Opinion and emotion in movies: a modular perspective to annotation

Abstract

We hereby present work in progress aimed at the development and annotation of an audiovisual corpus resource with respect to sentiment and opinion. A modular annotation schema has been employed following the specifications of existing schemas with certain extensions and/or modifications that cater for the peculiarities of the specific data.

1. Introduction

This paper presents the first version of a new specialized audiovisual corpus resource that comprises movies coupled with both orthographic transcriptions 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) the focus being on the identification of opinion and emotion in oral discourse, elaborating on specific semantic and pragmatic phenomena. We describe the specialized corpus focusing on the pilot annotation procedure, and the results of an inter-annotator agreement study.

2. Project scope and aims

The audiovisual data were initially selected in order to guide translatorial research that touches upon the language depicting a specific type of biased opinionated ideological stance/attitudes, namely that of racist discourse, and its transfer from source language (SL) to the target language(s) during subtitling. This type of discourse is full of opinionated and emotional speech that makes it an ideal pool for annotating opinions, beliefs, thoughts, feelings, emotions, goals, evaluations, and judgments. From another perspective, the annotations were also oriented towards populating a lexical resource that contains opinion and emotion words with new entries adhering to oral data. This work, was therefore, integrated into a larger initiative aimed at the development of a multilayered corpus for sentiment analysis.

3. Corpus description

Being a product of so-called prefabricated orality (Baños-Chaume, et al., 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 (racist discourse). To date, the corpus comprises 5 movies that amount to 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 audio. Finally, transcripts have also been aligned with their official subtitles in [el] and [es], the latter provided by expert translators, and being a specialized type of translation in the sense that subtitling conforms to certain time and space restrictions.

More precisely, following standard procedures, that ensure conformance with standards for audio-visual material, and, thus re-usability of resources, video segmentation and transcription were performed using the ELAN tool (Brugman, et al., 2004). TEI specifications for spoken language transcription were taken into account (Schmidt 2001, TEI Consortium 2011). Segmentation has been performed at the utterance level following intonation and pause clues, while quite long stretches of speech are further segmented into C-Units to facilitate alignment with target-language subtitles that follow written discourse conventions and are typically formed of short sentences. Repetitions, hesitations, repairs and overlapping utterances that are inherent to 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:

![Figure 1: transcribed text](image)

The external structural annotation (including text classification) of the corpus also adheres to the IMDI metadata scheme. IMDI metadata elements for catalogue descriptions 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

Much work has focused on sentiment classification at the document, sentence or even phrase and word level. MPQA (Wiebe, et al., 2005) define attitudes as private states and propose an annotation schema catering for the following conceptualizations or types of attitude: sentiment, agreement, arguing, intension, and speculation. They also retain a general type other attitude for all the remaining private states. A value of positive or negative is
also assigned specific classes (sentiment, agreement, arguing).

Asher et al (2009) 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 opinion and emotion. This scheme is argued to be appropriate for calculating the overall opinion expressed in a text on a given topic. Other annotation studies (Polanyi et al., 2006), (Neviarouskaya et al., 2010), etc., provide detailed guidelines that are useful for polarity or sentiment/opinion determination.

5. Opinion Annotation in Movies

The annotation schema employed caters for the identification of two broad categories: (a) sentiment, that expresses the psychological state toward something usually based on feeling or emotion rather than reasoning; and (b) opinion, that is attitude, speculation, beliefs, thoughts, etc.

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

The schema also caters for a more fine-grained classification of opinion and sentiment. More precisely, the following opinion classes are defined: evaluation, judgement, recommendation, and other. Sentiment classification is centred around a set of 6 basic emotions (Plutchik, R., 1991): anger, fear, sadness, disgust, surprise, anticipation, acceptance, and joy.

Emotions were further assigned feature strength with possible values: low, medium, high, uncertain. The schema is depicted in Figure 2, below:

**Figure 2: Annotation Schema**

### 5.1 Annotating pragmatic phenomena

Movies depict situations in which dialogue participants make use of a wide range of rhetorical devices. To render the subjectivity annotation as complete as possible, annotation of pragmatic phenomena was also in order, and irony was the first one to be treated.

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 Grice’s 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 so are able to recognize the ironic dimension in speech.

From another perspective, irony has been proved to function in both a positive and negative way. In (Alba Juez, 1995) two main kinds of irony were proposed: Positive Irony (intended to praise) and Negative Irony (intended to criticise). The annotation attempted hereby takes this double classification in account, too.

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. Finally, contextual and/or world and situation-specific knowledge also guided annotation.

### 5.2 Annotation methodology

After initial specifications were set, 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/word level first assigning a polarity. Nouns, adjectives, adverbs, verbs and multi-word expressions were treated. Further annotation was then performed at the sentence/clause level.

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. That way they were aware of the situational context of every utterance.

6. Discussion

For the time being, only two movies have been treated in two languages: en-transcripts along with their es-subtitles. To ensure annotation quality in terms of consistency, and in view of identifying problematic cases, inter-annotator agreement was calculated using the Cohen’s kappa coefficient for qualitative items. In an evaluation experiment involving 50 utterances, the inter-annotator
agreement between 2 separate annotators on the word/phrase level was counted to 0.92, whereas it dropped significantly at the sentence level (0.67 when all features were considered, and 0.86 when only polarity was taken into account).

As a matter of fact, annotating opinion and emotion in text is not a trivial task. Agreement was achieved in clear cut cases, as in the following examples:

1. Sweeney’s a good teacher. (opinion – positive)
2. I’m telling you, man, this kid is smart, (opinion – positive)
3. this kid is a genius. (opinion – positive)
4. Sweeney is a nigger on a power trip, (opinion – negative) Vinyard.
5. They’re a burden to the advancement of the white race. (opinion – negative)
6. The gangs are like a plague. (opinion – negative)
7. I hate the fact that it’s cool to be black these days. (emotion – negative)

However, in cases which seem to be the most problematic, sentiment is not directly concluded from co-text. The following utterance:

8. “Not right now, honey”, The emotional state of irritation is evident only when the situational context is provided by the audiovisual channel. Similarly, regarding the utterance
9. “Fucking pervert, Dan!”, one cannot tell whether it is an evaluation, an insult or an expression of negative emotion without being aware of the context.

7. Conclusions and Future Work

We have presented a new specialized multimodal resource and the pilot annotations aimed at the identification and representation of opinions/emotions from a multi-modal perspective. Future work involves the annotation of the remaining video material, including the [es] subtitles, as well as extending the annotations to new text types and genres. Moreover, a more fine-grained annotation of audio and video material especially with respect to pragmatic phenomena will be further and more systematically pursued. Finally, following common practices, additional features will be implemented, as for example the identification of opinion/sentiment frames that consist of the opinion-holder/sentiment-experiencer and the target of opinion/sentiment respectively, etc.

8. References