Annotating opinions and emotions in text

Specifications and detailed guidelines

Version 0.8

Giouli, Voula

December 2023

History of the document

27/03/2012 version 0.1
19/04/2012 version 0.2
24/04/2012 version 0.3
05/10/2012 version 0.4
26/02/2013 version 0.5[1]
22/07/2013 version 0.6
20/11/2013 version 0.7[2]
18/12/2023 version 0.8[3]

---

1 Annotation is applied separately on the distinct channels and modalities (text, audio, video). Specifications and guidelines presented in this document may be equally applied to written texts and dialogues, the assumption being that the computational treatment of emotion in multi-modal data relies on the processing of each modality as separate yet inter-related channel.

[2] Annotation of emotion/opinion event participants within the sentence is outlined.

[3] Updates wrt to GenAI and new trends

Table of Contents

1. Introduction

The rapid growth of Web 2.0 platforms has led to an explosion of user-generated content across blogs, social media, forums, and other online spaces. This vast amount of data contains valuable insights, especially in the form of *subjective language*, which includes *opinions*, *emotions*, and personal experiences. Processing and identifying subjective language is crucial for various applications, from market research to social science studies, and it provides significant benefits across multiple domains. By processing subjective data, we unlock deeper understandings of human emotions, opinions, and experiences. This allows for more informed decision-making, enhanced user experiences, and more targeted and effective strategies across various sectors. While challenges such as ambiguity, linguistic variation, and evolving language persist, advances in natural language processing (NLP) and machine learning are continuously improving our ability to analyze subjective data accurately. Embracing these technologies enables us to harness the full potential of the vast amounts of user-generated content available today.

This document presents detailed guidelines for annotating *opinion* and *emotion* in textual and multimodal data. More specifically, the annotation task aims at (a) identifying subjective sentences in textual data (as opposed to factive or objective ones) and (b) classifying them based on their (perceived) meaning within a predefined typology that builds on the notions of opinion and emotion.

The document is structured as follows: in section (2) the rationale and usefulness of the annotation are outlined; in section (3), the necessary definitions of the terminology used are provided, whereas section (4) presents an overview of the methodologies employed for the tasks at hand. Previous work that is viewed as background of this work

2. Rationale

Processing and identifying subjective language in textual or multimodal data is a critical task with wide-ranging applications. From business intelligence and social media monitoring to political and social research, healthcare and entertainment, sentiment and emotion analysis provide valuable insights that drive decision-making and improve user experiences across multiple domains. These may be summarized as follows:

Market Research and Consumer Insights / Business Intelligence:

- **Product Reviews and Feedback:** Companies mine and analyze customer reviews to understand consumer sentiment about their products or services. By identifying positive and negative feedback, businesses can improve their offerings, boost customer satisfaction, and refine their marketing strategies and improve customer engagement.
- **Brand Perception:** Social media monitoring enables businesses to gauge how the public perceives their brand. This information is vital for reputation management and strategic decision-making.
- **Competitive Analysis:** Monitoring customers' sentiment about their competitors helps companies in identifying strengths, weaknesses, opportunities, and threats in the market.

Customer Service:

- **Sentiment Analysis in Feedback:** Customer service teams can use sentiment analysis to prioritize and address customer complaints more effectively. Identifying negative sentiments on specific aspects of their products/services allows for prompt resolution and enhances customer experience.
- **Chatbots and Automated Responses:** Integrating sentiment analysis into customer service chatbots enables more empathetic and contextually appropriate responses, enhancing user interaction.

Social Media Monitoring:

- **Trend Analysis:** Identifying trending topics and sentiments on social media platforms helps organizations stay ahead of public interests and concerns. This is especially important during crises or for timely marketing campaigns.
- **Influencer Identification:** By analyzing subjective language, businesses can identify influential users who shape public opinion, enabling targeted and more effective influencer marketing strategies.

Political and Social Research:

- **Public Opinion:** Analyzing subjective data from social media, blogs, and forums provides deep insights into public opinion on political issues, policies, and candidates. This information can influence or even guide campaign strategies and policy-making.
- **Social Movements:** Monitoring sentiments related to social movements helps researchers and policymakers understand public support and dynamics behind various causes.
- **Policy Making:** Sentiment analysis of public opinion on policies aids in shaping decisions that align with citizens' preferences and concerns.
- **Election Campaigns:** Analyzing voter sentiment helps political campaigns strategize and communicate more effectively.

Healthcare:

- **Mental Health Monitoring:** Analyzing language used by patients in online forums, social media, or electronic health records can reveal early signs of mental health issues such as depression or anxiety, allowing for timely intervention and support.
- **Patient Feedback:** Understanding patient sentiment from feedback on medical services helps healthcare providers enhance the quality of care and patient satisfaction.

Media and Entertainment:

- **Content Recommendation:** Understanding viewer sentiment towards different types of content allows media companies to personalize recommendations and enhance user experience.
- **Audience Engagement:** Identifying emotions in user comments and feedback helps in creating more engaging and relevant content.

Despite challenges such as ambiguity, linguistic variation, and evolving language, advanced techniques in NLP and machine learning continue to enhance our ability to accurately analyze subjective data. By leveraging these technologies, we can unlock the full potential of user-generated content and gain deeper understanding of human emotions and opinions.

3. *Definitions*

Prior to outlining the annotation guidelines, definitions of the terminology used are in order. Note that sometimes, the notions of sentiment and opinion are used interchangeably.

Subjectivity. Some sentences express mere facts. For example, the sentence in (1) is true in the world we live. In this case, the knowledge is seen to be independent of the subject who is producing it. This is not the case with sentence (2), where the perspective of an individual human subject is highlighted. In this regard, subjectivity refers to beliefs, biases, experiences, judgements, stance and opinions of the locutors.

(1) Athens is the capital of Greece.

(2) John feels sad.

Factivity. Contrary to subjective utterances, some snippets of text express a fact or ground truth. Therefore, factivity is defined as the opposite of subjectivity, that is, an utterance or piece of text that expresses a fact.

Sentiment. Sentiment refers to the expression of feelings or attitudes toward a particular subject or entity. Sentiments are often categorized as positive, negative, or neutral. Sentiment analysis aims to identify and classify the emotional tone behind a body of text.

Emotion. Emotion refers to a complex psychological state that involves a subjective experience, a physiological response, and a *behavioral* or expressive response. Emotions can be categorized into basic types such as *happiness*, *sadness*, *anger*, *fear*, *surprise*, and *disgust*. See (Γιούλη, 2020 for a detailed overview of emotion typologies).

Private states. Private states are internal states that cannot be directly observed by others. They include emotions, beliefs, and other forms of subjective experience. See Quirk

Opinion. An opinion is a subjective statement or thought about an issue or topic. Opinions are often based on personal beliefs, feelings, and preferences.

Stance. Stance refers to the attitude or position of the speaker or writer in relation to a particular issue. It can be identified through expressions of agreement, disagreement, doubt, certainty, etc.

In the literature, similar terms are used, namely *Affect*, *Sentiment*.

4. Methods for identifying Subjective Language

4.1 Lexicon-Based Approaches:

- **Sentiment Lexicons:** Predefined lists of words associated with positive, negative, or neutral sentiments can be used to analyze text. Examples of such lexicons include SentiWordNet (ref) and AFINN (ref).
- **Emotion Lexicons:** Lists of words associated with specific emotions (e.g., joy, sadness, anger) help in identifying the emotional tone of the text.

4.2 Machine Learning Approaches:

- **Supervised Learning:** Training machine learning models on labeled datasets to classify sentiment and emotion. Techniques include support vector machines (SVM), logistic regression, and neural networks.
- **Deep Learning:** Advanced methods like convolutional neural networks (CNN) and recurrent neural networks (RNN) (including LSTM and GRU) are used to capture complex patterns and dependencies in the text.

4.3 Hybrid Approaches:

- **Combining Lexicon and Machine Learning:** Using lexicon-based features as inputs to machine learning models can improve performance by leveraging the strengths of both approaches.

5. Background

5.1. Theories of emotion and emotion typologies

Emotion has been studied in different fields and disciplines, ranging from philosophy (Spinoza 1675, James 1884), psychology (Plutchik 1962, Turner 1996) and linguistics (Wierzbicka 1996, Kövecses 2000), to biology (Darwin 1859) and computer science (Ortony et al. 1988, Picard 1995). And, although there is little or no consensus among different theories on the emotion classes, a small number of primary emotions are commonly assumed. These are often referred to as primary emotions (Plutchik, 1991), as opposed to other, secondary emotions which are perceived of as the mixtures of primary emotions.

Most theories of emotion treat recognition of a triggering cause event as an integral part of emotion (Descartes 1649, James 1884, Plutchik 1962, Wierzbicka 1996).
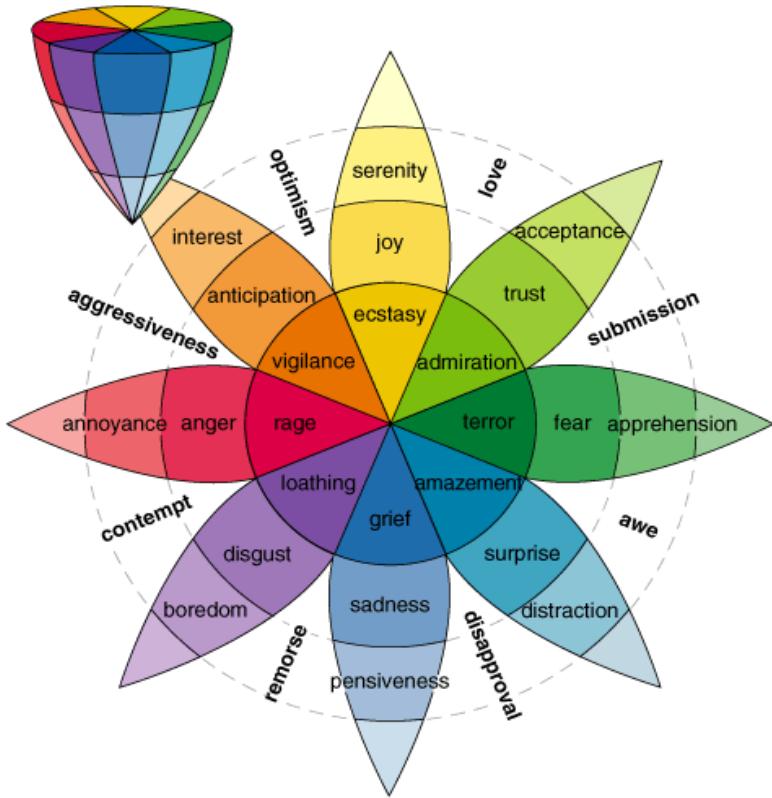
Figure 1. *Emotion wheel.*

5.2. Appraisal theory

Appraisal theory is a framework within the field of linguistics and psychology that focuses on the language of evaluation, attitude, and emotion. It provides a structured approach to understanding how individuals express their subjective feelings, judgments, and evaluations in discourse. Developed within the broader context of systemic functional linguistics, appraisal theory offers valuable insights into how language constructs interpersonal meaning and how speakers and writers position themselves and others in social interactions.

5.3. Corpus annotation

5.3.1. Subjectivity Annotation: identification of private states in text

The annotation scheme that we are proposing takes into account prior similar efforts. Background work has for the most part focused on sentiment classification, at the document, sentence or even phrase and word level. The MPQA corpus of news documentation (Wiebe et al., 2005; Wilson, 2008) defines attitudes as *private states* and proposes an annotation schema catering for the following conceptualizations or types of attitudes: *sentiment*, *agreement*, *arguing*, *intension*, and *speculation*. A general type, posited as *other attitude* is retained for all the remaining private states and a value of *positive* or *negative* is also assigned to the specific classes, as well as fine-grained intensity values. Expressive subjective elements, subjective speech events and explicit mentions of private states are annotated separately. Sources of

private states (Agents) and targets are also considered. Somasundaran & Wiebe (2010) explore further the *arguing* type as a means to investigate ideological stance. Opinion-target pairs are created, encoding also what the opinion is about, on the basis that opinions combined with targets are more informative than either of them in isolation.

Asher at al. (2009) have worked on a corpus compiled by movie reviews, letters to the editor and news reports to define a fine-grained annotation scheme that builds on the semantics of a wide class of opinion expressions at the sub-sentential level, the latter ultimately mapped onto a top-level typology of *reporting* (indicated by verbs)*, judgement* (that builds on the semantics of a wide class of opinion expressions at the sub-sentential level)*, advise* and *sentiment* expressions. This scheme is argued to be appropriate for calculating the overall opinion expressed in a text on a given topic.

The *Emotiblog* annotation model has been used on a corpus of various textual genres (news articles, news titles and a corpus of real-life self-expressed emotion) (Boldrini et al., 2010) and a corpus of blogs (Balahur et al., 2010) and distinguishes between objective and subjective speech. Polarity is assigned to adjectives/adverbs, verbs, nouns, anaphora and orthographic features. Interestingly, it takes into account two attributes (reader and author interpretation), annotating cases where apparently objective statements are used as indirect expressions of opinion.

As far as polarity is concerned, Polanyi & Zaenen (2006) examine how lexical valence is context-dependent and how valence shifters, such as negatives/intensifiers, modals, irony and various discourse structures influence the polarity and/or the strength of the opinion expressed. Furthermore, Neviarouskaya et al. (2010), based on the *Appraisal Theory* (Martin & White, 2005) present a scheme that includes polarity (positive, negative, neutral) on the top level, which is further divided into three types (affect, judgment and appreciation). Affect is further subdivided into 8 basic types. The authors propose an algorithm to decide how polarity is affected by a set of attitude-conveying terms, modifiers, functional words and modal operators. Using the *compositionality principle* the overall meaning of a sentence is determined.

Finally, the Blogoscopy corpus (Daille et al., 2011), containing personal thematic blogs, is annotated according to five types of evaluations: *opinion* (conviction/supposition), *appreciation*, *acceptance-refusal*, *agreement-discord* and *judgement.* Implicit and explicit cases are taken into account, as well as positive/negative polarity.

5.3.2. Beyond semantics: annotating pragmatic phenomena

Authentic texts usually make use of a wide range of *communicative* and *rhetorical devices.* To render the subjectivity annotation as complete as possible, pragmatic phenomena were also taken into account and *irony* was the first one to be annotated. Ironic/sarcastic utterances were also identified and marked as appropriate.

Irony is generally defined as a form of non-sincere speech, as a means to convey a meaning which is opposite or different to the literal one, and has been treated as a violation of the Gricean Maxims, principally of that of Quality (Alba Juez, 1995). According to the Maxim of Relevance, listeners attempt to interpret non-explicitly relevant utterances in a manner that fulfils the expectation of relevance and are thus able to recognize the ironic dimension in speech. From another perspective, irony has been proven to function in both a positive and a negative way. In Alba Juez (1995) two main kinds of irony were proposed: Positive Irony (intended to praise) and Negative Irony (intended to criticize).

6. The annotation scheme

Subjectivity annotation is seen as a three-step process: (a) identification of expressions that are not factive, (b) classification of those expressions in one of the types of subjective expression, and (c) annotation of the participants to the subjective event. In this respect, our annotation exercise is defined as *Subjectivity Annotation*, that is, identification and classification of subjective events and annotation of the respective event participants.

In this section we will elaborate further on the annotation scheme that we propose. The scheme provides for the identification of two broad categories of subjective discourse: (a) *emotion*, which refers to the psychological state of a *speaker* or an **agent** towards somebody or something - usually based on feeling or sentiment rather than reasoning; and (b) *opinion*, that is an expression of attitude, speculation, belief, thought, etc. Existing annotation afforts were taken into account **with certain modifications** – where needed - so as to accommodate emotion and opinion in **both oral and written discourse,** the ultimate goal being to provide an annotation scheme that can be employed to both written texts and oral discourse (dialogues).

The scheme, therefore, comprises two basic elements, namely, *emotion* and *opinion.* Yet, it also considers a more fine-grained classification of opinion and sentiment which are encoded as feature-value pairs of the aforementioned elements. We have opted for a modular approach to opinion/emotion classification, in the sense that it is applied across three axes: (a) **polarity**, (b) opinion/emotion **type** and (c) **intensity**. During manual annotation, values pertaining to all three axes are encoded; on the contrary, during analysis, classification might be targeted at any of the three axes/aspects above separately or combinations thereof.

6.1. Mandatory Features of elements *emotion* and *opinion*

Some elements in the annotation schema are mandatory, whereas others are optional. For an annotation to be complete, all mandatory features need to be completed. We elaborate on all features in the paragraphs below.

6.1.1. Feature *polarity*

Polarity of sentiment/opinion was also assigned to the selected text spans (being either sentences/clauses or phrases/words) as a mandatory feature assuming one of the following values: *positive*, *negative*, *neutral*, and *uncertain*.

6.1.2. Feature *type*

The two broad categories or basic elements provided for by our annotation scheme are further grounded with respect to a fine-grained classification scheme. Two distinct sets of values specified *Opinion* and *Emotion* classes are assigned to the feature *Type.* These are presented in the next paragraphs.

Emotion classification

Expressions denoting an emotion are assigned a feature *type* on the basis of a classification that is centred around a set of 8 basic sentiments (Plutchik, 1991): *anger*, *fear*, *sadness*, *disgust*, *surprise*, *anticipation*, *acceptance*, *joy* and *other*.

Opinion classification

The following types of opinion are encoded: *argument, evaluation, belief, recommendation, intention* and *other*. More precisely:

Argument. Under the umbrella term *argument,* we classify expressions denoting the point of view of the speaker, of what he believes to be true, possibly used to convince others or to make a statement.

Evaluation. An evaluation is a more precise type of belief and is specified as an estimation of the value of a person, object, action, etc., an assessment of behaviour or of phenomena, and involves both ethic and aesthetic values. Normally, predicative constructions are used to express evaluation, as in "John is a good student", "He's a good guy", "*The gangs are like a plague*. In dialogues, elliptical sentences are also used: "You, idiot!". Other expressions denoting evaluation: "Alexander found in Derek his shining prince" (=Derek is a shining prince for Alexander).

Intention. Additionally, *intentions* encompass aims, plans and other overt expressions denoting an act or instance of determining mentally upon some action or result. Similarly, intentions are normally introduced with "I intend to", "I am going to", "I plan to", "It is in my plans", "Future work involves / will be", "I am (not) ready to", etc. If any of the above is included in the sentence, then the feature inferred is not added to the markable. Attention should be paid, however, in discriminating senses, as for example in the case of the examples below:

(3) "I am going to leave tomorrow" (intention)

(4) "Your rhetoric and your propaganda aren't going to save you out there" (argument).

Recommendation. *Recommendations* are defined as expressions denoting a proposal for an appropriate course of action intending to advise or urge the interlocutor to take or to avoid an action. Recommendations are normally introduced with "I recommend", "should", "have to", "ought to", "need to" or imperatives. If any of the above is included in the sentence, then the feature inferred is not added to the markable. Note, however, the difference between: "You should avoid unhealthy foods" and "I should expel you" (argument)??.

Speculations are normally introduced with conditional clauses, one of the verbs "I suppose", "I assume", "I suspect", etc., or they contain adverbs and adjectival phrases of the type "supposedly", "possibly", "probably", "it is possible", "it is probable", etc.

Finally, a type ***opinion-other*** has been provided for encoding utterances that do not fall in any of the types mentioned above.

Note, however, that the listing of the verbs and phrases above is only indicative, and although they normally signal objective language as opposed to language describing facts, sense discrimination should always be based on context.

6.1.3. Feature *Strength*

Emotions and opinions were further assigned the mandatory feature of *strength* the possible values of which are: *low, medium, high.* An extra value *uncertain* was also provided for, in order to make annotators assign a value only if they are sure, leaving difficult or ambiguous cases for future treatment. The default value for this feature is *medium.* Lexical choices and/or paralinguistic features (especially for emotions) denote the strength of the opinion or emotion and are therefore taken into account during the annotation process.

(5) His brother *probably* put him into it

The presence of the adverb *probably*

(6) I think the street would kill you.

The annotation schema for elements emotion and opinion depicting mandatory features and their possible values is depicted schematically in Figure 1 and Figure 2 respectively.
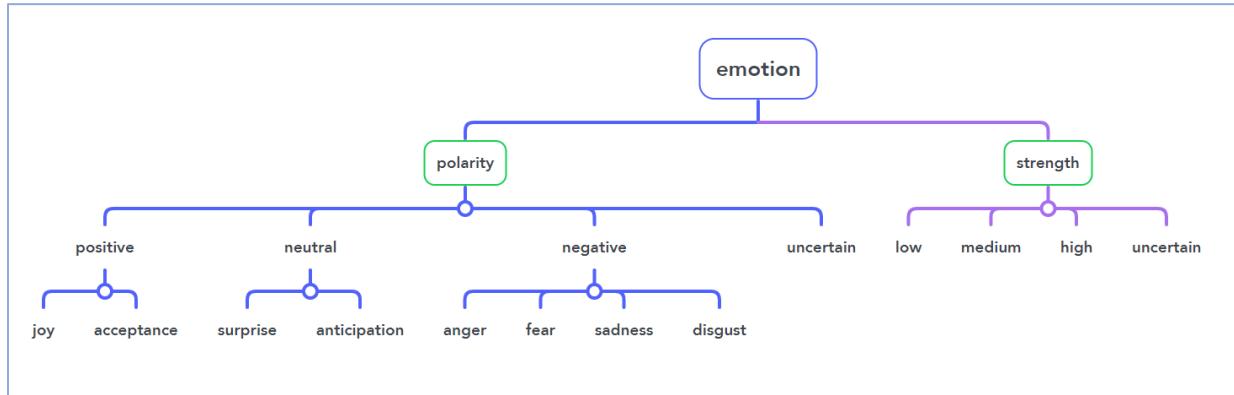
Figure 2. *Emotion annotation schema – mandatory features*
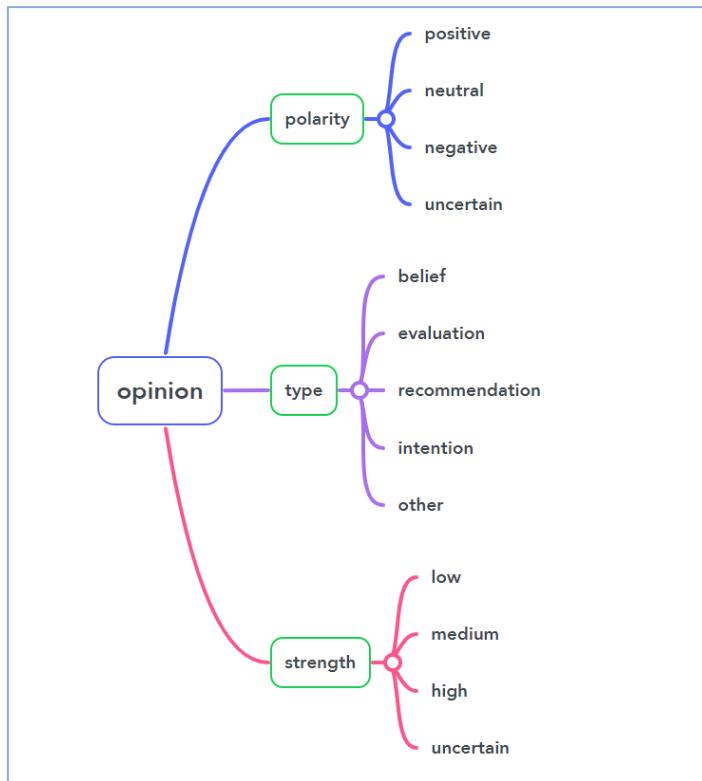

Figure 2. *Opinion Annotation Schema – Mandatory features*

6.2.   Optional features of the elements *emotion* and *opinion*

Additionally, three optional features are also proposed in our schema, namely: *inferred, repetition* and *sarcasm*.

6.2.1. Feature *inferred*

Speakers express their own or other persons' emotional states, opinions, evaluations, etc.,

either **explicitly** or **implicitly**. The feature *inferred* has been employed to distinguish utterances for which the opinion or emotion values are assigned on the basis of **non-verbal evidence** or paralinguistic cues. Possible values for this feature are: *audio, video, audio-video, text, that* is, the modality contributing to the interpretation of the utterance.

For example, in utterances (7) - (9) below, the speaker's emotional state is expressed directly, using an emotion expression:

    (7) Danny, no, <emotion>I feel sorry for you, Danny </emotion>.

    (8) <emotion> I hate anyone that isn't white Protestant</emotion>.

    (9) <emotion> I am angry</emotion> all the time.

Opinion expressions introduced with lexical (believe, think, should, etc.) or syntactic cues (i.e., imperatives) as outlined above are considered to convey the underlying meaning in a direct and explicit manner. In (11), the speaker explicitly expresses his belief or speculation, the correct interpretation of which is based on the modal *wouldn't*, whereas in (10) and (12), the verb "*think*" further reinforces the interpretation of the utterance as an opinionated one:

    (10)   I think I know why {opinion, type= "argument", polarity="negative"}

    (11)   Derek wouldn't let us visit him in prison {opinion, type="speculation", polarity="negative"}

    (12)   I think the street would kill you {opinion, type= "argument", polarity="negative"}

However, emotional states or opinionated discourse may be expressed implicitly as well. Interlocutors usually make use of implicit lexical choices to express their attitudes, as for example in utterance (23), or they make use of paralinguistic cues to express their emotional states, as in (14). To accommodate these cases, the feature *inferred* subsuming one of the possible values "*audio*", "*video*", "*audio-video*" "*text*" has also been added in order to retain information on the modality that contributes to the correct interpretation of the utterance.

    (13)   Good to have you back {emotion type="joy", inferred="text"}.

    (14)   Danny! {emotion type="anger", inferred="audio"} Shut the door! {emotion type="anger", inferred="audio"}

The value "text" is the default value and is used if contextual evidence is taken into account in order to distinguish opinionated sentences from factual ones, as in utterance (14). Note that the other three values, namely, "*audio*", "*video*", "*audio-video*" are used only when the other two modalities are used.

The identification of implicitly expressed emotions is not a trivial task. This task becomes even harder in cases of overt opinions which also convey the speaker's emotional load. In this work, we have attempted to annotate utterances which are opinionated and at the same time express the emotional state of the speaker (see also section 7.2):

    (15)   You hate this child (= I assert that you hate this child) {opinion, inferred="text", type= "argument", polarity="negative"}

    (16)   This is typical. (opinion-evaluation, emotion-sadness)

(17)   This country is becoming a heaven for criminals. (opinion-belief, emotion-anger)

(18)   This isn't fair! (opinion-evaluation, emotion-anger)

(19)   They are the fucking enemy. (opinion-evaluation, emotion-disgust).[4]

6.2.2. Feature *repetition*

Additionally, the feature *repetition* has also been used to mark cases where the repetition of an utterance or part of it (on word, syntax or phrase level) is used to express the strength of an emotion or opinion (that is *set to yes if true, otherwise it is assigned the value no).*

6.2.3. Feature *sarcasm*

The annotation scheme that we have developed, takes this double classification into account, and irony is encoded as being either <positive> or <negative>:

(20)   I'm not gonna stay in my room the rest of my life, you know. {opinion, type="intention", irony="negative"}

(21)   This country is becoming a heaven for criminals. {opinion, type="evaluation", irony="negative"}

(22)   Alexander found in Derek his shining prince {opinion, type="evaluation", irony="negative"}

(23)   What's the matter, afraid you're going to get a B? (=You are an excellent student, you are not going to get a B) {opinion, type="evaluation", irony="negative"}

(24)   Give yourself a raise, will you? (while depreciating the work of his colleague)

(25)   Hey, that's a great color on you, you know? Now you can get a white woman's job, bitch.

Annotation of irony in the corpus must be performed on the basis of contextual and/or world and situation-specific knowledge. Moreover, since our data involves the oral modality, identification of ironic utterances was also aided by acoustic features. On the basis of the assumption that speakers provide prosodic disambiguation cues when using verbal irony and that listeners use prosodic information, in addition to context information, to interpret ironic utterances (Bryant & Fox Tree, 2002), intonation was also used as a cue for disambiguation.

6.3.   Annotation of subjective event participants

Participants of the emotion/opinion event are identified and annotated. The source or the entity experiencing the emotion event is one of the most salient entities. Similarly, the emotion Target, that is the entity towards which the emotion is directed is also salient in the event. Since our goal is the construction of large lexica, we are also interested in the syntactic configurations used as a mechanism to denote the participants in the respective event. The annotation schema provides for the following event participants:

---

[4] We need to check these cases again.

Source/Experiencer/Cogniser[5]: The entity that experiences the emotion referred to in the text, or the entity that expresses their opinion

(26)   ***Η κυβέρνηση Μπους*** **ανησυχεί** για την επέκταση στη Λατινική Αμερική ενός ριζοσπαστικού μπλοκ υπό την ηγεσία του Τσάβες.

(27)   ***Με*** απογοήτευσε η πολλοστή επιβεβαίωση ότι δεν είμαι εκλεκτός της τύχης

Target: The entity the emotion/opinion is targeted at:

(28)   Η Θάτσερ **αντιπαθούσε** <u>*τον Παπανδρέου*</u> γιατί ήταν αντίθετοι πολιτικά

Cause: The event or entity that triggers the emotion

(29)   Η κυβέρνηση Μπους **ανησυχεί** <u>*για την επέκταση στη Λατινική Αμερική ενός ριζοσπαστικού μπλοκ υπό την ηγεσία του Τσάβες*</u>

6.4.   Nested vs entailed opinions and emotions

Finally, the annotation specifications allow the annotation of nested opinions and emotions, as shown in the following examples. The following utterance, in (30) also includes at least one more opinion/evaluation, based on the entailment depicted in (31). Notice, however, that no similar interpretation can be attempted for social parasites, which is therefore, not marked as an opinionated clause.

(30)   Decent, hardworking Americans, like my dad are getting rubbed out by social parasites

(31)   My dad was a decent, hardworking American

Two markables are identified:

Decent, hardworking Americans, like my dad are getting rubbed out by social parasites {opinion, type= "argument", polarity="negative", intensity="high"}

Decent hardworking Americans, like my dad = My dad was a decent, hardworking American {opinion, type="evaluation", polarity="positive"}.

(32)   I know you don't believe any of this shit, right?

I know you don't believe any of this shit {opinion, type= "argument", polarity="negative", intensity="high"}

this shit {opinion, type= "evaluation", polarity="negative", intensity="high"}

(33)   But <opinion-belief> if a white person sees two black men walking towards her, and she turns and walks in the other direction, <opinion-evaluation> she's a racist </opinion-evaluation> </opinion-belief>, right?

(34)   <opinion-belief> I know you don't believe any of <opinion-evaluation> this

---

[5] In our initial pilot annotations, we used the value Source for this attribute. We need to discuss and consolidate. The problem is that sometimes the author/speaker of the utterance is the experiencer or source of the emotion; in this case, the source/experiencer is not referred to in the text. In other cases, the

shit </opinion-evaluation></opinion-belief>, right?

(35)   He learned these nonsense, and he can unlearn it.

In this utterance, the speaker expresses his belief that the protagonist acquired some kind of knowledge, and at the same time, he makes an evaluation regarding the quality of what was learned. This

(36)   He learned these <opinion polarity=negative type=evaluation inferred=yes>

nonsense/opinion>

(37)   This isn't a way to talk to a guest, a friend of the family.

However, the issue in this case is how to compute the overall opinion value (at the sentence level), i.e., on the basis of a compositional technique. In other words, the question here is whether the value of the whole is the union of the values of its parts.

(38)   <negative>Fucking pigs</negative> <negative>telling us we should

<positive>get along</positive></negative>

For the time being, annotators judge intuitively the overall polarity value of the utterance. In the future, we should check whether we could define some hierarchy and an algorithm for computing the overall polarity of the clause. At the level of annotation, we should define levels of annotation: word, phrase and sentence.

6.5.   **Ambiguous cases**

In some cases, the same utterance seems to be both opinionated and at the same time express the emotional state of the speaker/writer.

(39)   This is typical (opinion – evaluation, emotion – sadness)

(40)   This country is becoming a heaven for criminals (opinion – evaluation, emotion – anger, irony=yes)

(41)   This isn't fair!

(42)   This is crap. It has been annotated as opinion-evaluation.

(43)   You are saying that you think maybe your father's murder was race-related? ⮞ opinion (= I doubt that your father's murder was race-related) and emotion – surprise????

(44)   Looks good. ⮞ Opinion – positive

(45)   Have a little self-respect! {opinion, polarity=negative, intensity=high type=evaluation inferred=text}, {emotion polarity=negative intensity=high type=disgust inferred=text}

It has been annotated as denoting both an emotion and an opinion. On the one hand, this utterance conveys the opinion of the speaker that the listener has no self-respect. At the same

time, the speaker also expresses a negative emotion/attitude towards his conversant, and this is identified on the basis of context, and world knowledge??

Equally problematic are utterances which express the speaker's opinion on a topic, and at the same time, they contain an emotion expression

(46)   You *hate* this child.

The speaker believes that his conversant hates the child. To this end, the whole sentence should be annotated as {opinion, type= "argument", polarity="negative", intensity="medium"}. What about the emotion expression? One solution would be to annotate it, but do not take it into account since it co-exists with an opinion. Should we devise a new feature for this type of elements (conditional emotions?)

6.6. Special cases involving certain syntactic constructions. Sentiment analysis and Speech Act Theory

Special attention needs to be paid to sentences or utterances that are not factive or declarative but which denote a speech act. Note that the annotation of speech acts in corpora is useful for various NLP applications. We detail some cases here, but final decisions will not be made now.

6.6.1. Utterances with *want – need*

Are they emotions or opinions? Is desire an emotion?

(47)   I want him out of my house

(48)   What do you want me to do it on?

(49)   (I want you to do it on) Your brother.

(50)   I want you to analyze and interpret all the events surrounding Derek's incarceration.

6.6.2. Questions

We interpret/analyze questions on the basis of their underlying meaning.

(51)   Derek, can we please go eat? = I want us to go for dinner (is it an opinion or emotion?)

(52)   Why are you tripping on me? (=I think that you are tripping on me and that you are doing so for some reason) – we have to agree on the extend of the markable

(53)   Who gives a shit? (= Nobody gives a shit)

(54)   How do you think I feel? (=I feel as expected good/bad)

(55)   What are you scared of? (= I think that you are scared of something)

But what about

(56)   Are you scared?

6.6.3. Imperatives

Imperatives are normally annotated as recommendations. But what about:

    (57)   But don't insult my intelligence, Murray!

Is it a recommendation or an argument? (=I think that you insult my intelligence, Murray.)

6.6.4. Other types of utterances/sentences/phrases

Similarly, other types of utterances, such as *greetings*, comparative structures (comparisons), discourse markers and their contribution to the opinion or emotions denoted, interjections, expressions of agreement or disagreement, etc. need to be studied further.

    (58)   Derek, if you could come down here please for just a second.

    (59)   This place is smaller than the old living-room.

    (60)   I don't understand (used as n interjection)

These are left aside for future study. Annotation of speech acts in dialogues is also useful for a number of applications.

7. Challenges in Identifying Subjective Language

7.1. Multiword expressions

See Fotopoulou and Giouli (2015).

7.2. Ambiguity and Context Dependence

Sometimes, the context is not enough to disambiguate the utterance.

    (61)   Honours English. → opinion, type="evaluation"?

- **Polysemy:** Words or phrases can have multiple meanings depending on the context. Disambiguating these meanings is crucial for accurate sentiment analysis.
- **Sarcasm and Irony:** Subjective language often includes sarcasm or irony, making it challenging to detect the true sentiment without understanding the context.

Other cases of ambiguity:

Comparatives seem to be problematic. The speaker thinks that the place is smaller. But smaller entails some type of appraisal.

7.3. Linguistic Variation

- **Slang and Informal Language:** User-generated content frequently contains slang, idioms, and informal language that standard natural language processing (NLP) tools might not recognize. However, this type of language seems to bear emotional
- **Multilingual Data:** Processing subjective language in multilingual datasets requires robust language-specific tools and techniques to maintain accuracy.

7.4. Evolving Language

- **Neologisms:** The continuous evolution of language, with the creation of new words and phrases, especially on social media, necessitates constant updating of NLP models and lexicons.

7.5. Metaphors in discourse

(62)   We 've washed ourselves in niggers' blood and all the mongrels, too.

(63)   The white man marches on

7.6. Markable extent: the problem of "What to annotate?" or "How much to annotate?"

The purpose of this annotation task was to spot and classify text snippets that bear a meaning and at the same time have an emotional load or snippets that instead of expressing facts, denote an opinion or stance held by individual persons, teams, organizations, etc. Determining what to annotate and how much to annotate is a significant challenge in corpus annotation. The goal is to ensure that annotation captures all meaningful elements while maintaining clarity and consistency. At the same time, annotation is also intended to reveal the linguistic elements that contribute to the expression of sentiment, emotion and opinion.

Units beyond the word and phrase level that convey a complete meaning are considered markables. These units include phrases below the sentence level that bear distinct semantic loads and punctuation marks, which play a crucial role in the interpretation of the text. Deciding on the extent of these markables involves several considerations:

- Complete Semantic Units: Annotate phrases and clauses that convey complete thoughts or arguments. This ensures that the annotations capture the full meaning intended by the text.
- Punctuation: Punctuation such as commas, question marks, and exclamation points should be included if they contribute to the interpretation of the utterance. Punctuation marks within an utterance often affect its meaning and should be considered part of the markable.

In this regard, in (64), the phrase **what I think, this is something we got to keep an eye on** is marked as it conveys a complete thought, argument or opinion expressed. On the contrary, the connective "**now**" does not add to the semantic completeness of the utterance and remains unmarked:

(64)   Now, **what I think, this is something we got to keep an eye on**.

By carefully considering what to annotate and how much to annotate, we can ensure that the annotation captures all relevant linguistic information, providing a rich resource for linguistic analysis and NLP applications. The balance between thoroughness and clarity is crucial, as over-annotation can lead to confusion, while under-annotation might omit important contextual, structural or semantic information.

8. Corpus description

The annotation scheme described above was applied to two corpora. At the first stage, pilot annotation was performed on a corpus of movies transcripts in English (EN) and their subtitles in Greek (EL) and Spanish (ES) (Mouka et al. 2012). The movies were selected according to external and internal criteria: (a) topic, i.e., inter-racial relations; (b) time, i.e., contemporaneity of production and reference; (c) realistic approach to events; and (d) their content (assumed racist discourse). The corpus comprises five (5) movies with a total playtime of 09:05 hours of quasi-spontaneous oral speech. The EN audio-visual material was transcribed and segmented, and utterances were synchronized (time aligned) with the movie audio. The EN transcripts were also aligned with EL and ES subtitles obtained from the official distribution of the movies. The subtitle material is a specialized type of translation corpus, in the sense that subtitling conforms to certain time and space restrictions. Following best practices for corpus development, video segmentation and transcription were performed using ELAN (Brugman & Russell, 2004), to ensure conformity with standards for audio-visual material processing, and thus re-usability of resources. The textual material was segmented at the utterance level, guided by intonation and pause clues. To further facilitate the alignment of EN transcripts with the EL and ES subtitles, segmentation at longer stretches of speech was performed at the level of C-Units; the latter follow written discourse conventions and typically comprise short sentences. Repetitions, hesitations, repairs and overlapping utterances that are inherent in oral discourse are retained in the corpus – yet they were not considered while annotating for sentiment and opinion. The pilot annotation was meant to test whether the schema could be applied to languages other than Greek.

Annotation proper was then performed on a corpus of Modern Greek (Giouli and Fotopoulou, 2013; Giouli, 2020). The Greek Sentiment Corpus comprises texts selected manually from various sources over the web. More particularly, the online edition of two newspapers along with a news portal were searched daily for the identification and selection of news articles dealing with and commenting on a set of predefined topics. The so-retrieved volume of texts comprises the newswire counterpart of our corpus. Additionally, letters to the editor and other opinionated articles focusing on the current affairs (as to the time of corpus compilation) were also collected from these online newspapers' editions. Similarly, blogs and forums rated as popular or influential were also used as sources for text collection. All the texts collected adhere to the sane topics (Greek financial crisis, international conflicts) and sub-topics (strikes, health, education, Arab spring, etc.).

Annotation was performed using GATE platform (Cunningham et al., 2002) on semi-automatically tokenised and POS-tagged data using a Greek POS-tagger (Papageorgiou et al., 2000).

9. Conclusions

We have presented ongoing work aimed at the development of a scheme for sentiment and opinion annotation on text and preliminary guidelines for applying it to textual data. The goal is to come up with a fine-grained annotation schema for sentiment, emotion and opinion. The schema has been applied to textual data and inter-annotator agreement has been calculated.

Future work has already been planned towards the following axes: (a) consolidation of the annotation guidelines, (b) elaboration of strict guidelines to overcome difficult cases and enhance reproducibility of the annotation, (c) addition of examples and counterexamples to make the annotation guidelines complete, and (d) corpus enrichment. At another level, future

work will be undertaken towards annotating Speech Acts in relation to sentiment and emotion.

References

Alba Juez, L. (1995). Irony and Politeness. *Revista Española de Lingüística Aplicada*, 10, pp 9-16.

Asher, N., Benamara, F., Mathieu, Y.Y (2009). Appraisal of opinion expressions in discourse. *Lingvisticae Investigationes*, 32(2).

Balahur, A., Boldrini, E., Montoyo, A., Martinez-Barco. P. (2010). IBEREVAL OM: Mining Opinions from the new textual genres. Procesamiento del Languaje, 45, pp. 267-271.

Boldrini, E., Balahur, A., Martínez-Barco, P., Montoyo, A. (2010). EmotiBlog: a finer-grained and more precise learning of subjectivity expression models. In *Proceedings of the 4th Linguistic Annotation Workshop (LAW IV), satellite workshop to ACL 2010*. Uppsala University.

Brugman, H., A. Russell (2004). Annotating Multimedia / Multi-modal resources with ELAN . In: Proceedings of LREC 2004, 4th International Conference on Language Resources and Evaluation (LREC 2004), Lisbon.

Bryant, G.A., & Fox Tree, J.E. (2002). Recognizing verbal irony in spontaneous speech. *Metaphor and Symbol*, 17(2), pp. 99-117.

Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V. (2002). GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02).* Philadelphia.

D. Das et al. 2010. Labelling Emotion in Bengali Blog Corpus – A Fine Grained Tagging at Sentence Level. 8th Workshop on Asian Language Resources.

Daille, B., Dubreil, E., Monceaux, L., Vernier, M. (2011). Annotating opinion - evaluation of blogs: the Blogoscopy corpus. *Language Resources and Evaluation*, 45(4), pp. 409-437.

Fotopoulou, A. and V. Giouli. 2018. MWEs and the Emotion Lexicon: Typological and cross-lingual considerations. In Manfred Sailer & Stella Markantonatou (eds.), Multiword expressions: Insights from a multi-lingual perspective, 63–91. Berlin: Language Science Press. DOI:10.5281/zenodo.1182591

Giouli, V. and Fotopoulou, A. (2014) Developing Language Resources for Sentiment Analysis in Greek. In G. Kotzoglou et al. (eds), *Selected Papers of the 11th International Conference on Greek Linguistics*, 501-511. Rhodes: University of the Aegean

Gao, W., Li, S., Lee, Y. M., Zhou, G., & Huang, C. (2013). Joint learning on sentiment and emotion classification. In *CIKM 2013 - Proceedings of the 22nd ACM International Conference on Information and Knowledge Management* (pp. 1505-1508) https://doi.org/10.1145/2505515.2507830

Lee, S.Y.M. , Chen, Y., Li, S., and Huang, C.R. 2013. Emotion Cause Events: Corpus Construction and Analysis. In *Proceedings of the 7th Language Resources and Evaluation Conference (LREC2010),* Malta.

Lee, S.Y.M. , Li, S. , and Huang, C.R. 2014. Annotating Events in an Emotion Corpus. In *Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014),* 26-31 May, Reykjavik, Iceland.

Lee. YM, et al. 2013. Detecting Emotion Causes with a Linguistic Rule-based Approach. *Computational Intelligence*.

Martin, J.R. & White, P.R.R. (2005). *The Language of Evaluation: Appraisal in English*. Palgrave, London, UK.

Mouka, E., Giouli, V., Fotopoulou, A., & Saridakis, I. E. (2012, May). Opinion and emotion in movies: a modular perspective to annotation. In *Proceedings of the 4th International Workshop on Corpora for Research on Emotion, Sentiment & Social Signals (ES^3 2012)*.

Neviarouskaya, A., Prendinger, H., Ishizuka, M. (2010). Recognition of affect, judgment, and appreciation in text. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*. Coling 2010 Organizing Committee, Beijing, China, pp. 806-814.

Papageorgiou, H., Prokopidis, P., Giouli, V., Piperidis, S. 2000. A unified taggingArchitecture and its Application to Greek. In *Proceedings of the 2nd International Conference on Language Resources and Evaluation*, Athens, Greece.

Plutchik, R. (1991). *The Emotions*. New York, NY: University Press of America.

Polanyi, L., Zaenen, A. (2006). Contextual Valence Shifters. In J. Shanahan, Y. Qu, and J. Wiebe (eds.) Computing Attitude and Affect in Text: Theory and applications. *The Information Retrieval Series*, Vol. 20, Springer, Dordrecht, The Netherlands, pp.1-9.

Qadir, A., & Riloff, E. (2011). Classifying Sentences as Speech Acts in Message Board Posts. *Conference on Empirical Methods in Natural Language Processing*.

Ren F. and Quan. 2012. Linguistic-based emotion analysis and recognition for measuring consumer satisfaction: an application of affective computing. Information Technology and Management.

Somasundaran, S. & Wiebe, J. (2010). Recognizing Stances in Ideological On-Line Debates. In *NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*. Los Angeles.

Wiebe, J., Wilson, T., Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2-3), pp. 165-210.

Wilson, T. A. (2008). *Fine-grained Subjectivity and Sentiment Analysis: Recognizing the Intensity, Polarity, and Attitudes of Private States*. Ph.D. thesis, University of Pittsburgh.