Using the Semantic-Syntactic Interface for Reliable Arabic Modality Annotation

Rania Al-Sabbagh
Department of Linguistics
University of Illinois at Urbana-Champaign
alsabba1@illinois.edu

Jana Diesner
School of Library and Information Science
University of Illinois at Urbana-Champaign
jdiesner@illinois.edu

Roxana Girju
Department of Linguistics
University of Illinois at Urbana-Champaign
girju@illinois.edu

Abstract

We introduce a novel modality scheme where triggers are words and phrases that convey modality meanings and subcategorize for clauses and verbal phrases. This semantic-syntactic working definition of modality enables us to design practical and replicable annotation guidelines and procedures that alleviate some shortcomings of current purely semantic modality annotation schemes and yield high inter-annotator agreement rates. We use this scheme to annotate a tweet-based Arabic corpus for modality information. This novel language resource, being the first, initiates NLP research on Arabic modality.

1 Introduction

Modality is the grammaticalized expression of the "speaker's subjective attitudes" (Bybee et al., 1994:176) and "psychological stances" (Mitchell and al-Hassan, 1994:7) towards propositions and events and their factuality status. In NLP applications and domains, modality is considered as one linguistic means to convey and detect attitudes and opinions (Wiebe at al., 2005; Abdul-Mageed and Diab, 2012), commitments and beliefs (Diab et al., 2009), power relations (Prabhakaran et al., 2012), uncertainties and speculations (Szarvas et al., 2008; Matsuyoshi et al., 2010).

We herein present the first work on Arabic modality annotation, which is part of a larger research project to **(1)** automatically identify modality triggers (i.e. words and phrases conveying modality meanings), holders (i.e. modality experiencers) and scopes (i.e. the range of linguistic constituents modified by the modality triggers), and **(2)** automatically detect power relations among participants in the social network of Twitter by using this modality information.

Despite solid work on Arabic modality in theoretical linguistics (Mitchell and al-Hassan, 1994; Brustad, 2000; Moshref, 2012), there are no Arabic corpora annotated for modality, not even the widely used Penn Arabic Treebank. However, there is a plethora of work and annotated corpora for modality in other languages, including English (Saurí et al, 2006; Baker et al., 2010; Prabhakaran et al., 2012; Rubinstein et al., 2013), Portuguese (Hendrickx et al., 2010; Avila and Mello, 2013), Japanese (Matsuyoshi et al. 2010) and Chinese (Cui and Chi, 2013).

Arabic modality annotation involves multiple challenges. First, the paradigm of Arabic modality triggers is complex as it includes auxiliaries, lexical verbs, nominals and particles - like many other languages as well. Second, triggers can be lexically and/or semantically ambiguous: a lexically-ambiguous trigger is a word or phrase that may or may not convey a modality meaning based on context. A semantically-ambiguous trigger is a word or phrase that may convey two or more modality meanings. Third, implicit scopes are common in Arabic and annotators have to be made aware of them. Fourth, Arabic word order flexibility allows triggers - especially adverbials - to occur in the scope's initial, medial or final positions, which makes it challenging for annotators to identify scope spans. Finally, modality scopes are not necessarily adjacent to their triggers, which furthermore complicates the detection of their spans.

The tweets genre on which we work poses an additional challenge due to language variation. We select a random sample of Arabic tweets from the YADAC corpus (Al-Sabbagh and Girju, 2012) posted in Egypt during the first six months

of the 2011 revolution. All selected tweets are about the political situation at that time. Tweets are not only in the Egyptian Arabic (EA) dialect, but also in Modern Standard Arabic (MSA), especially tweets from press agencies and celebrities. Therefore, our annotation scheme has to work on both MSA and EA modality.

Arabic and the tweets genre are not the only original aspects of this paper. We present a novel linguistically-motivated annotation scheme with a semantic-syntactic working definition of modality triggers as *words and phrases that convey modality meanings and subcategorize for clauses and verbal phrases*. Modality meanings are based on Palmer's (1989, 2001) cross-lingual typology of modality, which is proven valid for both MSA and EA (Mitchell and al-Hassan, 1994; Brustad, 2000; Moshref, 2012).

The semantic-syntactic interface between modality triggers and their scopes depicted in our definition is well-established in theoretical linguistics for Arabic (Moshref, 2012) and for English (Jackendoff, 1972; Brennan, 1993; Butler, 2003). Semantics and syntax work simultaneously such that semantics guarantees not to tag all words and phrases that subcategorize for clauses and verbal phrases as modality triggers. Syntax guarantees not to tag words and phrases that share some semantic meanings with modality triggers, but do not subcategorize for clauses and verbal phrases, as modality triggers.

Subcategorization frames of modality triggers are sporadically mentioned in the NLP literature on modality annotation. For English, Saurí et al. (2006) stated in their definition of event modality triggers that "they subcategorize for a that, gerundive or infinitival clause, but also an NP headed by event denoting nouns" (p.334). For Portuguese, Hendrickx et al. (2010) stated that "in the majority of the cases, the target [i.e. scope] is a subordinate clause or a verbal phrase ... in some cases, also main clauses can be targets [i.e. scopes]" (p. 1810). However, no prior work integrates the semantic-syntactic interface into a modality annotation scheme.

Our conceptualization and implementation of this semantic-syntactic interface provide annotators with practical annotation guidelines that yield highly-reliable results, as shown herein. Furthermore, they define modality in terms of concrete syntactic features which we use in our future work for the automatic identification of triggers and their scopes.

The rest of the paper is structured as follows: section 2 briefly reviews related work. Section 3

gives details about Arabic modality and its complexities. Section 4 presents our data, annotation guidelines and procedures. Section 5 reports the annotation results. Finally, we conclude with future work in Section 6.

2 Related Work

Recent work on modality annotation focuses on English, Portuguese, Japanese and Chinese. Baker et al. (2010) used an idiosyncratic categorization of English modality that distinguished 8 semantic meanings: requirement, permissive, success, effort, intension, ability, want and belief. They defined each type as a pattern of the form H (modal) P where H is the sentence's agent and P is the proposition (e.g. H permits [P to be true/false]). They obtained an average inter-annotator agreement rate of 0.82. The error analysis of their modality tagger showed that errors resulted primarily from the triggers' lexical ambiguity.

Prabhakaran et al. (2012) focused on 5 semantic meanings of English modality, and used the same *HP* patterns as Baker et al. (2010) for annotation guidelines. They reported an inter-annotator agreement rate of 0.95. Their modality tagger yielded a 0.44 F_1 score against a gold-standard and 0.79 and 0.91 F_1 scores against different testing sets from their crowdsourced data.

Rubinstein et al. (2013) used a more standardized typology of English modality that entailed (1) priority modality divided into bouletic, teleological and deontic triggers; and (2) non-priority modality divided into epistemic, circumstantial and ability triggers. Their purely semantic annotation scheme returned an alpha reliability score of 0.89 only when collapsing the subtypes of priority and non-priority triggers. The scheme yielded an alpha reliability score of 0.65 for scope span annotation.

Cui and Chi (2013) applied Rubinstein et al.'s (2013) scheme for modality annotation to the Penn Chinese Treebank. They obtained a reliability score of 0.94 for triggers' annotation using the collapsed binary typology of modality triggers as priority vs. non-priority. Their error analysis reported vagueness in the annotation guidelines as one disagreement factor.

The lack of previous NLP work on Arabic modality, modality annotation in tweets and syntactically-guided modality annotation schemes render direct comparisons to our work impossible. Yet, the two main distinguishing factors of our work are: (1) to guarantee the replicability of

our study, we avoid idiosyncratic typologies of modality that were used in some previous work; and (2) to better guide our annotators, we use practical guidelines that rely on both semantics and syntax rather than semantics only as in previous annotation schemes.

3 Arabic Linguistic Modality

3.1 Background

Among multiple typologies of modality, Palmer's (1989, 2001) was validated for both MSA and EA in theoretical linguistics (Mitchell and al-Hassan, 1994; Brustad, 2000; Moshref, 2012).

Palmer distinguishes two main classes of modality: propositional and event. Propositional modality is concerned with the speaker's attitude to the truth-value of a proposition, and includes:

- Epistemic modality, which expresses the speaker's judgment about the factual status of the proposition as well as the speaker's opinion and attitude towards that proposition.

- Evidential modality, which indicates the evidence the speaker has for his or her judgment or opinion. Evidence can be reported as in hearsay and quotes or sensory.

Event modality refers to events that are not actualized but are merely potential, and includes:

- Deontic modality, which relates to obligations and permissions that emanate from an external source, and commissives, which originate from an internal source as speakers lay an obligation on themselves for a potential event.

- Dynamic modality, which relates to ability, willingness and wishes.

3.2 Challenges

The challenges of Arabic modality annotation are attributed to (1) the complexity of the Arabic modality paradigm, (2) the lexical and semantic ambiguity of Arabic modality triggers, (3) implicit scopes, (4) word order flexibility and (5) potential long dependencies between triggers and their scopes.

The paradigm of Arabic modality triggers includes a large set of auxiliaries, lexical verbs, nominals and particles. Except for auxiliaries, adverbs and some particles, all modality triggers inflect for gender, number, person, tense, aspect and mood. Furthermore, generic modality patterns such as * mn Al-* >n (it is * that), where * is typically an adjective (e.g. من المهم أن

mn Almhm >n (it's important that)), are common.

Modality triggers can be lexically and/or semantically ambiguous. The noun   zmAn is one example of a lexically-ambiguous trigger because in 1 it is an epistemic with a clause scope. Yet, in 2 it is a non-modal standing for *era*.

.1 عندها خبرة   مبارك لسه بيحكمنا

*lw smEnA klAm AlnAs Ally EndhA xbrp kAn **zmAn** mbArk lsh byHkmnA*

If we'd listened to the elite, <u>Mubarak</u> **would have been still ruling us.**

.2   الحكم الفردي انتهى

zmAn AlHkm Alfrdy AnthY xlAS

The **era** of individual rulers has come to an end.

The MSA particle   lAbd is one example of semantically-ambiguous triggers because in 3 it is an epistemic with a clause scope; whereas in 4 it is an obligative with a verbal-phrase scope.

.3 أنه تذكر صدام حسين وهو يستلقي على سريره الطبي

*<n mbArk **lAbd** w>nh tzkr SdAm Hsyn whw ystlqy ElY sryrh AlTby fy qAEp AlmHkmp*

It must be that <u>Mubarak remembered Saddam Hussein</u> as he was lying on his medical bed in the court.

.4   يوضع

lAbd An ywDE mbArk fY Alsjn

Mubarak **must** <u>be put in jail</u>.

Implicit modality scopes are common in Arabic and come in different realizations. In 5, the scope of the permissive   nsmH (allow) is the deictic   *lk (that) which refers to the clause يهان المصري   *ln yhAn AlmSrywn* (Egyptians won't be humiliated). That is, the scope of *nsmH* is actually a clause.

.5 يهان مصري .   .

*ln yhAn AlmSrywn. ln **nsmH** b*lk.*

<u>Egyptians won't be humiliated</u>. We won't **allow** it.

In 6, the abilitives   *AErf* (can) and *EArf* (can) share the same verbal-phrase scope of   *A$wf* (see). To avoid redundancy, the speaker elides the scope of the second abilitive -   *AErf* (can) - and does not replace it with any deictic expression. Thus *AErf* modifies an implicit verbal-phrase scope.

.6

m$ EArf A$wf HAjp lmA AErf Hklmk

I can't <u>see anything</u>. When I **can**, I'll call you.

On the surface level, the obligative   *lAzm* (must) in 7 is followed by the noun phrase *a real reaction against military trials*. Yet, on a deeper level, the tweet is the short version of *we must*

(take) a real reaction against military trials. This means that *lAzm* has an implicit verbal-phrase scope.

7. ( ) العسكرية

lAzm (naxd) mwqf bjd Dd AlmHAkmAt AlEskryp

We **must** (take) a real reaction against military trials

Word order flexibility allows for some modality triggers - especially adverbials - to occur before, after or in the middle of their scope(s).

Long dependencies between modality triggers and their scope(s) are the last challenge with Arabic modality annotation. The obligative *ATlb* (ask; require) in 8 subcategorizes for a complement clause, which starts 9 words later (affixes excluded).

8. من عناصر الاجهزة الأمنية المصرية المتخفية في ملابس مدنية
انهم يتفرجو _____ عشان سلوكهم ولبسهم ونظرتهم مهروشة أوي

ATlb *mn EnASr AlAjhzp Al>mnyp AlmSryp Almtxfyp fy mlAbs mdnyp Anhm ytfrjw Ely kwnAn E$An slwkhm wlbshm wnZrthm mhrw$p >wy*

I **ask** Egyptian security individuals disguising in civil outfits to watch Conan because their behavior, outfit and looks are ridiculously revealed.

4 Arabic Modality Annotation

4.1 Corpus Encoding and Description

We randomly selected a corpus of 1,704 raw tweets (33,349 tokens and 11,013 unique types) from the YADAC corpus (Al-Sabbagh and Girju, 2012). The considered time span ranges from January 25, 2011 to June 30, 2011. All tweets were posted in Egypt by ordinary individuals, celebrities (e.g. politicians, actors, singers, TV hosts), and the press (e.g. newspapers, TV stations, NGOs, election campaigns).

The corpus includes tweets in both MSA and EA because press users always post in MSA, while celebrities and ordinary individuals frequently switch between MSA and EA. Based on our manual annotation of user types, we have 1,318 tweets posted by individuals, 369 tweets by celebrities and 17 tweets by the press.

4.2 Annotators and Annotation Units

Two EA native speakers performed the annotation. Being linguistics students, they can be assumed to master MSA. They were given a one-hour video tutorial covering the annotation guidelines and procedures in Sections 4.3 and 4.4, respectively, followed by a 30-minute workshop dedicated to training and discussion.

Each annotator is required to label each **(1)** modality trigger, **(2)** its semantic meaning, **(3)** its scope type(s), and **(4)** its scope span(s). We keep

holder annotation for future work as it poses additional challenges.

4.3 Annotation Guidelines

Our core annotation guidelines are summarized in the semantic-syntactic working definition of modality given in Section 1. We define modality triggers as words and phrases that (1) convey a modality meaning from Palmer's (1989, 2001) typology, (2) and subcategorize for clauses and verbal phrases; representing propositions and events, respectively. We also give the annotators a number of supplementary guidelines.

Annotators have to label each trigger and its scope(s). Multiple triggers may have the same scope as in 9 where the two epistemic triggers *EArf* (I know) and *mt>kd* (I'm sure) share the clause scope of *that Mubarak won't be executed*.

9. حيتعدم _____

AnA EArf wmt>kd An mbArk m$ HytEdm

I **know** and I'**m sure** that Mubarak won't be executed.

Annotators have to label all the scopes of the modality trigger for type and then identify their spans. In 10, the obligative *lAzm* (must) modifies three verbal-phrase scopes linked by the coordinating conjunction *w* (and).

10. مانسكتش عليه نجيبه من جدوره _____
يحصلش تاني

AHnA lAzm ntAbE AlmwDwE wmAnskt$ Elyh wnjybh mn jdwrh E$An mA yHSl$ tAny

We **must** follow up with this, not ignore it and investigate it well so it won't happen again.

Finally, annotators have to retrieve implicit scopes whether they are referred to in-text or using their own real-world knowledge.

4.4 Annotation Procedure

Annotation proceeded in four stages. For **Stage 0**, we used our novel, manually-built, large-scale Arabic Modality Lexicon (AML) to automatically pre-highlight candidate modality triggers. AML was built in three steps:

- First, we manually generated the person, gender, number, tense, mood and aspect inflections as well as the present and past participle derivations of 276 lemmas compiled from Mitchell and al-Hassan (1994), Brustad (2000) and Moshref (2012).
- Second, we added a list of triggers including particles, adverbs and multi-word generic expressions that do not inflect for person, gender, number, tense, mood and aspect.
- Finally, we labeled each entry for an English

<entry id="997" token="متهيألي" trans.="mthy>ly" gloss="I think" ambiguity="NA" dialect="EA" semClass="epistemic" features="NA" </entry>

<entry id="2032" token="   " trans="mn AlmHtm" gloss="it's essential that" ambiguity="NA" dialect="MSA" semClass="obligative" features="MWE" </entry>

<entry id="3423" token="   " trans="Ejz" gloss="failed to" ambiguity="lexical" dialect="MSA/EA" semClass="abilitive" features="inherentlyNeg,Quasi"</entry>

Table 1: An expert from the Arabic Modality Lexicon (AML)

gloss, ambiguities {lexical, semantic, both, NA}, dialects {MSA, EA, both}, modality semantic meaning and special features {quasi; inherently-negative, multi-word expression} as in Table 1.

Currently, AML has 7,584 entries, with the statistical distributions in Table 2. Despite the large size of AML, annotators were instructed to add any words or phrases that match our working semantic-syntactic definition of modality.

Semantic meanings		Ambiguity	
Epistemic	3,144	Lexical	2,363
Sensory	134	Semantic	155
Reported	427	Lexical/Semantic	116
Obligative	1,091	Unambiguous	4,950
Permissive	815		
Commissive	132		
Abilitive	957		
Volitive	884		
Dialects		**Special Features**	
MSA	2,268	Quasi	777
EA	3,100	Inherently-Neg.	788
MSA/EA	2,216	MWE	276

Table 2: AML statistics

For **Stage 1**, annotators labeled each pre-highlighted modality trigger for its modality semantic meaning. We defined modality semantic annotation as a synonymy judgment task where the annotators, given a number of synsets, had to decide to which synset the pre-highlighted trigger belongs. We used 8 synsets; each of which featured one modality semantic meaning from Palmer's (1989; 2001) typology. The average size of the synsets is 15 words/phrases to represent different shades of meaning. Yet, due to space limitations, we only included sample synsets in Table 3. To avoid fatigue, disinterest, and distraction effects, we used counterbalancing and prompted the annotators to provide their own synonym(s) for the pre-highlighted candidate trigger if none of the given synsets seemed synonymous.

For **Stage 2**, annotators labeled the syntactic type of the linguistic constituents modified by the pre-highlighted modality trigger (i.e. scope type) where applicable. Annotators had to choose whether the modified constituent was a clause, a verbal phrase, or another type of constituency (e.g. a noun phrase, an adjectival phrase). Once

the clause or the verbal phrase option was selected, annotators were prompted to extract that clause or verbal phrase.

Epistemic (opinion, conclusion, possibility)	- في رأيه	-	-	- -	-
Evidential (reported)	-	-	-	- -	-
Evidential (sensory)	شاهد بعينيه ـ سمع بنفسه ـ				
Obligative (and necessity)	يجب ـ -	-	-	-	
Permissive (and prohibitive)	- ـ نهى عن ...	-	-	-	-
Commissive	- ـ تعهد ـ عاهد ...	-			
Abilitive (incapability)	- ـ تسنى له -	-	-	...	
Volitive	- قد النية على	-			

Table 3: Sample synsets used for modality semantic annotation

For **Stage 3**, we automatically extracted the triggers that followed our semantic-syntactic working definition of modality. That is, triggers labeled as synonymous to one of the synsets in Table 3 AND as modifying a clause or a verbal phrase. In this stage, instances such as *AftkrwA* in 11 were automatically excluded: although it modifies a complement clause, it means *remember* which is a non-modality meaning.

.11 نمشي تحقيق مطالبنا بيبقى

AftkrwA <n kl mrp bnm$y tHqyq mTAlbnA bybqY <SEb

Remember that every time we leave, it becomes harder to achieve our demands!

Similarly, instances such as *>ElnwA* (they announced) in 12 were automatically filtered out. It is synonymous with the evidential reported synset, yet it modifies a prepositional phrase.

.12 أخيرا عن أسماء المعتقلين في السجن الحربي

>ElnwA >xyrA En >smA' AlmEtqlyn fy Alsjn AlHrby

They finally **announced** the names of all prisoners in the military jail.

Instances such as *qAdr* (able to) in 13 were automatically admitted as valid modality triggers. It belongs to the dynamic abilitive synset, and modifies the verbal phrase أن يصنع *>n ySnE* (to make).

.13 أن يصنع مصيره بيده

Al$Eb AlmSry qAdr >n ySnE mSyrh bydh

The Egyptian people **are able to** <u>make their own destiny</u>.

Our annotation procedure pinpoints the efficiency and applicability of each dimension of our definition of modality. These guidelines also speed up the annotation process and increase annotation reliability because they provide annotators with practical and concrete prompts, and elicit well-structured answers that can be automatically converted into the modality annotation profiles described in Section 4.5.

4.5 Modality Annotation Profiles

Twitter terms of services prohibit redistributing raw tweet texts. Thus at the end of the annotation process, a profile was built for each tweet with its user name, tweet ID, and modality-related information. Associated software is to be given to help reconstruct tweets using their IDs. Although at the time of writing this paper all tweet IDs are still active, there is a potential of degradation if users delete their tweets or make their accounts private. This does not affect the modality-related profile, however, the complete tweet text will not be available. Modality-related information presents chunks of the tweet texts that represent the trigger word/phrase and the scope clause/verbal phrase. Thus we assume that we are not violating the terms of services.

Triggers are marked with 4-character labels. The first character is *T* for Trigger. The second two characters indicate the semantic meaning of the trigger {*Ep*: epistemic, *Rp*: reported, *Sn*: sensory, *Ob*: obligative, *Pr*: permissive, *Cm*: commissive, *Ab*: abilitive, *Vl*: volitive}. The fourth character is an index to indicate whether the trigger is the 1st, 2nd and so on in the tweet and to relate the trigger to its scope(s).

Scopes are marked with 3-character labels. The first is *S* for Scope. The second represents the syntactic type of the scope - {*C*: clause, *P*: verbal phrase}. The last is an index matching that of its trigger. Table 4 shows the modality annotation profiles for examples 1 and 9, respectively.

```
user="alaa"
tweet_id=" 718574588884458240"
[[ ("zmAn", "TEP1"),
    (("mbArk lsh byHkmnA", "SC1")) ]]

user="eAiNet"
tweet_id="46316910177697792"
[[  ("EArf", "TEP1"),
        (("An mbArk m$ HytEdm", "SC1"))],
    [("mt>kd", "TEP2"),
        (("An mbArk m$ HytEdm", "SC2")) ]]
```

Table 4: Example modality annotation profiles

5 Annotation Results

5.1 Inter-Annotator Agreement Rates and Disagreement Factors

AML pre-highlighted 2,892 candidate triggers in our 1,704 tweets. We used the kappa statistics to measure the Inter-Annotator Agreement (IAA) rates for:

- **Modality semantic annotation**: this labels each candidate trigger as synonymous to one of the synsets in Table 3 featuring Palmer's (1989; 2001) typology.
- **Modality syntactic annotation:** this includes (1) **identifying the scope type** as to whether it is a clause, a verbal phrase, or none; and (2) **identifying the scope span** in terms of the beginning and the end of each scope.

Our macro kappa IAA rate for modality semantic annotation is 0.899 (Table 5). It is hard to measure if this rate is significantly higher than rates reported in the literature of modality semantic annotation because direct comparison with prior work is not possible as explained in Section 2. Yet, one point to highlight is that we do not use a collapsed typology of modality semantic meanings as in Rubenstein et al. (2013) and Cui and Chi (2013), who both collapsed modality semantic meanings into two major classes only: priority vs. non-priority.

	Kappa	Percent Agreement
Semantic annotation	0.899	0.918
Scope type	0.846	0.902
Scope span	0.929	0.973

Table 5: Macro kappa inter-annotator agreement rates for modality semantic and syntactic annotations

We attribute our high IAA rate for modality semantic annotation to: (1) the large-scale AML, which provides annotators with an extensive list of candidate triggers; and (2) using synonymy judgments to give annotators practical, self-evident annotation prompts instead of subjective guidelines, defining modality triggers as expressions of alternative states in which the world could be.

There are, however, two limitations to using synonymy judgments for modality semantic annotation. First, the quality of the annotation relies on the quality of the used synsets. It is important to select unambiguous triggers to represent the modality semantic meanings in different contexts. This is because triggers interact with other linguistic features such as modification, negation and grammatical mood. Second, to better guide

the annotators, especially when working on a morphologically-rich language such as Arabic, it is better to have the synset members inflected for the same person, gender, number, tense, mood and aspect as the candidate trigger.

It took us three iterations of annotations - each one with two different annotators - to come up with the best final synsets used in this paper. This process is time and labor consuming. Yet, once the synsets have been created, the annotation process is fast and replicable with a potential to be crowdsourced. We will examine this option in future work.

Highly-ambiguous lemmas are the first disagreement factor for modality semantic annotation. Epistemic lemmas such as *$Af* (saw), *Erf* (knew), فهم *fhm* (understood), *Sdq* (believed) and *qAl* (said) among others have multiple meanings of which one or more might be modality-related. This explains why most of the disagreement scores in Table 6 are between modality and non-modality meanings (i.e. NA).

	Ep	Rp	Sn	Ob	Pr	Cm	Ab	Vl	NA
Ep	**610**	10	3	1	9	0	9	0	33
Rp	7	**261**	0	13	0	0	0	0	20
Sn	2	0	**192**	0	0	0	0	0	5
Ob	0	8	0	**299**	3	0	1	1	0
Pr	0	0	0	1	**93**	0	6	0	0
Cm	0	0	0	0	0	**7**	0	0	0
Ab	0	0	0	0	0	0	**124**	0	0
Vl	0	0	0	0	0	0	0	**267**	1
NA	98	6	0	0	0	0	0	0	**802**

Table 6: Confusion matrix for semantic annotation

For one annotator نفهم *nfhm* in 14 is synonymous to *n$rH* (explain) and thus does not belong to any of the modality synsets in Table 3. For the other annotator, this trigger still means *explain,* but not as in explaining factual information, but as in making people adopt a specific point of view or a belief. Thus it is synonymous to نجعلهم يؤمنو *njElhm y&mnwn >n* (make them believe that) and is an epistemic trigger.

14. **نفهم** إن المجلس العسكري حاجة والجيش حاجة

lAbd nfhm AlnAs <n Almjls AlEskry HAjp wAljy$ Hajp

We must **explain to** people that the Supreme Council of Armed Forces is one thing and the army is another.

The same lemma may also have more than one closely-related modality meaning; (i.e. it is semantically-ambiguous). For one annotator بيقول *byqwl* in 15 is an evidential reported trigger meaning *is saying;* whereas it is an epistemic trigger meaning *is thinking* for the other.

15. قفشوا ناس معاهم سلاح وطلعوهم بره الميدان عشان بس اللي **بيقول** إننا بلطجية يعرف إنه غلطان

qf$wA nAs mEAhm slAH wTlEwhm brh AlmydAn E$an Ally byqwl <nnA blTjyp yErf <nh glTAn

They arrested some people with guns and kicked them out of the square so that those **saying** we're thugs realize that they are wrong.

Modality triggers not included in AML are the second disagreement factor for modality semantic annotation. A total of 168 triggers were identified as new; 85 of which were agreed upon by both annotators. For future modality annotation, agreed-upon new triggers will be added to AML and controversial ones are to be examined by experts prior to inclusion.

The macro kappa IAA rate for scope type identification is 0.846 according to Table 5. Main factors of disagreement are:

- **Clauses vs. verbal phrases:** in some contexts, triggers such as *mmkn* (may, it's possible that) and *Drwry* (must; it is necessary that) can be understood either as auxiliaries subcategorizing for verbal phrases or as adjuncts subcategorizing for clauses. Thus *Drwry nnzl nqwl l>* can be either *we must protest and say no* or *it's necessary that we protest and say no*.

- **Implicit scope recovery:** implicit scopes with deictic expressions or in-text reference were easy to retrieve unlike implicature-based scopes. For instance, احنا صدقنا خطابه *AHnA SdqnA xTAbh* (we believed his speech) was perceived by one annotator as *we believed what he said in his speech was true*. Thus the annotator selected the clause option for the scope type. The other annotator did not see such an implicature and thus selected *NA*, meaning that the scope is unrecognizable or is neither a clause nor a verbal phrase.

The macro kappa IAA rate for scope span recognition in Table 5 is 0.929 which is quite high. We attribute this to the simplicity of the tweet genre, which entails short sentences (140 characters or less) and a writing style that resembles short telegraphic notes more than formal and lengthy sentences. Interjections, adjuncts and subordinate conjunctions are the main reasons for disagreement. In 16, one annotator ends the span of the clause-based scope before *El$An* (so that); while the other includes the entire sentence into the scope span. We will add clearer guidelines for when interjections, adjuncts and subordinate clauses should be considered into the scope span in future work.

16. _____ (علشان اللي في الشارع يحسوا ان فيه حد جنبهم)

lAzm AlnAs tnzl (El$An Ally fy Al$ArE yHswA An fyh Hd jnbhm)

People **must** go to protests (so that those already protesting won't feel as if left alone).

As we implemented Stage 3 of our annotation procedure, we sought triggers that adhere to our semantic-syntactic working definition of modality. Triggers labeled as conveying a modality meaning AND subcategorizing for either clauses or verbal phrases were 1,746 and 1,619 triggers by Annotators 1 and 2, respectively. Triggers labeled as not conveying a modality meaning and/or not modifying a clause or a verbal phrase were 1,146 and 1,273 triggers by Annotators 1 and 2, respectively. Exact matches between the two annotators (i.e. triggers labeled similarly for modality meanings, scope type(s) and scope span(s)) amount to 1,343 valid modality triggers according to our definition.

	Modal	Non-Modal
Annotator 1	1,746	1,146
Annotator 2	1,619	1,273
Agreed-Upon (Exact-Match)	1,343	1,034
Total Exact matches	2,377	

Table 7: Exact-match modality annotated corpus

5.2 Annotated Corpus Statistics

In this section, we give statistics on candidate triggers labeled identically by both annotators: whether triggers eventually considered as valid modality triggers (i.e. 1,343 triggers) or triggers eventually rejected as invalid modality triggers (i.e. 1,034). Table 8 shows the correlation between modality semantic meanings and their scope types. We conclude that:

- Except for evidential sensory, modality triggers are more likely to modify clauses and verbal phrases than other linguistic constituents such as noun, adjective and adverb phrases. That is, modality triggers subcategorize for clauses and verbal phrases.

- Propositional modality (i.e. epistemic, evidential reported and sensory) subcategorizes more frequently for clauses; whereas event modality (i.e. deontic and dynamic) is more likely to subcategorize for verbal phrases.

- Triggers that were pre-highlighted as candidates by AML and later rejected for being invalid according to our definition correlate more frequently with linguistic constituents other than clauses and verbal phrases.

- Only 2% of scopes are implicit.

Based on AML dialect labels, valid modality triggers are: **(1)** 77.3% EA-exclusive such as متهيألي *mthy>ly* (I think) and عايز *EAyz* (I want), **(2)** 15.6% either MSA or EA based on context such as *qAl* (he said) and *Erf* (he knew),

and **(3)** 7% MSA-exclusive such as *Astwjb* (it necessitated) and *wddt* (I wanted).

	Clause	V. Phrase	Implicit	NA	Total
Ep	477	23	15	33	**548**
Rp	197	6	7	33	**243**
Sn	49	1	2	120	**172**
Ob	54	77	15	21	**167**
Pr	15	47	5	11	**78**
Cm	2	0	0	4	**6**
Ab	3	97	4	9	**113**
Vl	80	167	0	1	**248**
NA	36	6	0	760	**802**
Total	913	424	48	992	2377

Table 8: Modality semantic meanings and scope types

Ambiguity accounts for 37% and 5% of the valid modality lemmas for lexical and semantic ambiguity, respectively. Some of the most frequent lemmas from each ambiguity type are illustrated in Tables 9 and 10.

Lemma	Trans.	Modal Freq.	Non-Modal Freq.
	qAl	171	40
	nfs	56	186
	$Af	80	129
فهم	*fhm*	88	45

Table 9: Top frequent lexically-ambiguous lemmas

Lemma	Trans.	Modality meanings	Freq.
	$Af	epistemic (think)	47
		sensory (watch; witness)	95
	Erf	epistemic (know)	41
		abilitive (can)	40
	mmkn	epistemic (possible that)	28
		abilitive (can)	17

Table 10: Top frequent semantically-ambiguous lemmas

Finally, about 82% of the scope heads are adjacent to their triggers. This is expected given that tweets are typically short.

6 Conclusion and Outlook

We presented a novel modality annotation scheme and applied it to the Arabic language in the tweets genre. This work is part of a larger project to use linguistic modality to detect power relations among participants on Twitter. The presented scheme uses both semantics and syntax to increase annotation reliability. Results show that Arabic modality triggers have regular subcategorization patterns that yield high annotation agreement when used as guidelines.

Currently, we are working on an updated version of this corpus with improved guidelines to tackle disagreement factors that emerged here. The new version will also include annotations for modality holders and trigger-related features such as negation, modification and mood.

Acknowledgement

References

Muhammad Abdul-Mageed and Mona Diab. 2012. AWATIF: A Multi-Genre Corpus for Modern Standard Arabic Subjectivity and Sentiment Analysis. *The 8th International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, 21-27 May 2012.

Rania Al-Sabbagh and Roxana Girju. 2012. YADAC: Yet another Dialectal Arabic Corpus. *The 8th International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, 21-27 May 2012.

Kathrin Baker, Michael Bloodgood, Bonnie Dorr, Nathaniel W. Filardo, Lori Levin, and Christine Piatko. 2010. A Modality Lexicon and its Use in Automatic Tagging. *The 7th International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, 19-21 May 2010.

Virginia Brennan. 1993. *Root and Epistemic Modal Auxiliary Verbs*, PhD Thesis, University of Massachusetts, Amherst.

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.

John Butler. 2003. A Minimalist Treatment of Modality. *Lingua*, vol. 113 (10): 967-996.

Joan L. Bybee, R. D. Perkins and W. Pagliuca. 1994. *The Evolution of Grammar: Tense, Aspect and Modality in the Languages of the World*. University of Chicago Press, Chicago, USA.

Mona Diab, Lori Levin, Teruko Mitamura, Owen Rambow, Vinodkumar Prabhakaran and Wei Wei Guo. 2009. Commited Belief Annotation and Tagging. *Proceedings of the Third Linguistic Annotation Workshop (ACL-IJCNLP '09)*, August 2009, Suntec, Singapore.

Iris Hendrickx, Amália Mendes, and Silvia Mencarelli. 2010. Modality in Text: A Proposal for Corpus Annotation. *The 7th International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, 19-21 May 2010

Ray Jackendoff. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge, MA: The MIT Press.

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. *The 7th International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, 19-21 May 2010

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 Diaglossic Speech in Cairene Arabic. PhD Thesis. University of Illinois at Urbana-Champaign.

Frank R. Palmer. 1989. *Mood and Modality*. Cambridge University Press, Cambridge, UK.

Frank R. Palmer. 2001. *Mood and Modality*. 2nd Edition. Cambridge University Press, Cambridge, UK.

Vinodkumar Prabhakaran. 2012. Detecting Power Relations from Written Dialog. *Proceedings of the 2012 Student Research Workshop*, Jeju, Republic of Korea.

Vinodkumar Prabhakaran, Michael Bloodgood, Mona Diab, Bonnie Dorr, Lori Levin, Christine D. Piatko, Owen Rambow and Benjamin Van Durme. 2012. Statistical Modality Tagging from Rule-Based Annotations and Crowdsourcing. *Proceedings of the ACL-2012 Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics (ExProM-2012)*, Jeju, Republic of Korea.

Roser Saurí, March Verhagen and James Pustejovsky. 2006. Annotating and Recognizing Event Modality in Text. *The 19th International FLAIRS Conference*, pages: 335-339, Florida, USA, May 2006

György Szarvas, Veronika Vincze, Richárd Farkas, and János Csirik. 2008. The BioScope Corpus: Annotation for Negation, Uncertainty and their Scope in Biomedical Texts. *BioNLP 2008: Current Trends in Biomedical Natural Language Processing*, pages: 38-45, Columbus, Ohio, USA, June 2008

Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating Expressions of Opinions and Emotions in Language. *Language Resources and Evaluations*, 39:165-210