

Interactive Annotation for Event Modality in Modern Standard and Egyptian Arabic Tweets

Rania Al-Sabbagh[†], Roxana Girju[†], Jana Diesner[‡]

[†]Department of Linguistics and Beckman Institute

[‡]School of Library and Information Science

University of Illinois at Urbana-Champaign, USA

{alsabbal, girju, jdiesner}@illinois.edu

Abstract

We present an interactive procedure to annotate a large-scale corpus of Modern Standard and Egyptian Arabic tweets for event modality that comprises obligation, permission, commitment, ability, and volition. The procedure splits up the annotation process into a series of simplified questions, dispenses with the requirement of expert linguistic knowledge, and captures nested modality triggers and their attributes semi-automatically.

1 Introduction

Event modality, according to Palmer (2001), describes events that are not actualized but are merely potential. It comprises obligation, permission, commitment, ability, and volition. Both obligation and permission emanate from an external authority such as the law; whereas commitments are the obligations placed by speakers on themselves as in promises. Ability is the (in)capacity to do something. Volition is broadly defined as intensions, desires, wishes, and preferences. Event modality is used for several NLP tasks, including sales and marketing analysis (Ramanand et al. 2010, Carlos and Yalamanchi 2012), sentiment analysis (Chardon et al. 2013), the automatic detection of request emails (Lampert et al. 2010), and the classification of animacy and writers' emotions (Liao and Liao 2009, Bowman and Chopra 2012).

To-date, there are no large-scale Arabic corpora annotated for event modality compared to English (Baker et al. 2010, 2012; Rubinstein et al. 2013), Japanese (Matsuyoshi et al. 2010), Portuguese (Hendrickx et al. 2012), and Chinese (Cui and Chi 2013). One obstacle for the creation of modality-annotated corpora is the lack of consensus definitions of modality and its attributes to be rendered into annotation tasks and guidelines. Furthermore, most modality annotation schemes use sophisticated theoretical guidelines that need annotators with linguistic background; hence, annotation typically takes place in in-lab settings at small scales.

In this paper, we present an interactive annotation procedure to annotate event modality and its attributes of sense, polarity, intensification, tense, holders, and scopes in Modern Standard and Egyptian Arabic tweets. The procedure depicts the following ideas: first, it defines each annotation task as a series of questions displayed¹/hidden based on prior answers; second, it avoids lengthy theoretically-sophisticated definitions and uses the questions instead as simplified self-explanatory annotation prompts; and third, based on the elicited answers it automatically determines nested triggers and their attributes. The fact that our procedure does not require special linguistic background and consists of easy-to-administer questions makes it eligible for large-scale crowdsourcing annotation.

Our corpus comprises 9949 unique tweets, annotated for 12134 tokens that map to 315 unique types of event modality triggers and their attributes of sense, polarity, intensification, tense, holders, and scopes. The reason to work on the genre of tweets is that our corpus is part of a larger project to incorporate linguistic features, such as modality, with network-based features to automatically identify the key players of political discourse on Twitter for countries with fast-changing politics such as Egypt. The fact that our corpus is harvested from the Arabic Egyptian Twitter entails that the corpus is diglossic for Modern Standard Arabic (MSA), the

This work is licensed under a Creative Commons Attribution 4.0 International License. Page numbers and proceedings footer are added by the organizers. License details: <http://creativecommons.org/licenses/by/4.0/>

formal Arabic variety, and Egyptian Arabic (EA), the native Arabic dialect of Egypt. We evaluate the annotation results with Krippendorff's alpha (Krippendorff 2011). Results show high inter-annotator reliability rates, indicating that our annotation scheme and procedure are effective. The contribution of this paper, therefore, is twofold: first, we create a novel annotated resource for Arabic NLP that is larger than existing corpora even for languages other than Arabic; and second, we present an efficient and easy-to-administer annotation procedure with interactive crowdsourcing potentials.

The rest of this paper is organized as follows: Section 2 outlines the annotation scheme, guidelines and the interactive procedure; Section 3 gives examples for the final output representations; Section 4 describes corpus harvesting and sampling; Section 5 provides the annotation results and disagreement analysis; and Section 6 compares and contrasts our work with related work.

2 Annotation Scheme: Tasks and Guidelines

Our annotation scheme comprises six tasks to label sense, polarity, intensification, tense, holders, and scopes for each event modality. Prior to the beginning of the interactive procedure, we highlight all event modalities in each tweet using a string-match algorithm and the lexicons from Al-Sabbagh et al. (2013, 2014a). The algorithm finds all potential event modality triggers (i.e. words/phrases that convey event modality) within each tweet in our corpus and marks them as annotation units. A total of 12134 candidate triggers are highlighted in 9949 tweets.

2.1 Task 1: Sense

Sense annotation is to decide for each candidate trigger in context whether it actually conveys event modality given the tweet's context. The same present participle *حابب* *HAbb* in example 1 is a volition trigger meaning *I want/desire*; whereas in example 2 it is a non-modal present participle meaning *like/prefer/respect*.

1. ¹طبعاً أنا مش حابب [عمرو موسى يكسب]
*TbEA >nA m\$ **HAbb** [Emrw mwsY yksb]*
Definitely, I do not **want** [Amr Moussa to win].
2. عمرو أديب: رسمياً الكتاتني مش حابب أبو حامد
*Emrw >dyb: rsmYA AlktAtny m\$ **HAbb** >bw HAmd #egypt #qalyoum*
Amr Adeeb: Alkatatny does not officially **like** Abu Hamed #egypt #qalyoum

We define sense annotation as a synonymy judgment task, following Al-Sabbagh et al. (2013, 2014b). Each event modality sense is represented by an exemplar set manually selected so that: (1) each exemplar is an unambiguous event modality trigger; (2) exemplars are in both MSA and EA; (3) exemplars comprise both simple words and multiword expressions; (4) exemplars are both affirmative and negative; and (5) exemplars are of different intensities. Presented with a pre-highlighted candidate trigger in context and the exemplar sets, annotations are to decide whether the candidate trigger is synonymous with any of the exemplar sets. If not, the trigger is then assumed as non-modal.

If an annotator decides that a given candidate trigger is a non-modal, no further questions about polarity, intensification, tense, holders, or scopes are displayed. In order to guarantee that annotators do not select the non-modal option as an easy escape, they are not allowed to move forward without giving at least one synonym of their own to the candidate trigger.

2.2 Task 2: Polarity

Task 2 uses as input the candidates labeled as valid event modality triggers in Task 1 and label each as either affirmative (AFF) or negative (NEG). To decide, annotators are instructed to consider the absence/presence of:

- **Negation particles** such as *مش* *m\$* (not), *لا* *lA* (not), and *غير* *gyr* (not), among others.
- **Negation affixes**, especially in EA, like the circumfix *m...\$* in *مقدرش* *mqdr\$* (I cannot).

¹ Throughout the examples, modality triggers are marked in boldface, and scopes are in-between brackets.

- **Negative polarity items** like عمري *Emry* (never) and لم يعد *lm yEd* (no longer).
- **Negative auxiliaries** where negation is placed on the past tense auxiliary as in مكنتش عايز *mknt\$ EAyz* (I did not want).
- **Inherently-negative triggers** that encode negation in their lexical meanings such as عاجز *Eajz* (incapable) and يمنع *ymnE* (prohibit).
- **Embedding** under negated epistemic modality triggers as in لا أعتقد أنه يجب *lA >Etqd >nh yjb* (I do not think it is necessary) which entails that the speaker is not actually setting an obligation.

Annotators are instructed that using multiple negation markers results in an affirmative sense. Thus, لم يعجز *lm yEjz* (he was not unable to) means that he was actually able to. Annotators are required to give the reason for negation if they decide that a given trigger is negative.

2.3 Task 3: Intensification

Event modality triggers have different lexical intensities (i.e. intensities encoded in the lexical meaning of the word/phrase regardless of the context). In obligation triggers, for instance, even without a context, Arabic speakers know that ضروري *Drwry* (necessary) expresses a higher necessity than المفروض *AlmfrwD* (should). When used in context, the trigger's lexical intensity can be maintained as is, or amplified/mitigated by such linguistic means as:

- **Modification:** adverbs like تماما *tmAmA* (absolutely) amplify lexical intensity; whereas mitigation is invoked by such adverbs as غالبا *gAlbA* (most probably).
- **Categorical negation** typically amplifies lexical intensity as in مش المفروض أبدا *m\$ AlmfrwD >bdA* (it should never be).
- **Emphatic expressions** such as قد *qd* (indeed), والله *wAllh* (I swear), and من كل قلبي *mn kl qlby* (wholeheartedly), among others, lead to lexical intensity amplification.
- **Coordination** of two or more triggers typically results in intensity amplification as in لازم وضروري *lAzm wDrwry* (must and necessary).
- **Embedding** under epistemic modality triggers can affect the lexical intensities of event modality triggers. In أعتقد من الضروري أن *>Etqd mn AlDrwry >n* (I think it is necessary to) the strong obligation associated with الضروري *AlDrwry* (necessary) is mitigated by the moderate-intensity epistemic أعتقد *>Etqd* (I think), being embedded under it.

The annotators' task for intensification annotation is to decide for each candidate labeled as a valid event modality trigger in Task 1 whether its lexical intensity is amplified (AMP), mitigated (MTG) or maintained (AS IS). During interactive annotation, annotators are asked to provide the reason for their selection; that is, whether the lexical intensity is affected by modification, coordination, negation, embedding or any other reason whether listed above or not.

2.4 Task 4: Tense

In this version of our event modality corpus, we work on the present and past tenses only. Thus, Task 4 is to decide for each valid event modality trigger from Task 1 whether it is present (PRS) or past (PST). Annotators are required to give their reasons for selecting either PRS or PST.

2.5 Task 5: Holders

Holder annotation identifies the source of the obligation, permission, commitment, ability, or volition. In example 3, the source that sets the obligation that Egyptians have to learn the meaning of democracy is the Twitter user.

3. لازم [المصريين يتعلموا يعني إيه ديموقراطية الأول] *lAzm [AlmSryyn ytElmwA yEny <yh dymwqrATyp Al>wl]*
[Egyptians **have** to learn what democracy is first]

The holder is not always the Twitter user, however. In example 4, the Twitter user quotes Kamal Alganoury - a former Egyptian Prime Minister - stating that he does not want to

not want Shafiq to win the presidential elections. The trigger *اتمّنت* *Atmnyt* (wished) is tagged as synonymous with the volition exemplar set; therefore, it denotes a DESIRE. It is then labeled as a past tense (PST), negative (NEG) trigger. Furthermore, its lexical intensity is labeled as amplified (AMP) because of the categorical negation *ما عمري* *Emry mA* (never ever). Originally, *اتمّنت* *Atmnyt* (wished) is of moderate lexical intensity, being less intense than *اشتّهيت* *A\$thyt* (longed for) but more intense than *أردت* *>rdt* (wanted). Given the categorical negation, the lexical intensity of *اتمّنت* *Atmnyt* (wished) goes up the scale from moderate to strong (STRG).

10. *عمري ماتمّنت ان [شفيق يكسب]. الحمد لله ربنا محرمنيش من حاجة #مّرسي*
Emry mAtmnyt An [\$fyq yksb]. AlHmd Allh rbnA mHrmny\$ mn HAjp #mrsy
 I have never ever **wished** [Shafiq to win]. Thank God! #Morsi.
rep. USER, STRG PST NEG DESIRE (*\$fyq yksb*)

Example 11 reads as: the Twitter USER reports Hegazy stating that he has the ability to become the Muslim's caliphate. The trigger *أصلح* *>SIH* (can) is labeled as synonymous with the ability exemplar set. It is also labeled as a present (PRS), affirmative (AFF) trigger whose lexical intensity is maintained (AS IS) in the context. Therefore, its lexical intensity is maintained to its original level which is moderate (MOD).

11. *حجازي: أنا أصلح أن [أكون خليفة للمسلمين] وسنكون سادة العالم #Ikhwan*
HjAzy: >nA >SIH >n [>kwn xlyfp llmslmyn] wsnkwn sAdp AIEAlm #Ikhwan
 Hegazy: I **can** [be the Muslims' caliphate] and we will become the world's masters. #Ikhwan
rep. USER, report, (*HjAzy*, MOD PRS AFF ABLE, (*>kwn xlyfp llmslmyn*))

Example 12 shows a Twitter user who holds as true that the only thing Egypt needed was a wise politician to avoid the bloodshed. The trigger *تحتاج* *tHtAj* (needs) is labeled as an obligation trigger synonymous with *تتطلب* *tTlb* (requires). It is also labeled as past tense (PST) given the preceding past tense auxiliary *تكن* *tkn* (was). The assigned strong (STRG) lexical intensity label is attributed to the fact that the original moderate intensity of *تحتاج* *tHtAj* (needs) is amplified by the categorical negation structure *لا ... لم ... إلا* *la ... <lm ... <la* (nothing but).

12. *#مصر لم تكن تحتاج الا [رجل عاقل يخرج من الازمات بدون اراقة الدماء]*
#mSr lm tkn tHtAj ALA [rjl EAql yxrp mn ALAzmat bdwn ArAqp AldmA']
 #Egypt **needed** nothing but [a rational politician who solves crises without bloodshed]
rep. USER, true, (*mSr*, STRG PST AFF REQUIRE (*rjl EAql yxrp mn ALAzmat bdwn ArAqp AldmA'*))

Example 13 illustrates the representation of three-level nested holders. It reads as: the USER reports Obama's assumption as the latter holds as true that the Iranians do not want to confront other countries.

13. *اوباما: الشعب الايراني لم يعد يرغب في [المواجهة مع العالم الخارجي]*
AwbAmA: AISEb ALAyrAny lm yEd yrgb fy [AlmwAjhp mE AIEAlm AlxArjy]
 Obama: the Iranians no longer **want to** [confront other countries].
rep. USER, report, (*AwbAmA*, true, (*AISEb ALAyrAny*, MOD PRS NEG DESIRE, (*AlmwAjhp mE AIEAlm AlxArjy*)))

Example 14 shows how two conjoined triggers (i.e. *لازم* *LAzm* (must) and *ضروري* *Drwry* (necessary)) that share the same holder and scope are merged into one representation, and the conjunction leads to amplifying the intensity of the obligation set by them both.

14. *لازم وضروري [كلنا نكون قدام مقر المحكمة ومعانا صورة الرئيس]*
LAzm wDrwry [klnA nkwn qdAm mqr AlmHAKmp wmEAnA Swrp Alr}ys]
 We **must** and **it is necessary** that [we go to the court with President's pictures].
rep. USER, STRG PRS AFF REQUIRE, (*klnA nkwn qdAm mqr AlmHAKmp wmEAnA Swrp Alr}ys*)

4 Corpus Harvesting

Tweets are harvested from the Arabic Egyptian Twitter provided that (1) each tweet has at least one trendy political English or Arabic hashtag; and (2) each tweet has at least one candidate event modality trigger from the Arabic modality lexicons (Al-Sabbagh et al. 2013, 2014a). We harvest tweets from a variety of users such as newspapers, TV stations, political and humanitarian campaigns, politicians, celebrities, and ordinary people. Thus, our corpus comprises both MSA, the formal Arabic variety, and EA, the native Arabic dialect of Egypt. The harvested corpus comprises 9949 unique tweets, with 12134 tokens of event modality triggers that map to 315 unique types.

5 Annotation Results

5.1 Evaluation Methodology and Metrics

Our annotation tasks are of two types: (1) Tasks 1-4 are label-based where there is a pre-defined set of labels from which annotators choose; and (2) Tasks 5-6 are segmentation-based where the output of the annotation is a text segment. For the segmentation-based tasks, we use an all-or-nothing method to measure inter-annotator reliability: for segments to be considered as agreement, they must share both the beginning and end boundaries. We use Krippendorff's alpha α (Krippendorff 2011) as our inter-annotator reliability measure, following the most recent work on modality annotation for other languages including English (Rubinstein et al. 2013) and Chinese (Cui and Chi 2013). For more details on Krippendorff's alpha and α , we refer the reader to Artstein and Poesio (2008).

5.2 Results

We use the surveygizmo survey services² to implement our interactive annotation procedure given that their survey structure is one that uses conditional branching and skip logic. We distribute the survey on Twitter and we have three annotators participating. According to the short qualifying quiz given at the beginning of the survey, all three participants are native Egyptian Arabic (EA) speakers who have at least two-year experience with Twitter. They are also university graduates who, therefore, master MSA. None of the participants has a linguistics background. Table 1 shows alpha rates for each annotation task.

	Sense	Polarity	Intensification	Tense	Holder	Scope
Obligation	0.890	0.893	0.892	0.978	0.829	0.744
Permission	0.864	0.905	0.821	0.983	0.800	0.739
Commitment	0.760	0.794	0.783	0.947	0.702	0.654
Ability	0.895	0.914	0.905	0.950	0.828	0.763
Volition	0.921	0.921	0.867	0.982	0.858	0.779
Averages	0.866	0.885	0.854	0.968	0.803	0.736

Table 1: Krippendorff's alpha rates for inter-annotator reliability

5.3 Discussion and Disagreement Analysis

Among the factors that lead to high inter-annotator reliability are that: (1) the vast majority of negation is explicitly marked by negation particles that are easy to detect by human annotators; (2) the vast majority of triggers are used without any amplification or mitigation markers; and (3) punctuation markers are surprisingly informative for marking scope boundaries and direct quotations; and hence, holders.

Sense-related disagreement is attributed to: (1) nominal triggers, (2) highly-polysemous triggers, and (3) different interpretations invoked by the -RATIONAL (i.e. non-human) holders.

Typically, event modality triggers are adjunct constituents that add an extra-layer of meaning and can be removed without disturbing the syntactic structure. Yet, in example 15, واجب $wAjb$ (a must) and أوجب $>wjb$ (a more important must) have main grammatical functions as the predicates of the phrases they modify. Most of the exemplars from Section 2.1 are adjuncts; and, thus, none can substitute واجب $wAjb$ (a must) or أوجب $>wjb$ (a more important must) in such a context.

15. [التحفظ من اختطاف الثورة] واجب لكن [التوحد خلف مشروع] أوجب
[AlthfZ mn AxtTaf Alvwrp] $wAjb$ lkn [Altwhd xlf mSrwE] $>wjb$
[Being cautious about manipulating the revolution] is a **must** but [getting united for one project] is a more important **must**.

Highly-polysemous triggers invoke disagreement because in many cases even the context is ambiguous. In example 16, أقسم $>qsm$ (I swear) has two eligible interpretations: an epistemic trigger interpretation *I assure (you) that* and a commitment trigger interpretation *I promise (you)*

² <http://www.surveygizmo.com/>

that. Even the context is not enough to disambiguate the two interpretations and annotators go by the most common sense for the trigger according to their own opinions.

16. عمرو أديب: أقسم بالله [لن تنسقط #مصر]، احنا شعب 90 مليون ومش إشارة هتسقط بلد
Emrw >dyb: >qsm bAllh [ln tsqT #mSr], AHnA \$Eb 90 mlywn wm\$ <\$Arp htsqT bld
 Amr Adeeb: I **promise/assure** (you) by God that [#Egypt will not collapse]. We are 90 million Egyptians and we will not be defeated by a sign.

Non-human or –RATIONAL holders invoke disagreement, especially for obligation versus volition triggers. The most common sense of such triggers as عايزة *EAYzp* (want) is volition. Yet, when the holder is –RATIONAL like الانتخابات *ALAntxAbAt* (the elections) in example 17, annotators disagree as to whether عايزة *EAYzp* means *want* (i.e. a volition trigger) or *need* (i.e. an obligation trigger).

17. الانتخابات عايزة [مرشحين] وحملات الأحزاب تيجي براحتها.
ALAntxAbAt EAYzp [mr\$Hyn] wHmlAt Al>HzAb tyjy brAHthA
 Elections **want/need** [candidates] and later we can establish the political parties.

Intensity-related disagreement is attributed mostly to progressive verb aspect. Some annotators consider progressive verb aspect as indicated by the EA prefix *b* as a marker for lexical intensity amplification. Thus they tag the volition trigger يتمنى *btmnY* (I wish) in example 18 as amplified, especially it is modified by كل يوم *kl ywm* (everyday).

18. كل يوم يتمنى [سقوط حكم #مرسي].
kl ywm btmnY [sqwT Hkm #mrsy]
 Every day, I **wish** for [#Morsi's regime to fall].

Polarity-related disagreement is mainly caused by (1) negated holders and (2) contextual negation. In يقدر *yqdr* (no one can), annotators disagree as to whether يقدر *yqdr* (can) should be labeled as affirmative or negative. By contextual negation we mean examples like من الصعب أن تتمنى أن *mn AlSEb >n ntmnY >n* (it is hard to wish to), which entails negation due to the adjective الصعب *AlSEb* (hard).

Holder-related disagreement is attributed mainly to generic nouns and impersonal pronouns like الشعب *AlSEb* (the people) and الواحد *AlwAHd* (one), respectively. They are interpreted by some annotators as referring implicitly to the Twitter USER. Therefore, the annotators select the USER as the only holder with zero nesting. Other annotators interpret them as referring to people in general not necessarily the Twitter USER and thus they consider these as instances of nested holders.

Scope-related disagreement is attributed to (1) ambiguous subordinate conjunctions, (2) triggers' modifiers, and (3) absent punctuation markers.

Tense yields almost perfect inter-annotator reliability rates. Annotation disagreement does not show any particular pattern. Therefore, we attribute minor disagreement to random errors, resulting from fatigue.

5.4 Majority Statistics

Based on majority annotations, Table 2 gives the statistics for our corpus in terms of sense, polarity, intensification, and tense. As for holder annotations, approximately 60.5% of the triggers have zero-nested holders (i.e. the tweet's writer is the same as the holder).

	Sense		Polarity		Intensification			Tense	
	MD	NMD	AFF	NEG	AMP	MTG	ASIS	PRS	PST
Ability	1729	920	1047	682	348	308	1073	1175	554
Commitment	1048	495	599	449	221	220	607	639	409
Obligation	1786	848	1059	727	369	399	1018	1018	768
Permission	1699	980	1054	645	286	428	985	1053	646
Volition	1622	1007	974	648	341	292	989	1038	584
Totals	7884	4250	4733	3151	1565	1647	4672	4923	2961

Table 2: Token statistics for each annotation task per event modality sense where MD is modal, NMD is non-modal, AFF is affirmative, NEG is negative, AMP is amplified, MTG is mitigated, ASIS is as is, PRS is present, and PST is past

6 Related Work

Event modality is the focus of many annotation projects. Matsuyoshi et al. (2010) annotate a corpus of English and Japanese blog posts for a number of modality senses including volition, wishes, and permission. They annotate sense, tense, polarity, holders as well as other attributes that we have not covered in our scheme such as grammatical mood. They report macro kappa inter-annotator agreement rates of 0.69, 0.70, 0.66 and 0.72 for holders, tense, sense, and polarity, respectively.

Baker et al. (2010, 2012) simultaneously annotate modality and modality-based negation for Urdu-English machine translation systems. Among the modality senses they work on are requirement, permission, success, intention, ability, and desires. They report macro kappa inter-annotator agreement rates of 0.82 for sense annotation and 0.76 for scopes. They, however, do not annotate holders and do not consider nested modalities.

Hendrickx et al. (2012) annotate eleven modality senses in Portuguese, including necessity, capacity, permission, obligation, and volition, among others. They report a macro kappa inter-annotator rate of 0.85 for sense annotation.

Rubinstein et al. (2013) propose a linguistically-motivated annotation scheme for modalities in the MPQA English corpus. They annotate sense, polarity, holders, and scopes, among other annotation units. They work on obligation, ability, and volition among other modality senses. They attain macro alpha inter-annotator reliability rates of 0.89 and 0.65 for sense and scope, respectively.

Cui and Chi (2013) apply the same scheme of Rubinstein et al. (2013) to the Chinese Penn Treebank and get alpha inter-annotator reliability rates of 0.81 and 0.39 for sense and scope annotation, respectively.

Finally, Al-Sabbagh et al. (2013) annotate event modality in MSA and EA tweets. We attain kappa inter-annotator agreement rates of 0.90 and 0.93 for sense and scope annotation, respectively, for only 772 tokens of event modality triggers.

Our annotation results, therefore, are comparable to the results in the literature. Furthermore, our annotation scheme and its tasks are orthogonal to most of the aforementioned schemes. However, the key differences between our work and related work are:

- We use a standardized taxonomy of event modality - Palmer's (2001) - that has been proved valid for a variety of languages, including Arabic, according to Mitchell and Al-Hassan (1994), Brustad (2000), and Moshref (2012).
- We annotate nested holders unlike some of the aforementioned studies (e.g. Baker et al. 2010, 2012) and use a wider range of negation and intensification markers.
- We use crowdsourcing with simplified guidelines implemented interactively to annotate a larger-scale corpus of 12134 tokens for event modality and its attributes.

7 Conclusion and Outlook

We presented a large-scale corpus annotated for event modality in MSA and EA tweets. We use a simplified annotation procedure that defines each annotation task as a series of questions, implemented interactively. Our scheme covers a wide range of the most common annotation units mentioned in the literature, including modality sense, polarity, intensification, tense, holders, and scopes. We deal with nested holders - which are crucial in a highly interactive genre such as tweets where users frequently quote others and make assumptions about them. We also automatically merge triggers with shared holders and scopes based on elicited annotators' answers. The annotation procedure yields reliable results and creates a novel resources for Arabic NLP. The current version of our corpus does not, however, cover a number of issues including: the future tense, grammatical moods other than the declarative, and modality entailment. By modality entailment, we mean, for example, when a tweet's user criticizes the obligation of another quoted person, this entails that the user does not consider such an event as required. For a future version of the corpus, we plan to cover such points. Furthermore, we will use the corpus to train and test a machine learning system for the automatic processing of Arabic event modality.

References

- Rania Al-Sabbagh, Jana Diesner and Roxana Girju. 2013. Using the Semantic-Syntactic Interface for Reliable Arabic Modality Annotation. In *Proceedings of IJCNLP'13*, pages 410-418, October 14-18, 2013, Nagoya, Japan.
- Rania Al-Sabbagh, Roxana Girju and Jana Diesner. 2014a. Unsupervised Construction of a Lexicon and a Pattern Repository of Arabic Modal Multiword Expressions. In *Proceedings of the 10th Workshop of Multiword Expressions at EACL'14*, April 26-27, 2014, Gothenburg, Sweden.
- Rania Al-Sabbagh, Roxana Girju and Jana Diesner. 2014b. *3arif*: A Corpus of Modern Standard and Egyptian Arabic Tweets Annotated for Epistemic Modality Using Interactive Crowdsourcing. In *Proceedings of the 25th International Conference on Computational Linguistics*, August 23-29, 2014, Dublin, Ireland.
- Ron Artstein and Massimo Poesio. 2008. Inter-Coder Agreement for Computational Linguistics. *Computational Linguistics*, volume 34, issue 4, pages 555-596.
- Kathrin Baker, Michael Bloodgood, Mona Diab, Bonnie Dorr, Nathaniel W. Filardo, Lori Levin and Christine Piatko. 2010. A Modality Lexicon and its Use in Automatic Tagging. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC'10)*, pages 1402-1405, May 19-21, 2010, Valetta, Malta.
- Kathryn Baker, Michael Bloodgood, Bonnie J. Dorr, Chris Callison-Burch, Nathaniel W. Filardo, Christine Piatko, Lori Levin and Scott Miller. 2012. Modality and Negation in SIMT. *Computational Linguistics*. volume 38, issue 2, pages 411-438.
- Samuel R. Bowman and Harshit Chopra. 2012. Automatic Animacy Classification. In *Proceedings of the NAACL HTL 2012 Student Research Workshop*, pages 7-10, June 3-8, 2012, Montreal, Canada.
- Kristen E. Brustad. 2000. *The Syntax of Spoken Arabic: A Comparative Study of Moroccan, Egyptian, Syrian and Kuwaiti Dialects*. Georgetown University Press, Washington DC, USA.
- Cohan Sujay Carlos and Madulika Yalamanchi. 2012. Intention Analysis for Sales, Marketing and Customer Service. In *Processing of COLING 2012: Demonstration Papers*, pages 33-40, December 2012, Mumbai, India.
- Baptiste Chardon, Farah Benamara, Yannick Mathieu, Vladimir Popescu and Nicholas Asher. 2013. Sentiment Composition Using a Parabolic Model. In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)*, pages 47-58, March 20-22, 2013, Potsdam, Germany.
- Yanyan Cui and Ting Chi. 2013. Annotating Modal Expressions in the Chinese Treebank. In *Proceedings of the IWC 2013 Workshop on Annotation of Modal Meaning in Natural Language (WAMM)*, pages 24-32, March 2013, Potsdam, Germany.
- Iris Hendrickx, Amália Mendes and Silvia Mencarelli. 2012. Modality in Text: A Proposal for Corpus Annotation. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'12)*, pages 1805-1812, May 21-27, 2012, Istanbul, Turkey.
- Klaus Krippendorff. 2011. Computing Krippendorff's Alpha-Reliability. Annenberg School of Communication, Departmental Papers: University of Pennsylvania.
- Andrew Lampert, Robert Dale and Cecile Paris. 2010. Detecting Emails Containing Requests for Action. In *Proceedings of Human Language Technologies: the 2010 Annual Conference of the North American Chapter of the ACL*, pages 984-992, June 2010, Los Angeles, California.
- Ying-Shu Liao and Ting-Gen Liao. 2009. Modal Verbs for the Advice Move in Advice Columns. In *Proceedings of the 23rd Pacific Asia Conference on Language, Information and Computation*, pages 307-316, December 3-5, 2009, Hong Kong, China.
- Suguru Matsuyoshi, Megumi Eguchi, Chitose Sao, Koji Murakami, Kentaro Inui and Yuji Matsumoto. 2010. Annotating Event Mentions in Text with Modality, Focus and Source Information. In *Proceedings of LREC'10*, pages 1456-1463, May 19-21, 2010, Valetta, Malta.
- T. F. Mitchell and S. A. Al-Hassan. 1994. *Modality, Mood and Aspect in Spoken Arabic with Special Reference to Egypt and the Levant*. London and NY: Kegan Paul International.

- Ola Moshref. 2012. *Corpus Study of Tense, Aspect, and Modality in Diglossic Speech in Cairene Arabic*. PhD Thesis. University of Illinois at Urbana-Champaign.
- Frank R. Palmer. 2001. *Mood and Modality*. 2nd Edition. Cambridge University Press, Cambridge, UK.
- J. Ramanand, Krishna Bhavsar and Niranjan Pedaneekar. 2010. Wishful Thinking: Finding Suggestions and "Buy" Wishes for Product Reviews. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to the Analysis and Generation of Emotion in Text*, pages 54-61, June 2010, Los Angeles, California.
- Aynat Rubinstein, Hillary Harner, Elizabeth Krawczyk, Daniel Simoson, Graham Katz and Paul Portner. 2013. Toward Fine-Grained Annotation of Modality in Text. In *Proceedings of the IWC 2013 Workshop on Annotation of Modal Meaning in Natural Language (WAMM)*, pages 38-46, March 2013, Potsdam, Germany.
- Roser Saurí and James Pustejovsky. 2009. FactBank: A Corpus Annotated with Event Factuality. *Language Resources and Evaluation*, 43:227-268
- György Szarvas, Veronika Vincze, Richárd Farkas and János Csirik. 2008. The BioScope Corpus: Annotation for Negation, Uncertainty and their Scope in Biomedical Texts. In *Proceedings of BioNLP 2008: Current Trends in Biomedical Natural Language Processing*, pages 38-45, June 2008, Columbus, Ohio, USA.
- Janyce Wiebe, Theresa Wilson and Claire Cardie. 2005. Annotating Expressions of Opinions and Emotions in Language. *Language Resources and Evaluation*, volume 39, issue 203, pages 1663-210.