# An Arabic Speech-Act and Sentiment Corpus of Tweets

# ArSAS: An Arabic Speech-Act and Sentiment Corpus of Tweets

**AbdelRahim A. Elmadany**[1], **Hamdy Mubarak**[2], **Walid Magdy**[3]
[1]Jazan University, Saudi Arabia
[2]Qatar Computing Research Institute, HBKU, Qatar
[3]School of Informatics, Edinburgh University, UK
aelmadany@jazanu.edu.sa, hmubarak@hbku.edu.qa, wmagdy@inf.ed.ac.uk

**Abstract**

Speech acts are the type of communicative acts within a conversation. Speech act recognition (aka classification) has been an active research in recent years. However, much less attention was directed towards this task in Arabic due to the lack of resources for training an Arabic speech-act classifier. In this paper we present ArSAS , an Arabic corpus of tweets annotated for the tasks of speech-act recognition and sentiment analysis. A large set of 21k Arabic tweets covering multiple topics were collected, prepared and annotated for six different classes of speech-act labels, such as expression, assertion, and question. In addition, the same set of tweets were also annotated with four classes of sentiment. We aim to have this corpus promoting the research in both speech-act recognition and sentiment analysis tasks for Arabic language.

## 1. Introduction

Understanding user's Speech act within conversations is an important research task of Natural Language Understanding (NLU). Speech act task could be defined as identifying the intention of a speaker in producing a particular utterance of few words (e.g. sentence in a conversation or a tweet) (Webb, 2010), where the intention can be expression of a feeling, asking question, recommending something ... etc. Speech act recognition is becoming an essential task for many NLU applications such as summarization (McKeown et al., 2007), question answering (Hong and Davison, 2009), and chat-bots (Feng et al., 2006).

Speech act recognition is usually applied to conversations, such as dialogues and chatting platforms, which is mainly called synchronous conversations (Dhillon et al., 2004; Jurafsky, 1997). Other work focused on asynchronous conversations, where the discussion is to open audience and sequence of conversation is not fully synchronized; e.g. forums and social media (Tavafi et al., 2013; Oya and Carenini, 2014; Vosoughi and Roy, 2016b). Recently, social media platforms, such as Twitter, have become a major mean of communication between users. Consequently, understanding the speech act of user posts on these platforms became of increasing importance (Vosoughi and Roy, 2016b). The main objective of classifying speech act of a social post goes beyond the literal meaning of text, where it considers how the context and intention contribute to the meaning of the post (Vosoughi and Roy, 2016b). Most of the work on this task focused on English, while almost no attention was directed towards highly inflected languages such as Arabic.

In this paper, we present ArSAS, a manually annotated **Ar**abic **S**peech **A**ct and **S**entiment corpus of tweets. To our knowledge, ArSAS is considered the first corpus of Arabic speech act on Twitter. The corpus consists of a set of more than 21k Arabic tweets that are manually annotated for six differences classes of speech-act {Assertion, Expression, Recommendation, Respect, Question and Misc}. In addition, tweets are also annotated for four different categories of sentiment {Positive, Negative, Neutral and Mixed}, which is considered the largest Arabic tweets corpus that is labeled for four categories of sentiment. ArSAS dataset is publicly available for free for research purposes[1].

## 2. Background

### 2.1. Speech Act

One of the most essential steps in human-computer interaction systems, aka dialogue Systems, is understanding user's need. This process is called "language understating component", "Dialogue Acts" or "Speech Acts". Speech act recognition (also know as classification) task is labeling the speaker's intention in producing a particular utterance. The speech act terminology is approximately the equivalent of the speech act of (Searle, 1969), where it was presented as a fundamental concept of linguistic pragmatics analyzing; for example, what it means to ask a question or make a statement. Although major dialogue theories treat dialogue acts as a central notion, the conceptual granularity of the used speech act labels/classes varies considerably among alternative analyses, depending on the application or domain (Webb and Hardy, 2005). Within the field of computational linguistics, recent work closely linked to the development and deployment of spoken language dialogue systems has focused on some of the conversational roles such acts can perform. Most of the previous research on speech act is widely used with data transcribed from telephone or face-to-face conversations, which is also known as synchronous conversations (Dhillon et al., 2004; Jurafsky, 1997).

Recently, social media platforms, such as Twitter, became a hub for users to communicate and discuss various topics. These communicative acts among social media users are seen as a kind of asynchronous conversations, which can include spreading news, asking questions, or expressing feelings, which all fall under the scope of "speech act". Classifying speech act of social media posts can provide a new dimension to study social media content as well as providing real-life data to validate or reject claims in the

[1]http://homepages.inf.ed.ac.uk/wmagdy/resources.htm

speech act theory (Zhang et al., 2012). Speech act classification of tweets is considered fairly new task. Recent work mostly focused on classifying speech act of tweets on trending topics. (Zhang et al., 2012; Zhang et al., 2011) proposed a speech act classification method to understand Twitter users' behavior through a set of word-based and character-based features. (Nemer, 2015) proposed a system for understanding celebrity (e.g. Oprah Winfrey and Britney Spears) speech act on Twitter. They investigated celebrities' speech patterns on Twitter and whether they mostly talk to fans, and how they communicate with different audiences. (Vosoughi and Roy, 2015; Vosoughi and Roy, 2016a) proposed two systems based on assertion speech act detection. The assertion act is an utterance that commits the speaker to the truth of the expressed proposition. For example, the tweet "*there is a third bomber on the roof*" contains an assertion, while the tweet "*I hate reporters!*" contains an expression. They proved that assertion is important to identify rumors and track stories about real-world events. Moreover, they showed that more than half of tweets about events do not contain assertions. (Vosoughi and Roy, 2016b) proposed automatic speech act classifier for tweets based on semantics features such as opinion and vulgar words, emotions, speech act verb, n-grams, syntactic features, Twitter-specific Characters (# and @ ), abbreviations, and dependency Sub-trees. They examined four classifiers: Naive Bayes (NB), decision trees, logistic regression, and SVMs.
All previous work mainly focused on speech act classification for English, while very limited work targeted speech act classification for Arabic.

## 2.2. Arabic Speech Act

To the best of our knowledge, there are two available corpora for Arabic speech act on synchronous conversations. First, TuDiCoI (Tunisian Dialect Corpus Interlocutor) consists of Railway information from the National Railway Company in Tunisia (SNCFT) which transcribed spoken Arabic dialogues and contains 12,182 utterances (Graja et al., 2013). Second, JANA corpus which is a multi-genre corpus of Arabic dialogues labeled for Arabic Dialogues Language Understanding (ADLU) at utterance level and comprising Spontaneous Speech Dialogues (SSD) and Instance Messages (IM) for Egyptian dialect. It contains 4,725 utterances and it is published on LDC (LDC2016T24) (Elmadany et al., 2016).
These two datasets were investigated in few research studies. (Graja et al., 2013) used the TuDiCoI corpus to develop a discriminative algorithm based on conditional random fields (CRF) to semantically label spoken Tunisian dialect turns which are not segmented into utterances. (Elmadany, 2016; Elmadany et al., 2018) utilized the JANA corpus to create a statistical dialogue analysis model for recognizing utterance's dialogue acts using a machine learning approach based on multi-classes hierarchical structure.
In addition, there are other few initiatives that studied Arabic speech acts classification, but on a much smaller scale using hand-crafted small dataset. (Shala et al., 2010) applied speech act classification for Arabic discourse using NB and decision trees classifiers on a dataset of about 400 utterances only collected from newspapers. (Bahou et al., 2008) proposed a method for the semantic representations of utterances of spontaneous Arabic speech based on the frame grammar formalism and tested on about 1,000 Tunisian national railway queries collected using Wizard-of-Oz technology. Another work (Lhioui et al., 2013) used the same Wizard-of-Oz technology but to collect a smaller set of 140 utterances only recorded from 10 speakers.
Previous work shows the huge limitation in the availability of annotated Arabic data for the task of speech act recognition. We believe that ArSAS would be the first stranded corpus for Arabic speech act classification for asynchronous conversations, which contains over 21k tweets labeled with fine-grained set of six different speech act classes.

## 2.3. Arabic Sentiment Analysis

Unlike speech act, there was some attention to Arabic sentiment analysis including few initiative to create standard corpora and lexicons for this task.
Early work on Arabic sentiment analysis focused on Modern Standard Arabic (MSA) (Abbasi et al., 2008; Abdul-Mageed et al., 2011). Later on, many initiatives started to focus on dialectal Arabic on social media (Mourad and Darwish, 2013; Abdul-Mageed et al., 2014). One of the initial work on sentiment analysis for Arabic tweets was presented by (Mourad and Darwish, 2013). They proposed expandable ArabSinti lexicon for both Modern Standard Arabic (MSA) news articles and dialectal Arabic tweets. They used 2,300 Arabic tweets annotated with five possible labels: neutral, positive, negative, both, or sarcastic. Another work by (Badaro et al., 2015) introduced a large-scale Standard Arabic sentiment lexicon (ArSenL) developed using a combination of English SentiWordnet (ESWN), Arabic WordNet, and the Arabic Morphological Analyzer (AraMorph). They developed a set of 28,760 words, but mainly in MSA. (Ibrahim et al., 2015) proposed a corpus of MSA and Egyptian dialect. The corpus is extracted from tweets, comments on hotel reservations and TV programs and product reviews annotated at the sentence level. It consists of 2,154 positive, 1,648 negative and 1,98 neutral texts. (Refaee and Rieser, 2014) proposed a corpus of Arabic tweets annotated for subjectivity and sentiment analysis consists of 6,894 tweets and annotated with four sentiment labels: positive, negative, neutral and mixed. More recent work in SemEval 2016 on a sentiment analysis task for multiple languages including Arabic (Kiritchenko et al., 2016) introduced a small dataset of 1,366 tweets. Another SemEval task for sentiment analysis for Arabic tweets was introduced in 2017 with a larger set of 9,455 Arabic tweets annotated with 3 labels: positive, negative, and neutral (Rosenthal et al., 2017). Another available copora on Arabic sentiment analysis was introduced by (Nabil et al., 2015), where they introduced the Arabic Sentiment Tweets Dataset (ASTD) which contains 10k Egyptian tweets annotated with four sentiment labels. Finally, (Al-Twairesh et al., 2017) developed a larger corpus that consists of 17k annotated tweets with the same four sentiment labels.
Our ArSAS corpus should be the current largest corpus for Arabic sentiment analysis with over 21k annotated tweets annotated with 4 different labels of sentiment.

| Type | Arabic Topic | Translation |
|---|---|---|
| Events | الانتخابات الرئاسية المصرية | Egyptian presidential election |
| | تفجيرات سيناء | Sinai bombings |
| | حادث الواحات | Wahat attack |
| | تصفيات كأس العالم | World cup qualifications |
| | محاربة الفساد في السعودية | Fighting corruption in KSA |
| | منتدى شباب العالم | World youth forum |
| | اعتقال سلمان العودة | Arresting Salman Aloda |
| Entities | عبد الفتاح السيسي | Abdelfattah Al-Sisi |
| | المحامي خالد علي | Khaled Ali |
| | حمدين صباحي | Hamdeen Sabahi |
| | محمد صلاح | Mohamed Salah |
| | مرتضي منصور | Mortada Mansour |
| | ولي العهد السعودي | KSA crown prince |
| | الحوثيون | Houthis |
| | خليفة حفتر | Khalifa Haftar |
| Long Standing | الربيع العربي | Arab spring |
| | تيران وصنافير | Tiran and Sanafir |
| | ثورة يناير | January Revolution |
| | أزمة الخليج | Gulf crisis |
| | حصار ومقاطعة قطر | Qatar siege and boycott |

Table 1: List of topics used to collect tweets

# 3. Corpus Creation

## 3.1. Data Collection

We used the Twitter API[2] to collect tweets on a set of topics we developed. We used (Zhao and Jiang, 2011; Vosoughi and Roy, 2016b) definitions for three different types of topic. A topic is an essay or article which discussed in one or more tweets. A type is the characteristic of topics, and is classified into:

- Long-Standing: Topics about articles that are commonly discussed over long period of time.
- *Entity*: Topics about celebrities or organizations.
- Event: Topics about an important thing that is happening.

We created a set of 20 topics of the three types above which covers controversial topics that potentially get discussion on social media, which would be highly suitable for both the tasks of speech act recognition and sentiment analysis. Table 1 shows the list of topics we developed and used to collect the tweets.

We collected a set of 62,690 tweets in the period 1-15 November 2017. We applied some data filtering by removing short tweets that contain fewer than three words (without counting hashtags, user mentions, and URLs). Then we randomly selected a set of 21,064 tweets for annotation, where 6151, 6146, 8767 tweets were covering the long-standing, entity, and event topics respectively.

## 3.2. Labels Schema

Each tweet in our collection prepared for annotation with two labels for speech act and sentiment. We used a list of six speech act tags based on Searleś speech act taxonomies (Searle, 1969; Searle, 1975) as follows:

[2]https://dev.twitter.com/

| Arabic Tweet | Speech act | Sentiment |
|---|---|---|
| الرياضة تنتخب - شريف عبد القادر : الزمالك في أزمة ولا أحد ينكر طفرة مرتضى منصور داخل النادي | Assertion | Mixed |
| الكرة الايطالية تحتاج لتركي ال الشيخ | Recommendation | Neutral |
| الربيع العربي رغم الالتفاف عليه الا أنه إشعاع من الحرية | Expression | Positive |
| ليه محمد صلاح؟ ☺☺ | Question | Positive |
| اطالب الرئيس ببناء مدينة لها سور عالى فى اقاصى الصحراء وليتم حجز كل المعاتيه الذين امنو بما اسموها ثورة يناير | Request | Negative |
| سأرفع الحد الأدنى للأجور إلى 2000 جنيه وسأنفذ حكم تيران وصنافير - وتعهد أيضا بالإفراج عن المحبوسين | Miscellaneous | Neutral |

Table 2: Samples of annotated tweets with speech act and sentiment

1. Assertion: user declares some proposition such as stating, claiming, reporting, or announcing.
2. Recommendation: user recommends something.
3. Expression: user expresses some psychological state such as thanking, apologizing, or congratulating.
4. Question: user asks a question such as why, what, or confirmation.
5. Request: user asks for something such as ordering, requesting, demanding, or begging.
6. Miscellaneous: user committed to some future action such as promising or offering.

For sentiment labels, we used the standard four sentiment tags: positive, negative, mixed (contains both positive and negative sentiment), or neutral (no opinion or sentiment disclosed).

Table 2 shows illustrative examples of each of the speech act and sentiment tags.

## 3.3. Data Annotation

For tweets annotation, we created a job on CrowdFlower crowdsourcing platform where we showed tweets to annotators and asked them to classify speech act and sentiment for each tweet into one of the above-described tags. Guidelines and examples of tweets for each tag were presented to annotators for better understanding. We restricted annotation to workers from all Arab countries who have "Arabic" language in their profile. Each tweet was judged by at least three annotators.

Quality of annotation was controlled by utilizing 70 hidden test questions; each has the correct answer(s) for both speech act and sentiment. These test questions were selected such that their answers, as selected by two language experts, are matched. Annotators on CrowdFlower were required to get at least 70% of the hidden test questions correctly to continue. Otherwise, they get excluded from the job and their work gets discarded. Around 500 annotators participated successfully in the annotation process which gives the diversity of opinions needed for such tasks.

Agreement among annotators was 87% for speech act, and 79% for sentiment, which indicates that annotation

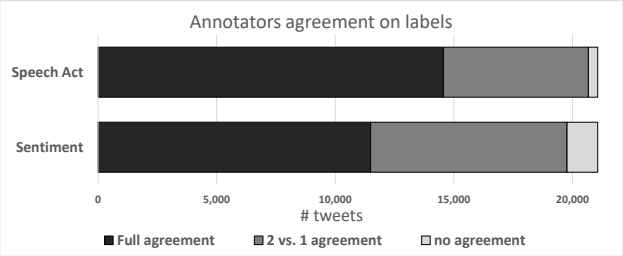


Figure 1: The number of tweets that have full, partial, or no agreement for each of the speech act and sentiment labels

of speech act might be simpler and more straightforward than the arguable sentiment. However, both inter-annotator agreements are considered high, especially for an annotation task with three annotators and 4+ choices.

Crowdflower provides a confidence score with each annotated tweet that represents the confidence in the quality of the label. For a three annotators per tweet setup, the confidence score would range between 0.3 and 1 according to two factors: 1) annotator quality level; and 2) agreement among annotators. A confidence score of 1 means that the three annotators selected the same label. A score around 0.6 refers to two annotators agreeing on a given label while the third selecting another one. A score around 0.3 refers to having the three annotators selecting three different labels, and in this case the label selected by the annotator with the highest quality value is considered. Figure 1 shows the number of tweets that have full, partial, or no agreement for each of the speech act and sentiment labels. As shown, the majority of the tweets got full agreement for both tasks (69% and 55% for speech act and sentiment labels respectively). The number of tweets that received different label from each annotator is very small (2% and 6% of all tweets for speech act and sentiment labels respectively). We could exclude these tweets, however we preferred to keep in our collection as an example of challenging tweets.

## 4. ArSAS Corpus Characteristics

ArSASreleased dataset contains the following information:

- `ID`: ID of the tweet.
- `Text`: the original unprocessed text of the tweet
- `topic`: topic type of the keyword used to collect the tweet.
- `Sentiment`: selected sentiment label.
- `Sentiment Conf.`: Confidence score of sentiment label.
- `Speech Act`: selected speech act label.
- `Speech Act Conf.`: Confidence score of the speech act label.

Figure 2(a) shows the distribution of speech act labels in our ArSAS corpus after annotation. As shown, the majority of the tweets are labeled as either Expression (56%) or Assertion (39%), and the remaining labels are used in only 5% of the cases. This highly unbalanced distribution is similar to the English tweets corpus used in (Vosoughi and Roy, 2016b).

Figure 2(b) shows the distribution of sentiment labels. Tweets having negative sentiment represent one third of the tweets, while those with positive sentiment represent around quarter of the tweets, and one third of the tweets have no (neutral) sentiment. Only 6% of the tweets have mixed sentiment.

Table 3 shows the fine-grained distribution of speech act tags by topics type and sentiment tags. As can be seen in the table, the majority of the Assertion tweets are coming from the 'Events' topics, while the tweets with other speech acts have less bias towards the topics. Also, it can be noticed that the largest two speech act tags, Assertion and Expression, have very different distributions for the sentiment, where most of the Assertion tweets have no sentiment (neutral), while most of the tweets with Expression speech act have polarized sentiment, most of them are negative. These observations show the value of having a corpus labeled for both speech act and sentiment, since one of the two tasks can be used as an effective feature to predict the other.

## 5. Conclusion

In this paper, we introduce ArSAS, a large dataset of Arabic tweets annotated for both speech acts and sentiment. ArSAS consists of 21k Arabic tweets written in multiple Arabic dialects as observed by examining different samples. The tweets in the corpus were extracted and collected using 20 controversial topics in different countries that are expected to have hot discussions among Twitter users. The tweets collection did not rely on emotions or sentiment keywords to avoid data bias to a given lexicon, especially for the task of sentiment analysis. The corpus is annotated with six speech act labels and four sentiment labels. The annotation process was applied using a crowdsourcing platform by having at least three annotators labeling each tweet. An inter-annotator agreement of 87% and 79% was achieved for the speech act and sentiment labels respectively. To the best of our knowledge, ArSAS is the largest annotated corpus of speech act and sentiment in Arabic. In addition, it is considered the first Arabic corpus annotated for the speech act recognition in tweets. We hope that our corpus would bring the attention to the speech act recognition task for Arabic and further promote the research in Arabic sentiment analysis. Moreover, it can be applied for applications that combines both tasks.

ArSAS corpus is freely available online as an open-source for researchers interested in Arabic speech act and sentiment analysis and could be downloaded from `http://homepages.inf.ed.ac.uk/wmagdy/resources.htm`.

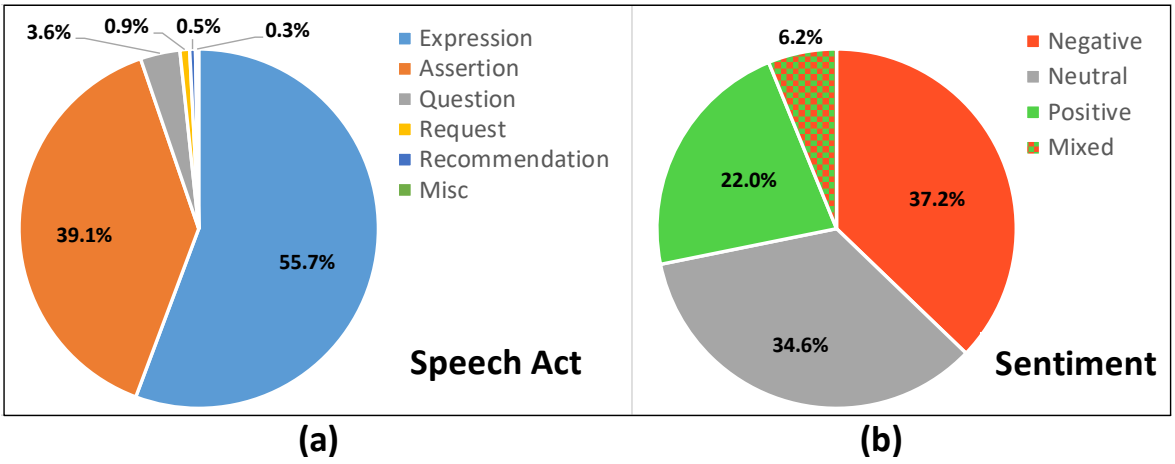


Figure 2: The distribution of the speech act tags (a) and sentiment tags (b) in the ArSAS corpus

| | Topics Types | | | Sentiment Analysis tags | | | |
|---|---|---|---|---|---|---|---|
| **Speech Act (# tweets)** | **Long-standing** | **Entities** | **Events** | **Neutral** | **Positive** | **Negative** | **Mixed** |
| Assertion (8,221) | 627 | 2,097 | 5,497 | 6,666 | 962 | 488 | 105 |
| Recommendation (107) | 18 | 55 | 34 | 23 | 36 | 36 | 12 |
| Question (751) | 327 | 219 | 205 | 248 | 47 | 403 | 53 |
| Request (180) | 23 | 94 | 63 | 30 | 63 | 66 | 21 |
| Expression (11,745) | 5,126 | 3,658 | 2,961 | 289 | 3,514 | 6,835 | 1,107 |
| Miscellaneous (60) | 30 | 23 | 7 | 23 | 21 | 12 | 4 |

Table 3: The distribution of speech act tags via topics type and sentiment analysis tags

## 6. References


Abbasi, A., Chen, H., and Salem, A. (2008). Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. *ACM Transactions on Information Systems (TOIS)*, 26(3):12.

Abdul-Mageed, M., Diab, M. T., and Korayem, M. (2011). Subjectivity and sentiment analysis of modern standard arabic. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 587–591. Association for Computational Linguistics.

Abdul-Mageed, M., Diab, M., and Kübler, S. (2014). Samar: Subjectivity and sentiment analysis for arabic social media. *Computer Speech & Language*, 28(1):20–37.

Al-Twairesh, N., Al-Khalifa, H., Al-Salman, A., and Al-Ohali, Y. (2017). Arasenti-tweet: A corpus for arabic sentiment analysis of saudi tweets. *Procedia Computer Science*, 117:63–72.

Badaro, G., Baly, R., Akel, R., Fayad, L., Khairallah, J., Hajj, H., Shaban, K., and El-Hajj, W. (2015). A light lexicon-based mobile application for sentiment mining of arabic tweets. In *Proceedings of the Second Workshop on Arabic Natural Language Processing*, pages 18–25.

Bahou, Y., Belguith, L. H., and Hamadou, A. B. (2008). Towards a human-machine spoken dialogue in arabic. In *6th Language Resources and Evaluation Conference (LREC 2008), Workshop HLT Within the Arabic World. Arabic Language and Local Languages Processing Status Updates and Prospects, Marrakech, Morocco.*

Dhillon, R., Bhagat, S., Carvey, H., and Shriberg, E. (2004). Meeting recorder project: Dialog act labeling guide. Technical report, INTERNATIONAL COMPUTER SCIENCE INST BERKELEY CA.

Elmadany, A., Abdou, S., and Gheith, M. (2016). Jana: A human-human dialogues corpus for egyptian dialect (ldc2016t24).

Elmadany, A., Abdou, S., and Gheith, M. (2018). Improving dialogue act classification for spontaneous arabic speech and instant messages at utterance level. In *11th edition of the Language Resources and Evaluation Conference*.

Elmadany, A. (2016). *Automatic Act Classification for Arabic Dialogue Context*. Thesis.

Feng, D., Shaw, E., Kim, J., and Hovy, E. (2006). An intelligent discussion-bot for answering student queries in threaded discussions. In *Proceedings of the 11th international conference on Intelligent user interfaces*, pages 171–177. ACM.

Graja, M., Jaoua, M., and Belguith, L. H. (2013). Discriminative framework for spoken tunisian dialect understanding. In *International Conference on Statistical Language and Speech Processing*, pages 102–110. Springer.

Hong, L. and Davison, B. D. (2009). A classification-based approach to question answering in discussion boards. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 171–178. ACM.

Ibrahim, H. S., Abdou, S. M., and Gheith, M. (2015). Mika: A tagged corpus for modern standard arabic and colloquial sentiment analysis. In *Recent Trends in Infor-*

*mation Systems (ReTIS), 2015 IEEE 2nd International Conference on*, pages 353–358. IEEE.

Jurafsky, D. (1997). Switchboard swbd-damsl shallow-discourse-function annotation coders manual. *www. dcs. shef. ac. uk/nlp/amities/files/bib/ics-tr-97-02. pdf*.

Kiritchenko, S., Mohammad, S., and Salameh, M. (2016). Semeval-2016 task 7: Determining sentiment intensity of english and arabic phrases. In *SemEval@ NAACL-HLT*, pages 42–51.

Lhioui, C., Zouaghi, A., and Zrigui, M. (2013). A combined method based on stochastic and linguistic paradigm for the understanding of arabic spontaneous utterances. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 549–558. Springer.

McKeown, K., Shrestha, L., and Rambow, O. (2007). Using question-answer pairs in extractive summarization of email conversations. *Computational Linguistics and Intelligent Text Processing*, pages 542–550.

Mourad, A. and Darwish, K. (2013). Subjectivity and sentiment analysis of modern standard arabic and arabic microblogs. In *WASSA@ NAACL-HLT*, pages 55–64.

Nabil, M., Aly, M. A., and Atiya, A. F. (2015). Astd: Arabic sentiment tweets dataset. In *EMNLP*, pages 2515–2519.

Nemer, D. (2015). Celebrities acting up: A speech act analysis in tweets of famous people. *Social Networking*, 5(01):1.

Oya, T. and Carenini, G. (2014). Extractive summarization and dialogue act modeling on email threads: An integrated probabilistic approach. In *SIGDIAL Conference*, pages 133–140.

Refaee, E. and Rieser, V. (2014). An arabic twitter corpus for subjectivity and sentiment analysis. In *LREC*, pages 2268–2273.

Rosenthal, S., Farra, N., and Nakov, P. (2017). Semeval-2017 task 4: Sentiment analysis in twitter. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 502–518.

Searle, J. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.

Searle, J. (1975). A taxonomy of illocutionary acts. *Language, Mind, and Knowledge: Minneapolis Studies in the Philosophy of Science*, 7:344–369.

Shala, L., Rus, V., and Graesser, A. C. (2010). Automated speech act classification in arabic. *Subjetividad y Procesos Cognitivos*, 14:284–292.

Tavafi, M., Mehdad, Y., Joty, S. R., Carenini, G., and Ng, R. T. (2013). Dialogue act recognition in synchronous and asynchronous conversations. In *SIGDIAL Conference*, pages 117–121.

Vosoughi, S. and Roy, D. (2015). A human-machine collaborative system for identifying rumors on twitter. In *Data Mining Workshop (ICDMW), 2015 IEEE International Conference on*, pages 47–50. IEEE.

Vosoughi, S. and Roy, D. (2016a). A semi-automatic method for efficient detection of stories on social media. In *ICWSM*, pages 707–710.

Vosoughi, S. and Roy, D. (2016b). Tweet acts: A speech act classifier for twitter. In *The 10th International Aaai Conference On Web And Social Media (ICWSM-16)*, pages 711–715.

Webb, N. and Hardy, H. (2005). Data-driven language understanding for spoken language dialogue. In *Proceedings of the AAAI Workshop on Spoken Language Understanding*, pages 23–29.

Webb, N. (2010). *Cue-based dialogue act classification*. Ph.D. thesis, University of Sheffield, Department of Computer Science.

Zhang, R., Gao, D., and Li, W. (2011). What are tweeters doing: Recognizing speech acts in twitter. *Analyzing Microtext*, 11:05.

Zhang, R., Gao, D., and Li, W. (2012). Towards scalable speech act recognition in twitter: tackling insufficient training data. In *Proceedings of the Workshop on Semantic Analysis in Social Media*, pages 18–27. Association for Computational Linguistics.

Zhao, X. and Jiang, J. (2011). An empirical comparison of topics in twitter and traditional media. *Singapore Management University School of Information Systems Technical paper series*.