Addressing Annotation Complexity: The Case of Annotating Ideological Perspective in Egyptian Social Media

Heba Elfardy
Department of Computer Science
Columbia University
New York, NY
`heba@cs.columbia.edu`

Mona Diab
Department of Computer Science
The George Washington University
Washington, DC
`mtdiab@gwu.edu`

Abstract

Automatically detecting the stance of people toward political and ideological topics –namely their "Ideological Perspective"– from social media is a rapidly growing research area with a wide range of applications. Research in such a field faces several challenges among which is the lack of annotated corpora and associated guidelines for collecting annotations. The problem is even more pronounced in situations where there is no clear taxonomy for the common community perspectives and ideologies. The challenges are exacerbated when the communities where we need to gather these annotations are in a state of turmoil causing *subjectivity* and *intimidation* to be factors in the annotation process. Accordingly, we present the process for creating a robust and succinct set of guidelines for annotating "Egyptian Ideological Perspectives". We collect social media data discussing Egyptian politics and develop an iterative feedback annotation framework refining the annotation task and associated guidelines attempting to circumvent both weaknesses. Our efforts lead to a significant increase in inter-annotator agreement measures from 75.7% to 92% overall agreement.

1 Introduction

With the rise of social media there has been a plethora of documented political and ideological discussions. These discussions typically represent polarizing topics and in doing so convey the participants' belief systems expressing their perspective (or stance) on contentious issues –namely their "Ideological Perspective". Identifying the perspective of users in such media is a challenging research problem that has a wide variety of applications from recommendation systems and targeted advertising to planning political campaigns, political polling and predicting possible future events. As a matter of fact, social media played a major role in the Arab Spring (2010– present). In Egypt, for example, activists and political leaders resorted to social media as an alternative to the censored and mostly biased state and privately owned media. Most of these activists used social media to make announcements, campaign for elections, spread awareness of important causes and conduct polls in order to predict election outcomes. After Egypt's Jan. 25th Revolution, alliances kept forming (and later breaking) between Islamist movements, Revolutionists, public figures from Mubarak's regime (the Old guard) and the Army. The formation and break-up of such alliances often triggered apparent perspective-shifts in the public sphere. These shifts in perspective can be best explained by Converse's concept of centrality in belief systems. Converse (2006) defines a belief system as the configuration of idea elements and attitudes that are bound together by some constraint. This constraint helps us in knowing that a person holds a specific attitude given knowledge that he/she holds another one (Converse, 2006). For example, if we know that an American citizen supports ObamaCare, can we predict that he/she supports gun control? While there are Americans who support ObamaCare and oppose gun control, the vast majority of people either support or oppose both issues because the stance toward these two issues is always backed by one's ideology or belief system, namely being of a Democratic Party leaning. Converse states that within a belief system, idea elements vary in "centrality". These variations always govern what happens when the status of one of the idea elements in a belief system changes. For example, what will a self-proclaimed Republican do if the Republi-

79

can Party decided to change its stance on universal healthcare and started to support it? The reaction of the person will depend on which is more central to the person's belief system – political party affiliation or stance on healthcare. Many Egyptians were faced with such choices post the Jan. 25th Revolution as the stance of political leaders toward major political entities such as the Military, the Police, Islamists, the Revolution, etc. kept changing. This change of stance among the leaders often triggered perspective shifts among the mass public toward the entities that are less central to their belief systems.

Collecting annotations of such perspectives is quite challenging in a dynamic political setting since many of the political stances are emergent and shifting. The problem is two fold: (1) pinning down what the perspectives are; and, (2) gathering annotations on such perspectives while circumventing the subjectivity of the annotators themselves. Due to the nature of the data, we need to use the help of annotators who understand the nature of the political landscape, hence they had to be Egyptians familiar with the recent events. But by being Egyptian, they are not themselves naturally divorced from the events, thereby having their own perspectives and biases. In this paper, we present our iterative approach to building effective guidelines for collecting annotations that aims at decoupling the annotation process from possible subjective assessment of the annotators. We build a list of major political events and sample a set of social media data that was posted within one week from the start of each of these events. We come up with a hypothesis on the most important elements governing the Ideological Perspective of most Egyptians and develop a set of guidelines and an annotation task to identify the perspective from which a given comment was written. Our hypothesis is that a person's perspective has two major underlying dimensions: (1) a person's stance on political reform versus stability; and, (2) a person's stance on the role Islam/religion should play in the public sphere, in politics. We run our first annotation experiment where we ask annotators to identify the stance of a given comment toward several political entities such as Jan. 25th Revolution, Mubarak's Regime, Military Rule, Islamists and Secularists. Based on the feedback and error analysis of this pilot annotation, we note some interesting observations most impactful of which is the annotators'

having significant reservations in making a judgment on comments. Accordingly, taking this feedback into consideration, we refine the guidelines and the annotation task and have the same set of comments annotated based on the refined guidelines. Given the new set of guidelines, annotators are asked to identify the top priorities expressed in the comment such as stability, supporting (or opposing) Islamists, supporting Jan. 25th Revolution, etc. The new task and guidelines yield better inter-annotator agreement and annotators give a more positive feedback on the clarity of the task.

2   Related Work

From a social-science viewpoint, the notion of "Perspective" is related to the concept of "Framing". Framing involves making some topics –or some aspects of the discussed topics– more prominent in order to promote the views and interpretations of the writer (communicator). (Entman, 1993). At the most basic level, these decisions are expressed in lexical choice. For example, a person who opposes gun rights is more likely to use words that emphasize "death" while a supporter is more likely to use ones that promote "self defense". As the saying goes, *"One man's terrorist is another man's freedom fighter"*. Perspective is also expressed on the syntactic and semantic levels. Greene and Resnik (2009) showed that the syntactic structure can be a strong indicator of a specific perspective, or bias. For example, using the passive voice puts less emphasis on the doer than using an active one. This is particularly important when the verb is sentiment bearing. In such case, the passive voice is less likely to associate the sentiment with the doer. Sentiment in itself serves as another important cue for identifying a person's perspective since it expresses one's opinion on different topics. In fact, from a computational point of view, the work on perspective-detection is closely related to subjectivity and sentiment analysis. One's perspective normally influences his/her sentiment toward different topics or targets. Conversely identifying the sentiment of a person toward multiple targets can serve as a cue for identifying this person's perspective. For example, we expect a typical Jan. 25th Revolutionist to express positive sentiment toward social justice, freedom of speech and the Revolution's public figures and negative sentiment toward the ousted ex-president Mubarak of Egypt and his regime.

Event	Date Range
1. Jan. 25th Revolution	Jan. 25 - Jan. 31, 2011
2. Battle of the camel	Feb. 2 - Feb. 8, 2011
3. Mubarak Stepping Down	Feb. 11 - Feb. 17, 2011
4. Referendum on amendments to old constitution	Mar. 19 - Mar. 25, 2011
5. Mohamed Mahmoud Protests *(Clashes between Army and Revolutionists)*	Nov. 19 - Nov. 25, 2011
6. Announcement of presidential election results	Jun. 24 - Jun. 30, 2012
7. Presidential decree and associated protests	Nov. 22 - Nov. 28, 2012
8. Ousting of President Mohamed Morsi	Jun. 30 - Jul. 6, 2013
9. Army calls for mandate to crack down on terrorism	Jul. 24 - Jul. 30, 2013
10. Rabia (Pro-Muslim Brotherhood) camp dismantling	Aug. 14 - Aug. 20, 2013

Table 1: List of events and their associated dates for which the data was selected.

Most of the currently available datasets that are annotated for Ideological Perspective are in English (Lin et al., 2006; Somasundaran and Wiebe, 2010; Abu-Jbara et al., 2012; Yano et al., 2010; Elfardy et al., 2015; Hasan and NG, 2012; Hasan and Ng, 2013). The only Arabic Ideological Perspective datasets that we are aware of are those of Abu-Jbara et al. (2013), Siegel (2014) and Borge-Holthoefer et al. (2015). Abu-Jbara et al. (2013)'s dataset is self annotated, and the annotations are more abstract –only provide the binary stance of each post toward the debate question. Siegel (2014) study whether Egyptian twitter users who are exposed to a more diverse twitter network become more tolerant toward people having different political and ideological leanings. On the other hand, Borge-Holthoefer et al. (2015) collect Arabic tweets posted between June and September 2013. They manually annotate a subset of 1000 tweets as either supporting, opposing or being neutral toward the Military and use this subset to build a classifier that they then apply to all unlabeled tweets. The authors then track users who change their position toward the Military across the studied timeline. As opposed to our work, Borge-Holthoefer et al. (2015)'s work focuses on a much shorter time-frame during which most people were polarized between either supporting the Military or supporting the Muslim Brotherhood (Islamists) hence the authors do not aim to identify the stance of the author toward other political entities such as Jan. 25th Revolution, Mubarak's regime, etc.

To the best of our knowledge, the presented work is the first attempt at creating guidelines for collecting fine-grained multidimensional annotations of Egyptian Ideological Perspectives that try to uncover the different underlying elements of a person's belief system.

3 Data Collection

We select a set of public social media discussion fora pages of renowned Egyptian activists and politicians of different political leanings and curate posts and comments from these pages. The "*post*" refers to some piece of content shared on a page while the "*comment*" is a response to this original piece of content. We filter spam/repetitive comments that do not respond to the original post. Moreover, only comments with no Latin words and that have a length of at least ten words were preserved.

After the initial cleanup of the data, we use a list of major events such as Jan. 25th demonstrations, major protests, Presidential elections, etc. to select our final dataset. Table 1 shows the list of events and the dates covered by the selected data. We split the data into two groups based on whether it was curated from a page that supports (1) Reform [RFM] (Supporting Jan. 25th Revolution); or, (2) Old Guard Rule [OGR] (ex. Supporting the ousted Egyptian President Mubarak and his regime, or supporting the current Egyptian President –Sisi– who was the ex-minister of Defense). We then select a sample of 31 comments per event for each of the two groups. It is worth mentioning that for the first event –Jan. 25th Revolution– no comments were posted in the pro-OGR pages accordingly we only have 31 pro-RFM comments for this event. This results in a total of 310 RFM and 279 OGR comments.

4 Egyptian Ideological Perspectives

Prior to collecting the annotations, we come up with a high level taxonomy for the most common

- All questions target the *comment*. (The *post* is meant to give you context)
- Please pay attention to the *post* and *comment* dates.
- Use your knowledge of the political events in Egypt when responding to the questions.
  ex. If a *comment* supports Jan. 25th Revolution and you know that this implies that it opposes Mubarak's regime then choose "*Oppose*" as an answer to Q4.
- If the answer to Q1 or Q2 is "*No*", then choose "*NA*" as an answer to all other questions
- Difference between "*NA/Does not apply*" and "*Not Sure*":
  - "*NA*" should be used when the *comment* does not discuss the subject of the question
    ex. If a given *comment* does not discuss Mubarak's regime then you should choose "*NA*" as an answer to Q4.
    If, on the other hand, the *comment* discusses Mubarak's regime but you are not sure whether it opposes it or supports it then choose "*Not Sure*"
- Q7 targets Military Rule at any point in time (not a specific Army leader)
- If a *comment* supports Islamists this does not necessarily mean that it opposes Seculars and vice versa. (Unless the author expresses anti-secular views)
- If you have any feedback, please respond to Q8.

Figure 1: Synopsis of annotation guidelines for Pilot annotation task

political leanings in Egypt for this timeframe. We base our taxonomy on the works of "The Hariri Center at the Atlantic Council",[1] and "Carnegie Endowment for International Peace".[2] As mentioned earlier, after Jan. 25th Revolution, the formation and breakup of alliances between different political entities resulted in a dynamic set of political leanings hence created a need for a dynamic classification. For the context of this paper, we reduce the very rich perspective map of a person to two underlying dimensions: (1) stance toward democracy and political reform versus stability at the expense of loss of civil liberties; (2) stance toward the role played by Islam/religion in the public sphere or politics, namely Islamist vs. Secular. Accordingly, we assume that these two dimensions constitute a person's perspective. So for example, a person can oppose involving Islam in politics and support political reform. Another person can focus on stability even if it brings autocracy while either supporting or opposing Islamists. As mentioned earlier, the dimension that is less central to a person's belief system is more likely to change over time.

5 Annotation

Noting how challenging the annotation will be, we wanted to get a sense of how to circumvent annotator bias. Accordingly we devise an iterative feedback loop for the annotation process. We first have the sampled comments annotated by four trained Egyptian annotators. We ask the annotators to self identify what their own positions are with respect to the two dimensions of interest. All annotators indicate that they support Jan. 25th Revolution. Additionally, three annotators (annotators 1-3) indicate that they are neutral toward the role of Islam in politics while the fourth annotator indicates support toward the Army's leadership in ousting Islamists. An annotation lead managed the process of (1) training the annotators, (2) relaying their feedback about the clarity of the task to the authors. Based on the feedback and inter-annotator agreement (IAA) from this round, we refine the guidelines and annotation task before having the same data annotated by the same set of annotators.

5.1 First Annotation Experiment

For each task, we present annotators with a post and an associated comment. Except for one optional question that asks for feedback about the overall annotation task, all questions are formatted as multiple choice and require one answer to be provided. We do not reveal the leaning of the source page from which the comments were curated to the annotators so as not to bias their judgments. Annotators were asked to answer the following questions for each task:

- Q1: Does the given *comment* discuss Egyptian politics? (Yes/No)
- Q2: Is there enough context to determine the political leaning of the *comment*? (Yes/No)

---

[1] http://www.atlanticcouncil.org/blogs/egyptsource/egyptian-politics

[2] http://carnegieendowment.org/2015/01/22/2012-egyptian-parliamentary-elections/

Does the given *comment* Support/Oppose/Not Sure/Not Applicable:

- Q3: Jan. 25th Revolution?
- Q4: Mubarak's regime?
- Q5: Seculars?
- Q6: Islamists?
- Q7: Military Rule?

• Q8: Do you have any feedback or suggestions?

Questions 3-7 aim to identify the two previously discussed dimensions that define a person's perspective. Questions 3, 4 and 7 attempt to uncover the first dimension –the person's position on political reform and democracy while questions 5 and 6 aim to identify the second dimension –the person's view on the role of Islam/religion in the political sphere/government.

Since the task is quite subjective, we tried to cover most possible scenarios and to provide examples in our guidelines. Moreover we attempted to the best of our knowledge to avoid any bias in the way the questions were phrased. Figure 1 shows the guidelines for this first annotation experiment.

5.2 Error Analysis

We calculate the pairwise and overall IAA for all questions. Table 2 shows the results. The average pairwise IAA for all questions is quite high ranging from 84.1% to 88.4%. However, achieving a complete-row agreement (Row) by all annotators is quite challenging. The four annotators achieved a perfect row agreement –chose the same answers for all questions pertaining to a particular comment– on only 25.5% of the comments. We also note that Annotator 1 and 3 exhibit the most agreement.

In order to get better insights into the source of disagreement between annotators, we perform a manual error analysis by looking into the confusable comments and find that most of them fall under the following categories:

1. Comments that provide cues for both supporting and opposing the topic the question is addressing **ex.** (Event 2)

لازم نصبر ونشوف اللى حيحصل مفيش حاجة بتتغير بين يوم وليلة براحة علشان اللى عملناه ميتقلبش ضدنا يا جماعة خافوا على البلد شوية عايزين نعمرها تانى

Translation: *We have to be patient and wait and see what will happen. Nothing changes in a day and night. Take it easy so what we did does not backfire on us. Care about the country. We need to rebuild it.*

While above comment opposes the continued demonstrations, this does not necessarily mean that it opposes Jan. 25th Revolution since the author just prioritizes stability over immediate political reform.

2. Ambiguous pronouns **ex.** (Event 9)

وقيادتهم اللى دفعوا بهم للتجارة بدمهم اليسوا اول المسئولين؟ واول من يستطيعوا ان يوقفوا اهدار دمهم وعدم دفعهم للانتحار

Translation: *And their leaders that pushed them in order to sell their blood, aren't they the responsible ones? They could have stopped their bloodshed if they didn't push them to commit suicide.*

In this comment, although "their leaders" refers to leaders of the Muslim Brotherhood, it can be easily confused with the Army leaders.

3. Comments where the stance toward one entity is implied from the stance toward another entity **ex.** (Event 6)

يا ما انا فرحانة فيكوا يا وفي شفيق العتيه يا عبيييييد يا حرامية

Translation: *I am gloating over the loss of the idiot Shafik, you slaves and thieves*

In the above comment, the author gloats over the defeat of Ahmed Shafik (a key figure of the OGR) in the 2012 presidential elections.. While the comment clearly supports Jan. 25th Revolution and opposes Mubarak's regime, it is not clear whether the author actually supports or is indifferent toward the Muslim Brotherhood's candidate.

4. Authors that report the opinions of other people by quoting them instead of stating their own opinions in comments they post. **ex.** (Event 7)

	Q1 Egy. Politics	Q2 Context	Q3 Jan. 25th	Q4 Mubarak	Q5 Seculars	Q6 Islamists	Q7 Military Rule	Avg	Row
Ann.1-2	95.6	81.8	80.3	76.1	96.8	80.6	79.5	84.4	44.3
Ann.1-3	96.8	87.6	81	83.4	97.8	86.1	86.1	88.4	55.5
Ann.1-4	97.1	86.1	81.2	81.5	97.5	83	78.8	86.4	48.6
Ann.2-3	95.8	82.3	77.2	72.8	97.3	82.7	80.3	84.1	42.3
Ann.2-4	96.4	83.5	87.3	79.8	98.8	82.7	78.8	86.8	47.5
Ann.3-4	98	84.9	80	81.2	97.3	84.9	81.3	86.8	49.4
All Ann.	93.5	71.1	66.9	64.3	95.9	72.2	65.9	75.7	25.5

Table 2: Inter-Annotator agreement for the pilot annotation experiment

	Pro-RFM Pages				Pro-OGR Pages			
	Yes	No			Yes	No		
Q1. Egy. Politics	97.6	2.4			97.8	2.2		
Q2. Context	84.4	15.6			81.5	18.5		
	Support	Oppose	Not Sure	NA	Support	Oppose	Not Sure	NA
Q3. Jan. 25th Revolution	42.9	2.6	0.3	54.2	3.9	32.2	0.8	63.1
Q4. Mubarak	1.9	43.5	1.5	53.1	30.1	7.7	1.5	60.7
Q5. Seculars	0.2	2.6	0.2	96.9	0	1.3	0.3	98.4
Q6. Islamists	27.7	11.2	1.7	59.4	9.9	33.9	0.8	55.5
Q8. Military Rule	1.2	11.3	0.6	86.9	22.6	5.8	0.4	71.1

Table 3: Answer Distribution (averaged over all annotators) to each question in the pilot annotation split according to the leaning of the source page from which data is curated.

وكالات انباء عالية: عدد المتظاهرين المعارضين لمرسي في التحرير يفوق عدد مؤيديه عند الأتحادية.

Translation: *International News Agencies: "The number of anti-Morsi protestors in Tahrir exceeds the number of his supporters at the Heliopolis Palace"*

5. Sarcastic comments where the annotator judges the comment based on the literal and not the intended meaning;

6. Comments that oppose a certain group of Islamists (ex. Muslim Brotherhood) and oppose other ones (ex. Salafis). To handle these cases, the annotation task should provide a "Mixed Views" option to Q6 (a comment's stance on Islamists).

5.3 Qualitative Assessment

To perform a qualitative assessment of the annotations, we begin by calculating the distribution of the answers to all questions. We further split the comments according to whether the source pages they were collected from support OGR or RFM. One should note that even if a page supports democracy this does not necessarily mean that all people who comment on that page share the same views. However, we do expect a higher number of pro-RFM authors to comment on the pro-RFM pages and vice versa. Table 3 shows the distribution. By analyzing the responses, we find that the majority of the given comments (>97%) discuss Egyptian politics, which indicates that our filtration process works well in excluding spam and irrelevant comments. Moreover the majority of comments (>84%) provide enough context to determine their stance. Another observation is that annotators are very conservative in using "*Not Sure*" category. As expected, we find a much higher percentage of comments that support Mubarak's regime and Military Rule and oppose Jan. 25th Revolution among the ones collected from pro-OGR pages. On the contrary, the majority of comments from pro-RFM pages that express a stance toward the different political entities support Jan. 25th Revolution and oppose both Military Rule and Mubarak's Regime. While pro-RFM pages have a higher percentage of comments that support Islamists (27.7%) and pro-OGR pages have a higher percentage of anti-Islamists comments (33.9%), a considerable number of comments in each of these pages follow the opposite trend. 11.2% of comments in pro-RFM pages oppose Islamists and 9.9% of

those in pro-OGR pages support them.

We analyze the answers per event and find that the distribution of the answers aligns with our knowledge of the political events in Egypt. For example, we expect and find a higher percentage of "*NA*" for Q4 (Mubarak's Regime) as we move away from the start of Jan. 25th Revolution and more polarization on the stance toward Islamists for events 8 through 10. Almost all comments pertaining to the first three events do not convey any stance toward Islamists. In the days right after the start of Jan. 25th Revolution most of the discussions addressed political reform versus stability and not the role of religion in politics. For events 7 through 10 more comments express a stance toward Islamists. For "Event 6" (announcing the results of presidential elections in which the Muslim Brotherhood's candidate was elected) the majority of comments sampled from pro-RFM pages support Islamists indicating acceptance of the election outcome while the pro-OGR pages express negative stance toward Islamists indicating disappointment in election outcomes, namely, disappointment that the OGR candidate –former Prime Minister– Ahmed Shafik lost.[3]

5.4 Pilot Annotation Weaknesses

Based on the feedback collected from the annotators and our manual error analysis, we notice the following problems with the way the task is formulated:

- The main point of confusion among annotators is deciding when they should infer the stance of the comment toward an entity based on the stance toward another entity. For example, if a person opposes the Army during Morsi's presidency term, does it imply that he/she supports Islamists;

- The task does not model the people who mainly care about stability regardless of political reform or the role of religion in politics;

- Even though the comments were collected from a specific set of events, we do not present the annotators with the event each comment was discussing and rather relied on the comment-date and the annotators' knowledge of the timeline of political events in Egypt;

- Q7 (A comment's stance on Military Rule) relied to a great extent on each annotator's interpretation of the Military Rule. A better way to phrase the question is to simply ask about the comment's stance toward the Military leaders and tap into our knowledge of the political timeline in Egypt in order to identify the periods where the Army/Military was actually in charge of governance;

- Most of the comments we looked at expressed the author's top priority whether it is political reform, stability, supporting the army, opposing the intervention of religion in political governance, etc. but our task gives equal weight to all political entities and do not ask annotators the top priority that they think drives the author's stance on various issues;

- Annotators were tempted to choose "*NA*" for many comments because they were trying to identify the reason behind a comment's stance. For example, a comment might support Islamists during Rabia camp dismantling because the author is against civil rights infringement but not necessarily because that person is pro-Islamists in general. We clarified to the annotators that we are only interested in the stance of the given comment at the time of the event of interest, namely in the specific context of the comment, regardless of the reason behind this stance or the person' stance at other points in time. Hence changing the question from a confusable potential "why" question to a "what" question. As mentioned earlier, this might also reflect the annotators' own concern over expressing their opinion about the comments with such a contentious event, erring on the side of caution;

- Some annotators chose "Yes" as an answer to Q2 (Is there enough context to judge the comment) when they were able to identify the sentiment of the comment but not the target of the sentiment. We clarified that if knowing the target is needed to identify the leaning of the comment then they should choose "No" as the answer to Q2;

- The guidelines do not address the cases where a comment shows mixed views on different Islamist groups/parties;

- Finally, the task does not address how the cases of reported opinions should be handled.

---

[3]In the interest of space, we do not show the distribution per event.

	Q1 Egy Polit.	Q2 Context	Q3 Reported Op.	Q4 Priority	Q5 Jan.25	Q6 Mubarak	Q7 Army	Q8 Islamists	Avg	Row
Ann. 1-2	99.3	95.1	95.1	89.5	92.9	92.9	95.4	94.9	94.4	82.7
Ann. 1-3	99.2	97.5	97.1	92.9	94.2	94.2	96.3	95.6	95.9	86.2
Ann. 1-4	99.2	94.6	94.1	88.6	91.7	91.9	94.6	94.4	93.6	80.8
Ann. 2-3	99.5	95.6	95.4	89.5	94.4	94.4	96.1	94.7	94.9	83.9
Ann. 2-4	99.5	97.1	96.6	92.0	95.9	95.8	97.1	95.6	96.2	86.9
Ann. 3-4	100.0	95.4	95.2	91.5	95.1	94.9	96.4	96.3	95.6	87.9
All-Ann.	99.0	93.2	92.9	85.2	90.5	90.5	93.5	92.4	92.1	76.9

Table 4: Inter-Annotator agreement for the refined annotation experiment

5.5 Refined Annotation Experiment

In order to mitigate the sources of confusion in the original guidelines, we come-up with event-based guidelines where we clarify for each event whether or not the annotators should draw correlations between different entities. This is needed in order to rely less on each annotator's political leaning and more on the presented set of rules. Additionally, we ask annotators to identify the priority expressed by the comment and change the questions and answer choices as follows:

- Q1: Does the given *comment* discuss Egyptian politics? (Yes/No)

- Q2: Is there enough context to identify the political leaning of the *comment*? (Yes/No)

- Q3: Does the *comment* report the opinion of another person/entity and not the opinion of the author of the comment? (Yes/No/None)

- Q4: Which of the following do you think is the top priority for the comment: (1) Supporting Jan. 25th Revolution; (2) Stability; (3) Supporting Mubarak's Regime; (4) Supporting the Military; (5) Supporting Islamists; (6) Opposing Islamists; (7) Cannot determine the priority; (8) None.

- Q5: What is the *comment*'s stance on Jan. 25th Revolution? (Support/Oppose/None)

- Q6: What is the *comment*'s stance on Mubarak and his regime? (Support/Oppose/None)

- Q7: What is the *comment*'s stance on the Military leaders during the period the *comment* was posted in? (Support/Oppose/None)

- Q8: What is the *comment*'s stance on Islamists? (Support/Oppose/Mixed/None)

We split the comments according to the event they discuss and present the annotators with 10 sub-tasks for each one of the 10 events. Additionally we clarify the following in the refined guidelines:

- When choosing "No" as an answer to Q1 or Q2, choose "None" for Q3-Q8;

- For Q4, choose "Can't determine the priority" when there is more than one priority in the *comment* and you cannot choose between them;

- For Q5-Q8, choose "None" if you cannot determine the leaning of the comment toward the entity in question;

- For all questions if the *comment* expresses an opinion toward Jan. 25th Revolution or Mubarak's regime but not both of them, in most cases you can assume that supporting Jan. 25th Revolution implies opposing Mubarak's regime and vice versa;

- If a *comment* reports a opinion of another person/entity without opposing it, indicate in Q3 that it is a reported opinion then assume for all other questions that the reported opinion expresses the opinion of the author of the *comment*.

- For event 6:
  - Opposing the OGR candidate Ahmed Shafik does not imply supporting the Islamist candidate Mohamed Morsi while supporting Ahmed Shafik implies opposing Mohamed Morsi.
  - Similarly, opposing Mohamed Morsi does not imply supporting Ahmed Shafik while supporting Mohamed Morsi implies opposing Ahmed Shafik.

- For events 9 and 10, if a *comment* expresses an opinion toward the Military or Islamists (not both of them), in most cases you can assume that supporting Islamists implies opposing the Military and vice versa.

	Pro-RFM Pages		Pro-OGR Pages	
	Yes	No	Yes	No
Q1. Egy Politics	97.7	2.3	97.9	2.1
Q2. Context	85.9	14.1	87.7	12.3
	Yes	No	None	Yes	No	None
Q3. Rep. Opinion	2.3	83.6	14.1	0.4	87.3	12.3

	Support	Oppose	None	Support	Oppose	None
Q5. Mubarak	46	3.6	50.4	12.2	44.2	43.6
Q6. Army	3.6	45.9	50.5	44	12.4	43.6
Q7. Islamists	23.1	9.7	67.3	25	5.2	69.8

	Support	Oppose	Mixed	None	Support	Oppose	Mixed	None
Q8. Islamists	29.4	12.9	57.3	0.3	12	37.7	0	50.3

Table 5: Answer distribution (averaged over all annotators) to questions Q1-3 and Q5-Q8 in the refined annotation experiment.

	Pro-RFM	Pro-OGR
Jan. 25th Revolution	33.5	3.3
Support Mubarak	0.6	31.5
Support Stability	9.4	7.8
Support Army	1.1	6.8
Support Islamists	28.5	11.5
Oppose Islamists	11.7	26.5
Can't Tell	1	0.4
None	14.1	12.3

Table 6: Answer distribution (averaged over all annotators) to Q4 (Identify the priority of the comment)

It is worth mentioning that for Q4 except for opposing Islamists, we only address what a comment supports (not opposes). We did an exercise where we annotated 400 comments ourselves and found that for many comments the most central element to the belief systems of the authors is whether or not Islam/religion should be involved in politics. A person who supports RFM might temporarily support OGR if it guarantees ousting Islamists from the political scene and vice versa. Moreover for all other aspects (Jan. 25th Revolution, Mubarak, Army, etc.) one can infer what a person opposes based on what this person supports and the event that is being commented on.

5.6 Results of Refined Annotation

Table 4 shows the IAA for the second annotation experiment. As expected, Q4 has a lower IAA than all other questions. Overall the new task yields a much higher agreement. The complete row agreement (Row) jumps from 25.5% to 76.9%

and the average question agreement jumps from 75.7% to 92% comparing the pilot annotations to the refined annotations. Tables 5 and 6 show the distribution to all answers in the second annotation experiment. While the distribution of answers to Q1 almost remained the same, the distribution of Q2 changed. We attribute this to our emphasis on what constitutes enough context in the modified guidelines.

6 Conclusion

In this work we explain our process for collecting and annotating a dataset of social media commentaries discussing Egyptian politics. We propose a taxonomy of major Egyptian Ideological Perspectives, develop annotation guidelines and conduct a pilot experiment to collect annotations that try to uncover the underlying dimensions of the perspective from which a given comment was written. We refine the annotation task and the guidelines based on feedback collected from the annotators. In the refined task, in addition to asking about the comment's position on different ideological aspects such as Jan. 25th Revolution, the Military, Islamists, etc. we ask them to identify the priority expressed by the comment. Additionally to address the challenge of when they should imply a comment's stance on one political entity (ex. the Military) based on its stance toward another entity (ex. Islamists), we develop a set of event-based rules for these associations. IAA between all four annotators for the refined task ranges from 99% to 85.2% for the different questions. We pay close attention to annotator bias. We design the

second set of guidelines in such a way to circumvent the role of annotator subjectivity, decoupling the "why" from the "what" in annotation. We plan on further refinement of the proposed guidelines to alleviate the points of confusion among the annotators. Moreover we plan on collecting more annotations from other informal genres testing the robustness of our annotation framework.

Acknowledgments

References

Amjad Abu-Jbara, Mona Diab, Pradeep Dasigi, and Dragomir Radev. 2012. Subgroup detection in ideological discussions. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics.

Amjad Abu-Jbara, Ben King, Mona T Diab, and Dragomir R Radev. 2013. Identifying opinion subgroups in arabic online discussions. In *ACL (2)*.

Javier Borge-Holthoefer, Walid Magdy, Kareem Darwish, and Ingmar Weber. 2015. Content and network dynamics behind egyptian political polarization on twitter. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 700–711. ACM.

Philip E Converse. 2006. The nature of belief systems in mass publics (1964). *Critical Review*, 18(1-3).

Heba Elfardy, Mona Diab, and Chris Callison-Burch. 2015. Ideological perspective detection using semantic features. *Lexical and Computational Semantics (* SEM 2015)*.

Robert M Entman. 1993. Framing: Toward clarification of a fractured paradigm. volume 43. Wiley Online Library.

Stephan Greene and Philip Resnik. 2009. More than words: Syntactic packaging and implicit sentiment. In *Proceedings of human language technologies: The 2009 annual conference of the north american chapter of the association for computational linguistics*, pages 503–511. Association for Computational Linguistics.

Kazi Saidul Hasan and Vincent NG. 2012. Predicting stance in ideological debate with rich linguistic knowledge. In *Proceedings of the 24th International Conference on Computational Linguistics*.

Kazi Saidul Hasan and Vincent Ng. 2013. Extralinguistic constraints on stance recognition in ideological debates. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics.

Wei-Hao Lin, Theresa Wilson, Janyce Wiebe, and Alexander Hauptmann. 2006. Which side are you on?: identifying perspectives at the document and sentence levels. In *Proceedings of the Tenth Conference on Computational Natural Language Learning*. Association for Computational Linguistics.

Alexandra Siegel. 2014. Tweeting beyond tahrir: Ideological diversity and political tolerance in egyptian twitter networks.

Swapna Somasundaran and Janyce Wiebe. 2010. Recognizing stances in ideological online debates. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*. Association for Computational Linguistics.

Tae Yano, Philip Resnik, and Noah A Smith. 2010. Shedding (a thousand points of) light on biased language. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*. Association for Computational Linguistics.