Age	Dialect
Lebanon-Syria	1	0.924	1
Tunisia	1	0.742	1
Iraq	1	0.696	0.9
Saudi Arabia	1	0.856	0.9
Palestine-Jordan	0.989	0.856	0.9
Qatar	0.969	0.901	0.9
Kuwait	0.979	0.886	0.9
Sudan	1	0.734	0.8
UAE	0.979	0.848	0.8
Yemen	1	0.984	1
Morocco	1	0.954	0.9
Algeria	0.959	0.931	0.9
Libya	0.989	0.924	1
Egypt	1	0.969	1
Oman	0.989	0.931	0.9
Overall	0.99	0.88	0.92

Table 3: Kappa values per region for gender, age, and dialect annotation.

ture and norms. This was an issue for us as we sometimes depend on the Twitter profile photos to guess the age especially if the username is not a real name. Second, we notice that older Arabs often write in Modern Standard Arabic (MSA) on social media, and this made the task of finding dialect users above 35 years much harder. Third, we faced an issue with Maghrebi users who often tweet in French more than Arabic.

To address these issues, we put more effort into the user selection and we replaced any users who did not fulfill the criteria explained above.

Fourth, we noticed that some users use more than one Arabic dialect to write their tweets on Twitter. For example, in Arab regions with high immigration rates such as Gulf countries, the dialect of the residents is sometimes affected by the language of their host country. Consequently, these users tweet in multiple dialects and they might mix dialects even in one same tweet. Other users had similar issues because their parents come from different Arab countries. To address this issues, all tweets were manually reviewed and users tweeting in two dialects or more were replaced by mono-dialectal users.

Fifth, we worked on regions sequentially, i.e., we selected a large set of accounts from Twitter for one region, annotated them, and then retrieved their tweets. Then, we worked on the next region.

As a result, when we retrieved again the users or their tweets we noticed that some profiles were either made private or even closed. In such cases, we were no longer able to access the users profile and tweets. We addressed this problem by periodically checking the users' accounts and replacing any accounts that were deactivated or made private.

6 Conclusion

We presented in this paper a significantly improved version of a large-scale manually-annotated and fine-grained multi-dialectal corpus of Arabic tweets, which is composed of 5 millions dialectal Arabic tweets of about 3000 Twitter users from 17 Arab countries. To the best of our knowledge, our corpus is the most comprehensive dialectal corpus in terms of coverage for so many Arab dialects and the data volume. Moreover, our corpus is balanced with respect to dialect, gender, and age. The corpus was annotated manually based on well-defined annotation guidelines and it was fully evaluated using the inter-annotation agreement measures.

The corpus was already and it can further be used to promote research on Arabic NLP including author profiling (Rosso et al., 2018), authorship attribution, dialect identification, sentiment analysis in dialectal Arabic, and bots detection in dialectal Arabic.

Acknowledgments

We thank our dedicated annotators: Hoda Ibrahim, Safaa Ellaham, Syrine Guediche for their hard work on the corpus annotation. This publication was made possible by NPRP 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.

References

- Faisal Al-Shargi, Aidan Kaplan, Ramy Eskander, Nizar Habash, and Owen Rambow. 2016. Morphologically annotated corpora and morphological analyzers for moroccan and sanaani yemeni arabic. In *10th Language Resources and Evaluation Conference (LREC 2016)*.
- Nora Al-Twairish, Rawan Al-Matham, Nora Madi, Nada Almgren, Al-Hanouf Al-Aljmi, Shahad Alshalan, Raghad Alshalan, Nafla Alrumayyan, Shams Al-Manea, Sumayah Bawazeer, et al. 2018. Suar: Towards building a corpus for the saudi dialect. *Procedia computer science* 142:72–82.

- Israa Alsarsour, Esraa Mohamed, Reem Suwaileh, and Tamer Elsayed. 2018. Dart: A large dataset of dialectal arabic tweets. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*.
- Houda Bouamor, Nizar Habash, and Kemal Oflazer. 2014. A multidialectal parallel corpus of arabic. In *LREC*. pages 1240–1245.
- Houda Bouamor, Nizar Habash, Mohammad Salameh, Wajdi Zaghouni, Owen Rambow, Dana Abdulrahim, Ossama Obeid, Salam Khalifa, Fadhl Eryani, Alexander Erdmann, et al. 2018. The madar arabic dialect corpus and lexicon. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*.
- Mona Diab and Nizar Habash. 2007. Arabic dialect processing tutorial. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Tutorial Abstracts*. Association for Computational Linguistics, pages 5–6.
- Nizar Habash, Ryan Roth, Owen Rambow, Ramy Eskander, and Nadi Tomeh. 2013. Morphological analysis and disambiguation for dialectal arabic. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. pages 426–432.
- Nizar Y Habash. 2010. Introduction to arabic natural language processing. *Synthesis Lectures on Human Language Technologies* 3(1):1–187.
- Mustafa Jarrar, Nizar Habash, Diyam Fuad Akra, and Nasser Zalmout. 2014. Building a corpus for palestinian arabic: a preliminary study .
- Salam Khalifa, Nizar Habash, Dana Abdulrahim, and Sara Hassan. 2016. A large scale corpus of gulf arabic. *arXiv preprint arXiv:1609.02960* .
- Salam Khalifa, Nizar Habash, Fadhl Eryani, Ossama Obeid, Dana Abdulrahim, and Meera Al Kaabi. 2018. A morphologically annotated corpus of emirati arabic. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*.
- Mohamed Maamouri, Ann Bies, Tim Buckwalter, Mona T Diab, Nizar Habash, Owen Rambow, and Dalila Tabessi. 2006. Developing and using a pilot dialectal arabic treebank. In *LREC*. pages 443–448.
- Mohamed Maamouri, Ann Bies, Seth Kulick, Michael Ciul, Nizar Habash, and Ramy Eskander. 2014. Developing an egyptian arabic treebank: Impact of dialectal morphology on annotation and tool development. In *LREC*. pages 2348–2354.
- Mohamed Maamouri, Ann Bies, Seth Kulick, Wajdi Zaghouni, David Graff, and Michael Ciul. 2010. From speech to trees: Applying treebank annotation to arabic broadcast news. In *LREC*.
- Hamdy Mubarak and Kareem Darwish. 2014. Using twitter to collect a multi-dialectal corpus of arabic. In *Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP)*. pages 1–7.
- Martha Palmer, Olga Babko-Malaya, Ann Bies, Mona T Diab, Mohamed Maamouri, Aous Mansouri, and Wajdi Zaghouni. 2008. A pilot arabic propbank. In *LREC*.
- Arfath Pasha, Mohamed Al-Badrashiny, Mona T Diab, Ahmed El Kholy, Ramy Eskander, Nizar Habash, Manoj Pooleery, Owen Rambow, and Ryan Roth. 2014. Madamira: A fast, comprehensive tool for morphological analysis and disambiguation of arabic. In *LREC*. volume 14, pages 1094–1101.
- Paolo Rosso, Francisco M. Rangel Pardo, Bilal Ghanem, and Anis Charfi. 2018. ARAP: arabic author profiling project for cyber-security. *Procesamiento del Lenguaje Natural* 61:135–138.
- Clare R Voss, Stephen Tratz, Jamal Laoudi, and Douglas M Briesch. 2014. Finding romanized arabic dialect in code-mixed tweets. In *LREC*. pages 2249–2253.
- Wajdi Zaghouni and Anis Charfi. 2018a. Arap-tweet: A large multi-dialect twitter corpus for gender, age and language variety identification. *arXiv preprint arXiv:1808.07674* .
- Wajdi Zaghouni and Anis Charfi. 2018b. Guidelines and annotation framework for arabic author profiling. *arXiv preprint arXiv:1808.07678* .
- Wajdi Zaghouni, Mona Diab, Aous Mansouri, Sameer Pradhan, and Martha Palmer. 2010. The revised arabic propbank. In *Proceedings of the fourth linguistic annotation workshop*. Association for Computational Linguistics, pages 222–226.
- Wajdi Zaghouni, Abdelati Hawwari, and Mona Diab. 2012. A pilot propbank annotation for quranic arabic. In *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature*. pages 78–83.
- Rabih Zbib, Erika Malchiodi, Jacob Devlin, David Stallard, Spyros Matsoukas, Richard Schwartz, John Makhoul, Omar F Zaidan, and Chris Callison-Burch. 2012. Machine translation of arabic dialects. In *Proceedings of the NAACL 2012*. pages 49–59.