Opinion and emotion in movies: a modular perspective to annotation

Mouka, E.[2], V. Giouli[1,2], A. Fotopoulou[1], I.E. Saridakis[2,3]

[1]Institute for Language and Speech Processing, R.C. "Athena", Greece;

[2]Ionian University, Greece; [3]National and Kapodistrian University of Athens, Greece;

f10mouk@dflti.ionio.gr, voula@ilsp.athena-innovation.gr, afotop@ilsp.athena-innovation.gr, iesaridakis@gmail.com

Abstract

This paper presents an ongoing effort work focusing on the development of an audiovisual corpus resource and its annotation in terms of sentiments and opinions. A modular annotation schema has been employed based on the specifications of existing schemas and extending or adapting them to cater for the peculiarities of the corpus-specific data.

Keywords: annotation of emotion, annotation of opinion, movies corpus

1. Introduction

This paper presents the first version of a new specialized audiovisual corpus resource that comprises movies coupled with both orthographic transcriptions in English [en] and their official subtitles in Greek [el] and Spanish [es]. The corpus resource bears annotations at various levels of analysis (word/phrase/sentence, and also on the audio) while the focus is on the identification of opinions and emotions in oral discourse, elaborating on specific semantic and pragmatic phenomena. Cross-language issues were considered as well as textual vs. audiovisual cues. We describe the specialized corpus focusing on the pilot annotation procedure, and the results of an inter-annotator agreement study.

The paper is organized as follows: Section 2 presents the scope and aims of the research undertaken, as well as the multimodal corpus in terms of its content and typology. Section 3 includes descriptions of the metadata. Section 4 presents an overview of related works. Section 5 includes a detailed overview of the annotation scheme employed and the methodology adopted in the current annotation work. Our preliminary findings are presented and discussed in Section 6. Finally, Section 7 includes our conclusions and prospects for future work.

2. Project scope and aims

The audiovisual data was initially selected in order to guide translation-oriented research examining the language options that depicts a specific type of biased opinionated ideological stance/attitudes, namely that of racist discourse, and its transfer from the textual source language (SL) to the target language(s) (TL) through subtitling. This type of discourse is socio-culturally marked (Waugh, 1982). This makes the corpus an excellent pool for annotating opinions, beliefs, thoughts, feelings, emotions, goals, evaluations, and judgments in oral discourse. In yet another aspect, the annotations were also oriented towards populating a lexical resource that is currently under development and contains opinion and emotion words with new entries adhering to oral data. The ultimate goal of this work, therefore, is to investigate the use of opinion and emotion expressions in oral discourse by means of a corpus annotation study that extends across modalities and languages.

Finally, this work which is still in progress may be integrated into a larger initiative undertaken by the Institute for Language and Speech Processing aimed at the development of a suite of language resources (corpora, lexica, tools) for sentiment analysis.

3. Corpus description

As a product of the so-called prefabricated orality (Baños-Piñero & Chaume, 2009), movies were selected according to external and internal criteria: (a) topic (centered around inter-racial relations); (b) time, i.e., contemporaneity of production and reference; (c) realistic approach to events; and (d) their content (assumed racist discourse). To date, the corpus comprises 5 movies with a total playtime of 09:05 hours of quasi-spontaneous oral speech. The [en] audio-visual material has been transcribed and segmented, and utterances have been synchronized (time aligned) with the movie audio. Finally, the transcripts were also aligned with "[el] and [es] subtitles 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.

More precisely, following standard procedures, so as to ensure conformity with standards for audio-visual material, and, thus re-usability of resources, video segmentation and transcription were performed using ELAN (Brugman & Russell, 2004). The spoken language transcription relied on the TEI specifications for (Schmidt, 2011; TEI Consortium, 2011). The segmentation was performed at the utterance level, following intonation and pause clues, while quite long stretches of speech were further segmented into C-Units to facilitate alignment with TL subtitles that follow written discourse conventions and typically comprise short sentences. Repetitions, hesitations, repairs and overlapping utterances that are inherent in oral discourse have been retained in the corpus. Each utterance is assigned a time slot and a speaker. The final output is a TEI-conformant .xml document. An example of the resulting representation is depicted in the Figure 1 below:

```
- <div>
- <u who="#SPK16">
  <anchor synch="#T835" />
     Who do you hate, Danny?
  <anchor synch="#T836" />
     </u>
- <spanGrp type="subtitles" lang el">
  <span from="#T835" to="#T836"> Ποιον μισείς,
     Ντάνι ;</span>
     </spanGrp>
- <spanGrp type="subtitles es">
    <span from="#T835" to="#T836">¿A quién
     odias, Danny?</span>
     </spanGrp>
     </div>
- <div>
```

Figure 1: transcribed text

The external structural annotation (including text classification) of the corpus also adheres to the IMDI metadata scheme (IMDI Team, 2003). IMDI metadata elements for catalogue descriptions (IMDI Team, 2009) were also taken into account to render the corpus, and adaptations proposed specifically concerning Multimodal Language Resources have been taken into account. This type of metadata descriptions was added via the ELAN interface and stored in XML format.

4. Opinion and emotion: background

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 attitude: *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 mention of private states are annotated separately. 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 Boloscopy 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 long as positive/negative polarity

5. Opinion Annotation in Movies

In this section we will elaborate further on the annotation schema employed that caters to the identification of two broad categories: (a) *emotion*, expressing 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, beliefs, thoughts, etc. The schema, therefore, comprises two basic elements, namely, *emotion* and *opinion*.

The schema also considers a more fine-grained classification of opinion and sentiment. *Emotion* classification is centred around a set of 8 basic sentiments (Plutchik, 1991): *anger*, *fear*, *sadness*, *disgust*, *surprise*, *anticipation*, *acceptance*, *joy* and *other*. Moreover, the following opinion classes are defined: *evaluation*, *belief*, *recommendation*, *intention* and *other*. More precisely, an evaluation 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. Under the umbrella term *belief* we classify expressions denoting the point of view of the speaker, of what he believes to be true, possibly used as an argument. Additionally, *intentions* encompass aims, plans and other

overt expressions of intention, while *recommendations* are further defined as expressions intenting to urge the interlocutor to take an action.

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*.

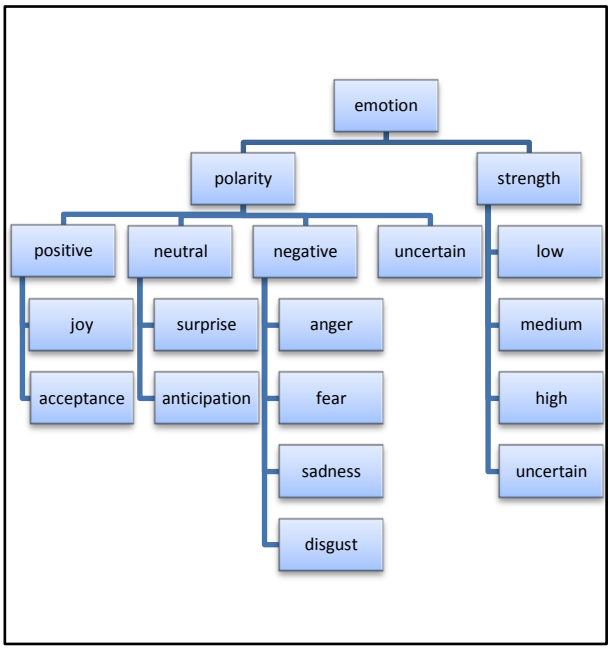

Figure 1: Emotion annotation schema

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

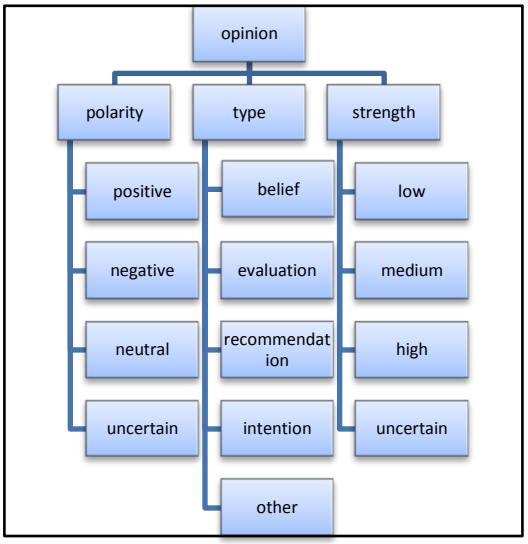

Figure 2: Opinion annotation Schema

ambiguous cases for future treatment. 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. 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.

Moreover, three more features that are optional are proposed in the schema, namely *inferred*, *repetition* and *irony*. 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*, *text*, that is, the modality contributing to the interpretation of the utterance. Additionally, the feature of *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).* The feature *irony* is used to encode pragmatic phenomena (see section 5.1).

The proposed schema builds on existing annotations of emotion and opinion in discourse with certain modifications that were deemed necessary so as to accommodate the peculiarities of the data at hand, namely phenomena that are inherent to oral discourse.

Speakers express their own or other persons' emotional states, opinions, evaluations, etc, either explicitly or implicitly. For example, in utterances (1) - (3) below, the speaker's emotional state is expressed directly, using an emotion expression:

(1) Danny, no, <emotion>I feel sorry for you, Danny </emotion>.
(2) <emotion> I hate anyone that isn't white Protestant</emotion>.
(3) <emotion> I am angry</emotion> all the time.

In (4) the speaker explicitly expresses his belief or speculation, the correct interpretation of which is based on the modal *wouldn't*, whereas in (5) the verb "*think*" further reinforces the interpretation of the utterance as an opinionated one:

(4) <opinion – speculation> Derek wouldn't let us visit him in prison </opinion – speculation>
(5) <opinion – speculation> I think the street would kill you</opinion – speculation>

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 utterances (6), or they make use of paralinguistic cues to express their emotional states, as in (7):

(6) <emotion-joy>
Good to have you back </emotion-joy>.
(7) Danny! Danny! <emotion-anger> Danny</emotion–anger>!
<emotion–anger>Shut the door</emotion–anger>!

To accommodate these cases, a further feature *inferred*

with possible values "*audio", "video", "context"* has also been added in order to retain information on the modality that contributes to the correct interpretation of the utterance (see above). The identification of implicitly expressed emotions is not a trivial task. In fact, this is even harder in cases of overt opinions which convey also the speaker's emotional load. In this work, however, we have attempted to annotate utterances which are opinionated and at the same time express the emotional state of the speaker:

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

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

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

(11) They are the fucking enemy. (opinion-evaluation, emotion-disgust)

Finally, the annotation specifications allow the annotation of nested opinions and emotions, as shown in the following examples:

(12) <opinion-belief>
<opinion-evaluation>Decent, hardworking Americans, like my dad </opinion-evaluation> are rubbed out by <opinion-evaluation> social parasites</opinion-evaluation>
</opinion-belief>.

(13) 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?

(14) I know <opinion-belief> you don't believe any of <opinion-evaluation> this shit </opinion-evaluation></opinion-belief>, right?

5.1 Annotating pragmatic phenomena

Movies comprising our corpus depict situations in which dialogue participants 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). The annotation scheme that we have developed, takes this double classification into account, however, only one instance of positive irony has been identified so far, and marked appropriately as "irony-positive", in (15) where irony is used to imply that the boy is an excellent student. Examples in (16), (17), and (18) are all cases of negative irony/sarcasm which clearly show how the literal lexical meaning is altered by irony

(15) What's the matter, afraid you're going to get a B?

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

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

(18) This country is becoming a *heaven* for criminals.

Annotation of irony in the corpus was 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.

5.2 Annotation methodology

After the initial specifications were formulated, annotations as outlined above were applied by three expert linguists separately for each language and modality in a *modular* way. More precisely, annotation was initially performed on the [en] transcripts at the phrase and word level, first assigning a polarity. Nouns, adjectives, adverbs, verbs and multi-word expressions were treated. Further annotation was then performed at the sentence and clause level. We did not provide annotators with any predefined grammatical categories and the span of every annotation corresponds to the extended units of meaning (Sinclair, 1996; Hunston, 2007), i.e. what fragment of text is considered to express each emotion or opinion in the communicative instance

At the next level, cues beyond lexis that were provided by the audiovisual material were also taken into account with respect to the speakers' emotional state. To this end, a second round of annotation was initiated with annotators taking into account acoustic and visual cues, such as intonation, gestures and body language to interpret utterances.

As noted already, this procedure has been conceived of as a modular approach to annotation. Each level (word/phrase, clause/sentence) or modality contributes separately to the overall emotional load or attitude expressed either in a film or in any given film scene, shot, etc. To keep track at any given point of the contributing level or modality, however, each text span has been coupled with information on the level or modality from which the opinion or emotion is inferred. Many applications would benefit from being able to determine not just whether a film or scene is opinionated or emotionally overloaded, but also the contributing level or modality.

The source of every speech event is by default the

corresponding speaker and therefore it is not explicitly identified. Although we consider identifying targets, annotation at this level has not been implemented yet.
Finally, annotation was performed using the GATE (version 7.0) platform (Cunningham et al., 2002). The tool was selected for its user-friendliness and its versatility in fulfilling all the requirements of our annotation model.

6. Discussion

For the time being, only two movies have been annotated in two of the languages involved in the study: en-transcripts along with their el-subtitles.
It should be noted that, as one might expect, the data included in our corpus is quite different from the data usually treated in similar efforts that have been reported in the literature (see section 4), in that our data is oral and includes a significant amount of implicit speech events (not triggered by expressions, such as "I said", as in other works), conversational and includes highly colloquial discourse elements. This is a unique feature of our textual evidence.
To ensure annotation quality in terms of consistency, and in view of identifying problematic cases, inter-annotator agreement was calculated using Cohen's kappa coefficient (Cohen, 1960), i.e. a statistical measure of inter-rater measure for qualitative items. In an evaluation experiment involving 50 utterances, the inter-annotator agreement between 2 separate annotators on the word/phrase level was 0.92, dropping significantly at the sentence level (0.67 when all features were considered, and 0.86 when only polarity was taken into account).
Admittedly, annotating opinion and emotion in text is not a trivial task. Agreement was achieved in clear-cut cases, as in the following examples:

(19) Sweeney's *a good teacher*. (opinion – positive)
(20) I'm telling you, man, this kid is *smart*. (opinion – positive)
(21) this kid is *a genious*. (opinion – positive)
(22) Sweeney is *a nigger on a power trip*, (opinion – negative) Vinyard.
(23) They're *a burden* to the advancement of the white race. (opinion – negative)
(24) The gangs are *like a plague*. (opinion – negative)
(25) You've got to draw the line. (recommendation-neutral)
(26) I'm not ready to give up on him yet. (intention positive)
(27) I can guarantee you one hundred per cent his brother did not put him up to this. (opinion-belief, polarity-positive, strength-high)
(28) You hate this child (emotion – negative)

Instances presenting a disagreement between annotators must be further analysed so as to explicate the reasons underlying this difference. However, in cases which seem to be the most problematic, sentiment is not directly concluded from the co-text of the utterance examined.

7. Conclusions and Future Work

This paper has presented our ongoing work, aimed to develop a new specialized multimodal resource and to implement a pilot annotation scheme, so as to identify and represent opinions and emotions from a multi-modal perspective. The resource will tentatively be useful in a variety of cross-language studies and applications
Future work involves annotating of the remaining video material, including the [es] subtitles and developing and implementing a more fine-grained annotation scheme for our audio and video material, especially with respect to pragmatic phenomena, in order to facilitate a comparison between source and target texts and draw conclusions on translational norms and behaviours (Toury, 1995; Saridakis, 2010) with regard to subtitling practices in Greece and Spain.
Moreover, following common practices (Wiebe et al., 2005; Wilson, 2008), additional features will be implemented, as for example the identification of opinion/sentiment frames that consist of the opinion-holder or sentiment-experiencer and the target of opinion/sentiment respectively, etc.
Our future plans include also the annotation of the textual material at the various levels of linguistic analysis, with the focus being on the syntax and semantics of verbs, nouns, and adjectives that are indicative of emotions and/or opinions.
In conclusion, the present work might also prove useful for other researchers interested in the multimodal annotation, in the fields of sentiment and subjectivity analysis.

8. Acknowledgements

We must thank the anonymous reviewers for their useful suggestions and comments. Part of this work was supported by the Greek State Scholarships Foundation (IKY) through a Ph.D. scholarship awarded to E. Mouka.

9. 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.
Baños-Piñero, R. & Chaume, F. (2009). Prefabricated orality. A challenge in audiovisual translation. *Intralinea*, Special Issue: The translation of dialects in multimedia.
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.

Cohen, J. (1960). A coefficient of agreement for nominal scales, *Educational and Psychological Measurement*, 20 (1), pp.37-46.

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.

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.

Hunston, S. (2007). Semantic prosody revisited. *International Journal of Corpus Linguistics.* 12.2, pp. 249–268.

IMDI Team, (2009). IMDI Metadata Elements for Catalogue Descriptions, Version 3.0.13, MPI Nijmegen.

IMDI Team. (2003). IMDI Metadata Elements for Session Descriptions, Version 3.0.4, MPI Nijmegen.

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

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 .

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.

Saridakis, I.E. (2010). *Text Corpora and Translation. Theories and Applications*. Athens: Papazisi (in Greek).

Schmidt, T. (2011). A TEI-based Approach to Standardising Spoken Language, *Journal of the Text Encoding Initiative*, 1.

Sinclair, J. (1996). The search for units of meaning. *Textus*, 9.1, pp. 75–106.

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.

TEI Consortium, (eds.) TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 1.3.0. May 2011. TEI Consortium.

Toury, G. (1995). *Descriptive Translation Studies and Beyond*. Amsterdam: John Benjamins.

Waugh, L. (1982). Marked and Unmarked: A Choice Between Unequals in Semiotic Structure. *Semiotica* ,38, pp. 299-318.

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.