Annotating Targets of Opinions in Arabic using Crowdsourcing

Noura Farra
Columbia University
New York, NY 10027, USA
noura@cs.columbia.edu

Kathleen McKeown
Columbia University
New York, NY 10027, USA
kathy@cs.columbia.edu

Nizar Habash
New York University
Abu Dhabi, UAE
nizar.habash@nyu.edu

Abstract

We present a method for annotating targets of opinions in Arabic in a two-stage process using the crowdsourcing tool Amazon Mechanical Turk. The first stage consists of identifying candidate targets "entities" in a given text. The second stage consists of identifying the opinion polarity (positive, negative, or neutral) expressed about a specific entity. We annotate a corpus of Arabic text using this method, selecting our data from online commentaries in different domains. Despite the complexity of the task, we find high agreement. We present detailed analysis.

1 Introduction

An important task in subjectivity analysis of text is the identification of targets - also often called *topics* or *subjects* - of opinionated text. Knowledge of the target is important for making sense of an opinion (e.g in '*The* ***will of the people*** *will prevail over the* ***regime's brutality***', the opinion is positive towards 'the people' and negative towards 'the regime'). An opinion system which can identify both targets and polarities of opinions, and which can summarize the opinions of writers towards different targets, will be more informative than one which only identifies the overall sentiment of the text. This problem has started gaining interest in the product review domain (Hu and Liu, 2004; Qiu et al., 2011), news and social media (Kim and Hovy, 2006; Jiang et al., 2011), and in general language and discourse (Wilson, 2008; Ruppenhofer et al., 2008; Somasundaran and Wiebe, 2009).

Annotating targets of opinion is a difficult and expensive task, requiring definition of what constitutes a target, whether targets are linked to opinion expressions, and how the boundaries of target spans should be defined (e.g **'the people'** vs. **'the will of the people'** or **'the regime'** vs. **'the regime's brutality'**), a problem which annotators often disagree on (Pontiki et al., 2014; Kim and Hovy, 2006; Somasundaran et al., 2008). Additionally, it is not always straightforward to attribute a target to a specific opinion phrase. Consider for example the following statement:

> *'The Lebanese PM said he was convinced that there would be a consensus on* ***the presidential election****, because since the moment the US and Iran had reached an understanding in the region, things were starting to look positive.'*

Which is the opinion expression that leads us to believe that the PM is optimistic about the target **presidential election**? Is it *'convinced'*, *'consensus'*, *'reached an understanding'*, or *'look positive'*, or a combination of the above? Such decisions are difficult for annotators to agree on; many studies have noted these challenges (Stoyanov and Cardie, 2008; Ruppenhofer et al., 2008) which can make the task complex.

Compared to the amount of resources available for sentiment and subjectivity analysis, there is much less annotated data available for this more fine-grained type of analysis. Due to the difficulty of the task, most of the available datasets of fine-grained subjectivity have been annotated by trained annotators or expert linguists, making the process slower and more expensive.

In this work, we consider annotation of targets using a sequence of simple crowdsourced substeps. We focus on Arabic, where subjectivity analysis is of growing interest, and where there are no publicly available resources for fine-grained opinion analysis. We assume targets of opinions to be noun phrase entities: people, places, things or ideas. We develop a two-stage annotation process for annotating targets of opinions using Amazon Mechanical Turk. In the first, annotators list all important 'entities', and in the second, they choose the polarity expressed (positive, negative, or neutral) towards any given entity. We select online data from multiple domains: politics, sports, and culture; and we provide a new publicly available resource for Arabic by annotating it for targets of opinions along with their polarities. Finally, we evaluate the quality of the data at different stages, obtaining majority agreement on sentiment for 91.8% of entities in a corpus of 1177 news article comments. We also find that the morphology and grammar of Arabic lends itself to even more variations in identifying the boundaries of targets.

Section 2 describes related annotation work. Section 3 describes the Amazon Mechanical Turk tasks design, the data selection, and the annotation process. In Section 4, we examine and analyze the annotations, evaluate the inter-annotator agreement, and provide detailed examples. We conclude in section 5.

2 Related Work

2.1 Annotating Targets in English

Fine-grained subjectivity annotation in the English language has recently started gaining interest, where annotation can include opinion targets, opinion sources, or phrase-level opinion expressions. One of the early datasets collected for identifying opinion targets is that of (Hu and Liu, 2004), where product features (e.g price, quality) were annotated in customer reviews of consumer electronics. These consisted of mostly explicit product features annotated by one person.

Also in the product review domain, the SemEval Task on aspect feature mining in 2014 (Pontiki et al., 2014) was concerned with finding aspect features of products with the polarities towards them. The products (e.g 'restaurant') and coarse-grained features (e.g 'service') were provided to annotators, who identified the aspects (e.g 'waiter') and the corresponding sentiment.

The MPQA corpus is an in-depth and general-purpose resource for fine-grained subjectivity annotations (Wiebe et al., 2005; Wilson, 2008), containing annotations of opinion expressions at the phrase level while specifying polarities, sources, and target spans. The annotation scheme links each subjective expression to one or more attitudes, which in turn can have one or more or no targets. The target annotations include the full target spans, but do not necessarily identify target entities within the span. Stoyanov and Cardie (2008) extended part of the MPQA corpus by annotating it for 'topics', arguing that 'targets' refer to the syntactic span of text that identifies the content of an opinion, while 'topic' is the real-world object or entity corresponding to the primary subject of the opinion. Using trained annotators, they identify 'topic clusters', which group together all opinions referring to the same topic. In parallel with this work, part of the MPQA corpus was recently annotated for entity-level targets (Deng and Wiebe, 2015) by specifying target entities within the MPQA span, leading to the annotation of 292 targets by two annotators. The entities were anchored to the head word of the noun phrase or verb phrase that refers to the entity or event. In our work, we only consider noun phrase entities, and we consider the noun phrase itself as an entity.

Other fine-grained annotation studies include that of Toprak et al. (2010) who enrich target and holder annotations in consumer reviews with measures such as relevancy and intensity, and Somasundaran et al. (2008) who perform discourse-level annotation of opinion frames, which consist of opinions whose targets are described by similar or contrasting relations.

In these studies, the annotation was usually done by trained individuals or someone who has knowledge and experience in the task. Our study is different in that it utilizes crowdsourcing for the annotation process, and it focuses on the marking of important entities and concepts as targets of opinions in the more noisy online commentary genre. We view targets as 'real-world entities', similar to the topics discussed by Stoyanov and Cardie (2008), and the targets in (Deng and Wiebe, 2015), and we annotate multiple targets in the text.

Carvalho et al. (2011) also annotated targets in online commentary data; here targets were considered to be human entities, namely political and media personalities. This annotation was done by one trained annotator where agreement was computed for a portion of the data. Another related task was that of Lawson et al. (2010) who describe a Mechanical Turk annotation study for annotating named entities in emails, with favorable agreement results. The tasks for identifying the spans of and labeling the named entities were grouped in a single Human Intelligence Task (HIT).

2.2 Annotation Studies in Arabic

Abdul-Mageed and Diab (2011) performed a sentence-level annotation study for Modern Standard Arabic (MSA) newswire data which covered multiple domains including politics, sports, economy, culture, and others. Both the domains and the sentence-level sentiment were annotated by two trained annotators. Our data also comes from different domains, but it is from the genre of online commentaries, which have greater prevalence of dialect, imperfect grammar, and spelling errors. Also, to select less prevalent domains from our comments corpus, we used topic modeling.

There have been other MTurk studies in Arabic; among them Zaidan and Callison-Burch (2011) who annotated dialectness, Denkowski et al. (2010) who annotated machine translation pairs, and Higgins et al. (2010) who annotated Arabic nicknames. To the best of our knowledge, there are no known studies for target or topic annotation for Arabic.

3 Annotation Process

We describe the crowdsourcing process for annotating targets of opinions, including the choices which motivated our design, the tasks we designed on Amazon Mechanical Turk, and the way we selected our data.

3.1 Scope and Decisions

We assume targets of opinions to be nouns and noun phrases representing entities and concepts, which could be people, places, things, or important ideas. Consider for example:

'It is great ***that so many people showed up to the protest.****'*

The full target span is marked in bold, but the actual entity which receives the positive opinion is **'the protest'**. We are interested in such entities; for example, entities could be politicians, organizations, events, sports teams, companies, products, or important concepts and ideas such as 'democracy' or entities representing ideological belief.

Given the complexity of the task, we annotate targets without specifying opinion expressions that are linked to them, as in (Pontiki et al., 2014; Hu and Liu, 2004), although the dataset can be extended for this purpose to provide richer information for modeling. We assume the availability of an Arabic opinion lexicon, to identify the opinion words. We don't consider targets of *subjective-neutral* judgments (e.g *"I expect it will rain tomorrow"*). For this corpus, we are interested only in targets of polar **positive or negative** opinions; everything else we regard as neutral. Moreover, since our data comes from online commentaries, we assume that in the majority of cases, the opinion holder is the writer of the post.

3.2 Amazon Mechanical Turk Tasks

Instead of asking annotators to directly identify targets of opinions, which we believed to be a much harder task, we broke the annotation into two stages, each in a different series of HITs (Human Intelligence Tasks). The task guidelines were presented in Modern Standard Arabic (MSA) to guarantee that only Arabic speakers would be able to understand and work on them. Many of the insights in the task design were gained from an extensive pilot study.

Task 1: Identifying Candidate Entities Given an article comment, annotators are asked to list the main nouns and noun phrases that correspond to people, places, things, and ideas. This task, or HIT, is given to three annotators and a few examples of appropriate answers are provided.

The answers from the three annotators are then combined by taking the intersection of common noun phrases listed by all three responses. If they only agree on a subset of the noun phrase, we choose the maximal phrase among agreed entities in order to determine the entity span. For example, if two annotators specify *the president* and a third specifies *the election of the president*, we keep *the election of the president*. The maximal noun phrase was also chosen by Pontiki et al. (2014) when resolving disagreements on target spans.

We allowed annotators to list references in the comment to the same entity (e.g *'The president'* and *'President Mubarak'*) as separate entities.

Insights from Pilot We asked specifically for the **main** noun phrases, after we found that annotators in the pilot over-generated nouns and noun phrases, listing clearly unimportant entities (such as اليوم 'today/this day', and السلام 'hello/the greeting'), which would make Task 2 unnecessarily expensive. They would also break up noun phrases which clearly referred to a single entity (such as separating كرسي *'the seat'* and الرئاسة *'the presidency'* from كرسي الرئاسة *'the presidency's seat'*), so we instructed them to keep such cases as a single entity. These reasons also support choosing the maximal agreeing noun phrase provided by annotators. By making these changes, the average number of entities resolved per comment was reduced from 8 entities in the pilot study to 6 entities in the full study.

We paid 30 cents for Task 1, due to its importance and due to the time it took workers to complete (2-3 minutes on average).

Task 2: Identifying Sentiment towards Entities In the second task (HIT), annotators are presented with an article comment and a single entity, and are asked to specify the opinion of the comment towards this entity, termed a 'topic' موضوع. The entities are chosen from the resolved responses in Task 1. The question is in multiple-choice form where they can choose from options: positive, negative, or neutral. Each HIT is given to five annotators, and the entities which are specified as **positive or negative** with majority agreement of 3 are considered to be **targets**. Entities with disagreement, or with neutral majority, are discarded as non-targets. In this question, we tell annotators that opinions can include sentiment, belief, feelings, or judgments, and that the **neutral** option should be selected if the comment reveals either no opinion or an unbiased opinion towards this particular entity. We provide multiple examples. For this task, we paid workers 5 cents per HIT, which took 30 seconds to 1 minute on average.

Insights from Pilot In our pilot study, we had an additional question in this HIT which asks annotators to specify the holder of the opinion, which could be the writer or someone else mentioned in the text. However, we removed this question in the final study due to the low quality of responses in the pilot, some of which reflected misunderstanding of the question or were left blank.

Additionally, we found that some annotators specified the overall sentiment of the comment rather than the sentiment about the topic. We thus emphasized, and included an additional English translation of the instruction that the opinion polarity should be about the specific **topic** and not of the whole comment.

We completed the full annotation study in five rounds of a few hundred comments each. For the first two rounds of annotation, we rejected all HITs that were clearly spamming the task or were not Arabic speakers. After that we created task qualifications and allowed only a qualified group of workers (5 for Task 1 and 10 for Task 2) to access the tasks, based on their performance in the previous tasks.

Domain	**# Comments**	**Distribution(%)**
Politics	596	51
Culture	382	32
Sports	199	17
Total	1177	100

Table 1: Distribution of article comments by domain

3.3 Data Selection

Our data is selected from the Qatar Arabic Language Bank (QALB) (Mohit et al., 2014; Zaghouani et al., 2014), which includes online commentaries to *Aljazeera* newspaper articles.

Topic Modeling We initially selected a random sample of data from the *Aljazeera* corpus, which contains mostly political data. In our pilot study and first annotation round, we found that this data was biased towards negative sentiment. We thus used topic modeling (Blei et al., 2003; McCallum, 2002) to select data from other domains which we thought might contain more positive sentiment. Upon applying a topic model specifying 40 topics to the *Aljazeera* corpus, we found a general "sports" topic and a general "culture" (language, science, technology, society) topic among the other political topics. We chose sports and culture comments by taking the top few hundred comments having the highest probability score for these topics, to guarantee that the content was indeed relevant to the domain. Table 1 shows the distribution of the final data used for annotation, consisting of 1177 news article comments.

Data Characteristics The average length of comments is 51 words, spanning 1-3 Arabic sentences. We do not correct the data for spelling errors; we annotate the raw text because we want to avoid any alteration that may affect the interpretation of sentiment, and we would like to keep the data as real as possible. However, it is possible to correct this output automatically or manually.

We performed a manual analysis of 100 comments from a randomly selected subset of the dataset and having the same domain distribution. We found that 43% of the comments contain at least one **spelling error** including typos, word merges and splits,[1] 15% contain at least one **dialect word**, 20% contain a **run-on sentence** not separated by any conjunction or punctuation, and 98% contain **subjective opinions** on any topic. We believe this is a good dataset for annotation because it contains real-world data, and many strong opinions on controversial topics.

4 Experimental Results

This section describes results and analyses of the crowdsourced annotations. We report the inter-annotator agreement at each of the two annotation stages, the distribution of the sentiment of collected targets by domain, and a manual analysis of our target entities. We also provide examples of our final annotations.

4.1 Inter-annotator agreement

Task 1: Agreement on Important Noun Phrases To compute the agreement between annotators on important entities in a HIT, we compute the average precision p_{HIT}. p_{HIT} is then averaged over all HITs to obtain the agreement.

$$p_{HIT} = \frac{1}{3} \cdot \left(\frac{\#matches}{\#phrases\_a1} + \frac{\#matches}{\#phrases\_a2} + \frac{\#matches}{\#phrases\_a3} \right)$$

An average precision of **0.38** was obtained using exact matching of entities and **0.75** using subset matching: i.e a match occurs if the three annotators all list a sub-phrase of the same noun phrase. (Recall that the final entities were chosen according to subset agreement.)

Our noun phrase agreement numbers are comparable to the target span subset agreement numbers of Somasundaran et al. (2008) in English discourse data, and lower than that of Toprak et al. (2010), who annotated targets in the consumer review domain. Note that besides the language difference, the task itself is different, since we annotate important noun phrases rather than opinion targets; a lower agreement on this task essentially indicates that fewer entities are being passed on to the next task for consideration as targets, the assumption being that only important entities will be agreed upon by all three annotators. Since we had three rather than two annotators, the agreement using exact match is expected to be low.

Domain	**# Entities**	**Majority Agree (%)**
Politics	3853	91.2
Culture	2271	95.8
Sports	1222	87.6
Total	**7346**	**91.8**

Table 2: Agreement on entity-level sentiment annotation

Task 2: Sentiment agreement Table 2 shows the annotator agreement for the task of identifying sentiment towards given entities. A majority agreement occurs when 3 out of 5 annotators of an entity agree on whether the sentiment towards it is positive, negative, or neutral. We see that the agreement (91.8%) is reasonably high. Abdul-Mageed and Diab (2011) have reported overall agreement of 88% for annotating sentence-level Arabic sentiment (as positive, negative, neutral, or objective) using two trained annotators. We note that after assigning our task to only the qualified group of workers, the annotator agreement increased from 80% and 88% in the first two annotation rounds, to 95% in the remaining rounds.[2]

Sentiment Distribution Table 3 shows the distribution of the sentiment of the final targets by domain. The final targets of opinions correspond to entities which were agreed to be **positive or negative by majority agreement.** We can see that the politics and sports domains are biased towards negative and positive sentiment respectively, while targets in the culture domain have a mostly even distribution of sentiment. We also note that overall, 95% of all comments had at least one target of opinion, and 41% of those comments had multiple targets with both positive and negative sentiment. This verifies our hypothesis about the sentiment diversity and need for finer-level opinion analysis for this dataset.

Finally, we found that the majority of targets are composed of 2 words (38% of targets), followed by 1-word targets (25% of targets), 3-word targets (18%), and 4-word targets (9%), while 10% of all targets are composed of more than 4 words.

4.2 Manual Analysis

We manually examined 200 randomly selected targets from our final dataset, and found a num-

[1] We don't count the different variations of *Alef* ا, ى/ي, or ة/ه, forms, which are often normalized during model training and evaluation.

[2] In the final dataset, we include the annotations organized by each annotation round. We mark the entities with disagreement as 'undetermined'.

Domain	**# Targets**	**(%) Pos**	**(%) Neg**
Politics	2448	30	70
Culture	1149	48	52
Sports	748	79	21
Total	**4345**	**43**	**57**

Table 3: Distribution of sentiment in final targets

Class	**Example**
Spelling errors **2.5%**	ارادت الشعب *“the people’s will"*
Punctuation **5%**	منتجات ابل. *“Apple’s products."*
Prep & Conj clitics **8.5%**	لمانشستر يونايتد *“to Manchester United"*
Non-noun phrases **3%**	البرشا بطل الدور الاسباني *“Barcelona (is) the champion of the Spanish league"*
Targets with sentiment **5.5%**	الشعب السوري الحر *"the free Syrian people"*
Propositional entities **3%**	تشجيع الباحثين *"encouraging researchers"*

Table 4: Target phrase observations

ber of observations, many of which are language-specific, that deserve to be highlighted. They are summarized in Table 4.

We first note orthographic observations such as spelling errors, which come mostly from the original text, and punctuations attached to targets, which may easily be stripped from the text. The punctuations result from our decision to take the maximal noun phrase provided by annotators.

Prepositional and conjunctional clitics result from Arabic morphology which attaches prepositions such as *l+* ل *(to)* and *b+* ب *(in)*, or conjunctions *w+* و *(and)* to the noun preceding them. They can be removed by tokenization (Habash, 2010), but we preserve them for completeness and their usefulness for allowing us to distinguish between different mentions of the same target.

Non-noun phrases mainly come from nominal sentences specific to Arabic syntax جملة اسمية ; these are problematic because they may be interpreted as either noun phrases or full sentences that begin with a nominal. We also observed a number of verbal phrase targets (e.g “نبلبل بالديموقراطية" *“we confuse democracy"*), but these were very few; the majority of this class of observations comes from verbless nominal phrases.

Targets containing sentiment words appear since sentiment words can be part of the noun phrase and are not always independent of the topic itself. As for propositional entities, they result from process nominals مصدر which can have a verbal reading (Green and Manning, 2010) but are correctly considered to be nouns. We find that they occur mostly in the culture domain, where more discussions occur about ‘important concepts’.

We also found from our manual inspection that our final entity spans reasonably corresponded to what would be expected to be targets of opinions for the topic in context. From our 200 randomly selected targets, we found 6 cases where the polarity of the noun phrase potentially negated the polarity towards a shorter entity within the noun phrase. However, in most of these cases, the noun phrase resolved from the annotations correctly represents the actual target of opinion: e.g.“*depletion of* ***ozone***" ثقب الاوزون, “*bombing of* ***houses***" قصف المنازل, and “*methodology of teaching* ***Arabic***" اسلوب تعليم العربية. We found one case “*absence of* ***Messi****"* غياب مسي, labeled negative, where it could be argued that either *Messi* (positive) or his absence (negative) is the correct target. We generally preferred target annotations which correspond to the topic or event being discussed in the context of the comment.

Examples We provide examples of the annotations, shown in Table 5. Note that we have preserved all spelling errors in the original Arabic text. As it is common in Arabic to write very long sentences, we have added punctuation to make the English translation more readable.

Example (1) is from the culture domain. We see that it summarizes the writer’s opinions towards all important topics. Note that the direct reference to the target “*e-book"* الكتاب الالكتروني is the first mention (the second mention is preceeded by the preposition *to* ل). However, we generally assume that the opinion towards a target is deduced from the entire comment (i.e from both the phrase ‘*despite the popularity of the e-book*’ and the phrase ‘*there is no place for an e-book in my dictionary*’). Ideally, the annotators should also have marked *traditional book* الكتاب التقليدي as a positive target; although the opinion expressed towards it is less direct, it can also be inferred by co-reference with *paper book* الكتاب الورقي .

Example (2) lists an entity that doesn’t appear in the text *“(to) the Arab team the world cup"*

للمتخب العربي المونديال; this likely results from an error in Task 1 where the phrase got picked up as the maximal common noun phrase. The annotator might have meant that *Arab team in the world cup* is a topic that the writer feels positively about; however, our current annotation scheme only considers entities that strictly appear in the text. We also see that annotators disagreed on the polarity of the propositional entity *“either team qualifying"* تأهل الفريقين, likely because they were not sure whether it should be marked as neutral or positive. In addition, this example contains an over-generated target *“world cup"* المونديال, which would have been best marked as neutral.

Example (3) is from the politics domain. It correctly annotates multiple references of *the Iraqi government* and captures the sentiment towards important entities in the text. The target *“the only neighboring country"* الدولة الجارة الوحيدة can be considered an over-generation; a better interpretation might be to consider this phrase part of the opinion expression itself (*"the only neighboring country with whom we have ties that are not just based on interests* is Turkey"). Nonetheless, this extra annotation may provide helpful information for future modeling. Notice that the Arabic comment for this example, in addition to being long, has no punctuation other than the period ending the sentence. It is common in Arabic to encounter such constructions, whereby conjunctions and transitional words are enough to determine the separation between clauses or sentence phrases. We have added punctuation to the English translation of this example.

We generally found that the annotations were a good representation of the diverse opinions of online writers, correctly covering sentiment towards essential targets and mostly complying with our definition of entities. The annotations contain some errors, but these are expected in a crowdsourcing task, especially one that relies so heavily on subjective interpretation. We noticed that annotators tended to over-generate targets rather than miss out on essential targets. We believe that even annotation of secondary targets may prove useful for future modeling tasks.

5 Conclusions

We developed a two-stage method for annotating targets of opinions using Amazon Mechanical Turk, where we consider targets to be noun phrase entities. This method was applied to Arabic, yielding a new, publicly available resource for fine-grained opinion analysis.[3] We found high agreement on the task of identifying sentiment towards entities, leading to the conclusion that it is possible to carry out this task using crowdsourcing, especially when qualified workers are available.

Unlike some of the previous work, our focus was on annotating target entities rather than the full target spans; and we developed a unique approach for identifying these entities using Amazon Mechanial Turk. The first task involves marking important entities, while the second task involves finding targets by assessing the sentiment towards each entity in isolation. We found that although the agreement was generally high for both tasks, it was not as high for the entity identification task as it was for the second and easier task of finding sentiment towards entities.

We also found that the morphological complexity of Arabic, as well as the variation in acceptable syntax for noun phrases, creates additional annotation challenges for deciphering the boundaries of entities. We also anticipate that the long structure of Arabic comments will create interesting challenges for future modeling tasks.

In the future, we hope to extend this dataset by mapping the targets to specific opinion phrases and identifying which targets refer to repeated mentions (e.g *the team*) or aspects (e.g *defense*) of the same target (e.g *the Algerian team*), in addition to annotating conflicting sentiment towards the same entity. We also hope to create a manually reviewed version of the corpus corrected for spelling errors and non-noun phrase targets.

6 Acknowledgments

This work was made possible by grant NPRP 6-716-1-138 from the Qatar National Research Fund (a member of the Qatar Foundation). The statements made are solely the responsibility of the authors. We thank anonymous reviewers for their helpful comments. We would also like to thank Debanjan Ghosh, Owen Rambow, and Ramy Eskander for helpful discussions and feedback. We thank the AMT annotators for all their hard work, insightful questions, and for continuing to participate in multiple rounds of our annotation tasks.

[3] The corpus is available and can be downloaded from www.cs.columbia.edu/~noura/Resources.html

	Example Comment
Example (1) Domain: Culture	رغم انتشار **الكتاب الألكتروني** الا ان **الكتاب الورقي** اثبت وجوده. احب **الكتاب المطبوع** .. حتى تقليب صفحاته أجد بها متعة.. والأجمل عند قراءته وهو بين يدي .. لا أحتمل **قراءة الكتاب من خلال الشاشة** .. لا أستطيع الاستمرار في تحمل وهج الضوء والصداع.. الكتاب التقليدي أقراءه في المكتبة في القطار في الطائرة على الشاطئ في الحديقة في اي مكان أرتاح فيه .. لامكان للكتاب الألكتروني في قاموسي.
English Translation	Despite the popularity of **the e-book**, **the paper book** has proven itself. I like **the printed book**... I even find a pleasure in turning its pages ... and it is nice is to read it while it is in my hands ... I cannot stand **reading a book through a screen** ... I cannot bear the glare of light and the headaches...I can read a traditional book in the library on the train in the airplane on the beach in the garden in anywhere I am comfortable .. there is no place for the e-book in my dictionary.
Annotated Targets	**negative:** the e-book الكتاب الالكتروني **positive:** the paper book الكتاب الورقي **positive:** the printed book الكتاب المطبوع **negative:** reading a book through a screen قراءة الكتاب من خلال الشاشة
Example (2) Domain: Sports	**المنتخبان المصري والجزائري** هما منتخبان قويان. والدعم الدي حضي به **المنتخب الجزائري** بالمناسبة جعل الكل مثوثر ولايوجد فرق في تأهل الفريقين و ا تمنى ان يتأهل **الفريق الجزائري** الي **المونديال** لأنّني احب **الفريق الجزائري** الى جانب المنتخب المصري . والمهم التمثيل الجيد و ا تمنى ان يكون **للمتخب العربي** احسن تمثيل في **المونديال** .
English Translation	**The Egyptian and Algerian teams** are strong teams. The support gained by the **Algerian team** for this occasion has made everyone nervous and there is no difference in either team qualifying and I hope that **the Algerian team** gets qualified to **the world cup** because I like **the Algerian team** alongside the Egyptian team. The important thing is good representation and I hope that **the Arab team** will be best represented in **the world cup**.
Annotated Targets	**positive:** The Egyptian and Algerian teams المنتخبان المصري والجزائري **positive:** the Algerian team 'elect' المنتخب الجزائري **positive:** the Algerian team الفريق الجزائري **positive:** the world cup المونديال **positive:** (to) the Arab team the world cup للمتخب العربي المونديال **undetermined:** either team qualifying تأهل الفريقين
Example (3) Domain: Politics	مع الاسف **الحكومة العراقية** لا يفتهم من السياسة شيء لأن **الدولة الجارة الوحيدة** التي تربطنا معها اكثر من مصالح من الموارد الطبيعية كالمياه الى مصالح صناعية هي **تركيا** فعلينا ان نقوي علاقتنا معها لانها اصبحت تنافس الدول الاوربية لنستفاد منها ولكن **حكومة المالكي الفاشلة** لا يهمهم التطور وقد رجع **العراق** بظل هؤلاء مئات السنين الى الخلف.
English Translation	Unfortunately **the Iraqi government** understands nothing of politics because **the only neighboring country** with whom we have ties that are not just based on interests - such as natural resources like water and industrial interests - is **Turkey**, so we have to strengthen our relationship with it because it is now a competitor with European nations, we should benefit from it but **Maliki's failed government** cares nothing for progress and **Iraq** has gone back hundreds of years because of these people.
Annotated Targets	**negative:** the Iraqi government الحكومة العراقية **positive:** the only neighboring country الدولة الجارة الوحيدة **positive:** Turkey تركيا **negative:** Maliki's failed government حكومة المالكي الفاشلة **negative:** Iraq العراق

Table 5: Examples of Annotations. The original spelling errors are preserved.

References

Muhammad Abdul-Mageed and Mona T Diab. 2011. Subjectivity and sentiment annotation of modern standard Arabic newswire. In *Proceedings of the 5th Linguistic Annotation Workshop*, pages 110–118. Association for Computational Linguistics.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022.

Paula Carvalho, Luís Sarmento, Jorge Teixeira, and Mário J Silva. 2011. Liars and saviors in a sentiment annotated corpus of comments to political debates. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 564–568. Association for Computational Linguistics.

Lingjia Deng and Janyce Wiebe. 2015. Mpqa 3.0: An entity/event-level sentiment corpus.

Michael Denkowski, Hassan Al-Haj, and Alon Lavie. 2010. Turker-assisted paraphrasing for English-Arabic machine translation. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, pages 66–70. Association for Computational Linguistics.

Spence Green and Christopher D Manning. 2010. Better Arabic parsing: Baselines, evaluations, and analysis. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 394–402. Association for Computational Linguistics.

Nizar Y Habash. 2010. Introduction to Arabic natural language processing. *Synthesis Lectures on Human Language Technologies*, 3(1):1–187.

Chiara Higgins, Elizabeth McGrath, and Lailla Moretto. 2010. MTurk crowdsourcing: a viable method for rapid discovery of Arabic nicknames? In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, pages 89–92. Association for Computational Linguistics.

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177. ACM.

Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. 2011. Target-dependent twitter sentiment classification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 151–160. Association for Computational Linguistics.

Soo-Min Kim and Eduard Hovy. 2006. Extracting opinions, opinion holders, and topics expressed in online news media text. In *Proceedings of the Workshop on Sentiment and Subjectivity in Text*, pages 1–8. Association for Computational Linguistics.

Nolan Lawson, Kevin Eustice, Mike Perkowitz, and Meliha Yetisgen-Yildiz. 2010. Annotating large email datasets for named entity recognition with mechanical turk. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, pages 71–79. Association for Computational Linguistics.

Andrew K McCallum. 2002. {MALLET: A Machine Learning for Language Toolkit}.

Behrang Mohit, Alla Rozovskaya, Nizar Habash, Wajdi Zaghouani, and Ossama Obeid. 2014. The first QALB shared task on automatic text correction for Arabic. In *Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP)*, pages 39–47, Doha, Qatar, October. Association for Computational Linguistics.

Maria Pontiki, Haris Papageorgiou, Dimitrios Galanis, Ion Androutsopoulos, John Pavlopoulos, and Suresh Manandhar. 2014. Semeval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35.

Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. 2011. Opinion word expansion and target extraction through double propagation. *Computational linguistics*, 37(1):9–27.

Josef Ruppenhofer, Swapna Somasundaran, and Janyce Wiebe. 2008. Finding the sources and targets of subjective expressions. In *LREC*.

Swapna Somasundaran and Janyce Wiebe. 2009. Recognizing stances in online debates. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1*, pages 226–234. Association for Computational Linguistics.

Swapna Somasundaran, Josef Ruppenhofer, and Janyce Wiebe. 2008. Discourse level opinion relations: An annotation study. In *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, pages 129–137. Association for Computational Linguistics.

Veselin Stoyanov and Claire Cardie. 2008. Annotating topics of opinions. In *LREC*.

Cigdem Toprak, Niklas Jakob, and Iryna Gurevych. 2010. Sentence and expression level annotation of opinions in user-generated discourse. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 575–584. Association for Computational Linguistics.

Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, 39(2-3):165–210.

Theresa Ann Wilson. 2008. *Fine-grained subjectivity and sentiment analysis: recognizing the intensity, polarity, and attitudes of private states*. ProQuest.

Wajdi Zaghouani, Behrang Mohit, Nizar Habash, Ossama Obeid, Nadi Tomeh, Alla Rozovskaya, Noura Farra, Sarah Alkuhlani, and Kemal Oflazer. 2014.

Large scale Arabic error annotation: Guidelines and framework. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, May. European Language Resources Association (ELRA). ACL Anthology Identifier: L14-1721.

Omar F Zaidan and Chris Callison-Burch. 2011. The Arabic online commentary dataset: an annotated dataset of informal Arabic with high dialectal content. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 37–41. Association for Computational Linguistics.