Sentiment Analysis: From Opinion Mining to Human-Agent Interaction

Chloé Clavel and Zoraida Callejas

Abstract—The opinion mining and human-agent interaction communities are currently addressing sentiment analysis from different perspectives that comprise, on the one hand, disparate sentiment-related phenomena and computational representations, and on the other hand, different detection and dialog management methods. In this paper we identify and discuss the growing opportunities for cross-disciplinary work that may increase individual advances. Sentiment/opinion detection methods used in human-agent interaction are indeed rare and, when they are employed, they are not different from the ones used in opinion mining and consequently not designed for socio-affective interactions (timing constraint of the interaction, sentiment analysis as an input and an output of interaction strategies). To support our claims, we present a comparative state of the art which analyzes the sentiment-related phenomena and the sentiment detection methods used in both communities and makes an overview of the goals of socio-affective human-agent strategies. We propose then different possibilities for mutual benefit, specifying several research tracks and discussing the open questions and prospects. To show the feasibility of the general guidelines proposed we also approach them from a specific perspective by applying them to the case of the Greta embodied conversational agents platform and discuss the way they can be used to make a more significative sentiment analysis for human-agent interactions in two different use cases: job interviews and dialogs with museum visitors.

Index Terms—Sentiment analysis, opinion mining, embodied conversational agents (ECAs), human-agent interaction, affect, emotion, socio-affective interaction

1 Introduction

The domain of sentiment analysis has seen an upsurge of interest with the rapid increase of available text data containing opinions, critics and recommendations on the web (movie reviews, forum debates, tweets and other entries in social networks) [1]. The diversity of the data and of the industrial applications using sentiment analysis raises various scientific issues that have yet to be fully addressed by the existing systems. A challenging area is the development of opinion detection methods relying on these new sources. Opinion detection systems using sentiment analysis have been developed to target customers and evaluate the success of marketing campaigns [2], to know the user experience with certain products or their image of brands [3][4], or to predict stock price fluctuations [5].

Another growing research field is the development of embodied conversational agents (ECAs), virtual characters able to interact with humans. ECAs are involved in various applications. For example, they can play the role of an assistant such as the characters present on sales sites [6], of a tutor in Serious Games [7], or of a companion such as a health coach [8]. One of the scientific challenges of this research field is to integrate the affective component of the interaction. First, the ECA has to take into account the human emotional behaviors and social attitudes. Sentiment analysis of the user's verbal content is crucial for ECAs in order to determine the user's emotions and attitudes and to adapt its behavior accordingly. Second, they must be able to convey them appropriately.

The goal of this paper is to provide a state of the art on sentiment analysis from the perspective of the opinion mining and conversational agents communities, identifying the most relevant advances of both communities and discussing the open research questions and prospects. We have contextualized the discussion in the development of a sentiment analysis module and its integration in an ECA platform dealing with multimodal socio-emotional interactions. The final goal is to determine which reaction an ECA should have according to the user's detected socio-emotional behaviors and sentiments.

The communities analyzed in this paper have provided different state of the art studies and surveys following their particular vision of the topic. A variety of work addressing the aspects of emotion modeling can be found. Particularly paradigmatic and prolific are the studies by Scherer and colleagues presenting surveys on different aspects of defining, producing and modeling emotion [9], [10], [11], [12], [13]. Although there are currently many state of the art papers that address the concept of emotion and its various uses for different applications [14], the link between emotions, sentiments and opinions is not always clear [15]. With respect to the sentiment analysis community, [16] and [17] provide a state of the art of sentiment analysis in the context of data mining focusing on machine learning methods. These states of the art were recently updated by [18] and [19].

- *C. Clavel is with the Institut Mines-Télécom, Télécom ParisTech, CNRS LTCI75014, Paris, France. E-mail: chloe.clavel@telecom-paristech.fr.*
- *Z. Callejas is with the Depertment Languages and Computer Systems, University of Granada, CITIC-UGR18071, Granada, Spain. E-mail: zoraida@ugr.es.*

Manuscript received 4 Sept. 2014; revised 22 May 2015; accepted 8 June 2015. Date of publication 11 June 2015; date of current version 7 Mar. 2016. Recommended for acceptance by R. Mihalcea.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TAFFC.2015.2444846

With respect to the ECA community, special interest has been paid to emotion recognition. Notably, the work by Schuller and Batliner, who have organized special issues, challenges and a book in the topics of paralinguistics in speech and language [20], [21] and emotion in interaction [22].

Several authors have also presented specific proposals for affective interaction management. For example Pitterman et al. [23] proposed a dialog model including the dependencies between combined dialog and emotional states. Bui et al. [24] used statistical approaches that adapt the dialog strategy to the user's actions and emotional states, which are the output of an emotion recognition module. In our previous work we have also proposed statistical strategies to model the user's state considering their intentions and emotions [25]. Besides these general approaches, certain authors have focused on particular domains, for example D'Mello and Graesser [26] propose an intelligent tutoring system that automatically adapts its dialogs based on whether the learner is bored, confused, or frustrated, and tracks specific learners' behaviors such as disengagement. However, to our knowledge, there are no surveys presenting a comprehensive state-of-the-art of such approaches, focusing specifically on how to deal with affective interaction goals and manage the system responses according to the user's emotions.

The present paper aims at gathering all these aspects and at enriching the existing studies by providing a comparison of the underlying theoretical frameworks of *sentiment-related phenomena*[1] and the computational methods for sentiment analysis by addressing the following four key aspects:

- Which sentiment-related phenomena are relevant in order to determine ECA reactions?
- How relevant are the existing sentiment computational methods to be integrated in an ECA platform?
- How can sentiment analysis be considered by the affective interaction management system?
- Which interaction strategies can be used based on the user's perceived sentiment in order to conduct the interaction and decide the most appropriate ECA's behavior?

Our contribution is thus to confront the existing states of the art provided independently by each community and combine all these aspects together to integrate sentiment analysis in a human-agent interaction system. We focus on confronting the phenomena that are modelled in the existing computational methods and on the relevance of existing sentiment analysis algorithms for modeling the phenomena of interest for human-agent interaction.

Fig. 1 sums up the various scientific issues that are implied in the integration of sentiment analysis in human-agent interaction and presents how these issues are organized in this paper. Section 2 presents the various terminologies and underlying theoretical frameworks of sentiment-related phenomena. As the ECA-specific issues offer so far a scarce literature compared to the bank of literature offered by sentiment analysis for data mining, we choose to outline the existing sentiment analysis approaches concurrently with the existing socio-affective interaction and dialog management systems (Section 3) and debate some open questions and related works concerning the integration of user sentiment analysis in an interaction with an ECA (Section 4). Then, we show how to follow the guidelines proposed in the case of a particular ECA platform in two different scenarios (Section 5). Finally, Section 6 sums up the core avenues that have to be addressed for future ECAs to take advantage of the sentiment-related phenomena in their socio-affective interactions with the user.

1. A term which we will use here to regroup all the phenomena related to sentiment in the literature, from opinion to affect and emotion.

2 Terminologies and Theoretical Frameworks across Scientific Communities

2.1 Terminologies

Sentiments, opinions, emotions, moods, appreciations, social stances. . . Each of these terms refers to phenomena that present both specific features and overlaps. According to the scientific community involved and the targeted application, various terminologies are used. Scherer [14] defines the term *emotion* and its specificity according to other phenomena: preferences, attitudes, affective dispositions, and interpersonal stances. These definitions are references in the affective computing community. The natural language processing (NLP) community tends to use more frequently *opinion* and *sentiment*, whereas the ECA community tends to use *emotions*, but the actual studied phenomena overlap. Next sections discuss the subtle differences between the terms reported in the literature across scientific communities from computational linguistics to psychology.

2.1.1 Opinion Mining Perspective

Affects, emotions, sentiments, opinions: Krcadinac et al. [27] distinguish *affect* (or *emotion*) and *sentiment* analysis as two different research areas closely related to recognition. The authors state that sentiment analysis is focused on the classification of positive/negative valence, and affect analysis or emotion recognition in recognizing more fine-grained emotions. Many authors are consistent with this classification, even though they sometimes concede that the distinction between affect and sentiment is not clear in the community [28]. For instance, [29] denote as sentiment analysis their automatic analysis of written reviews of youtube videos in terms of positive or negative valence. Other authors attempt to recognize more fine-grained categories. They use two different appellations for their work: *emotion recognition* [30] and *affect analysis* [31], [28]. The Semeval evaluation campaign of 2007 [32] devotes a task to *emotion detection* in texts (headlines news) that relies on both emotional labels (the big-six) and the valence axis.

Despite this categorization, sentiment analysis has also been considered in a broader sense, as the computational study of sentiments, emotions and even opinions [33][34], and when it is focused on a positive versus negative distinction, they call it sentiment polarity. For example, [33] talk about "polarity sentiment analysis" but they present their study as a "sentiment analysis of microblog data". Similarly,

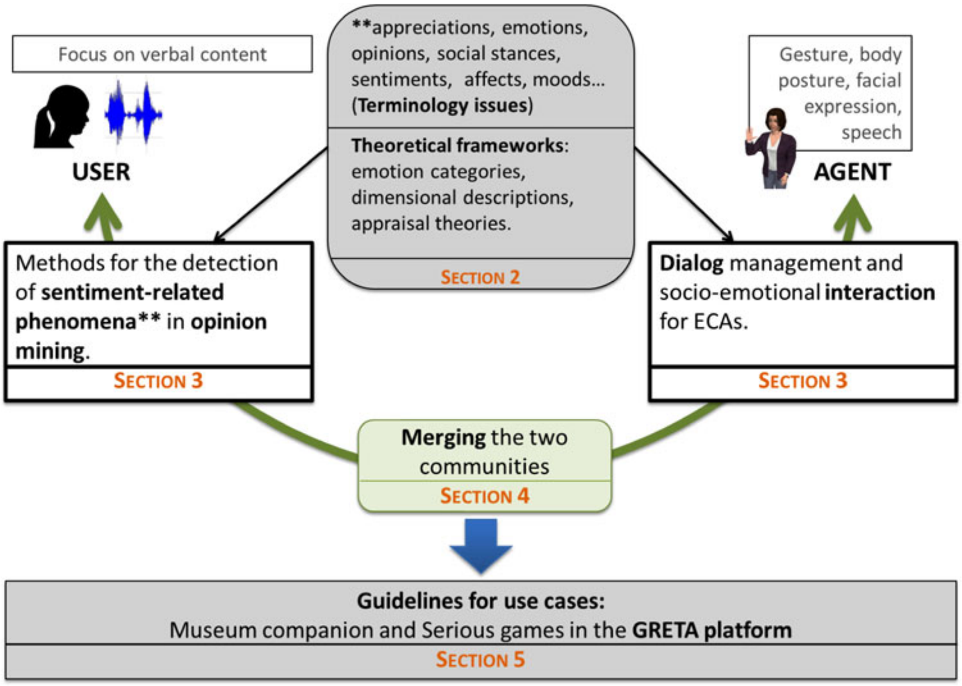

Fig. 1. Merging two communities—outline of the paper.

[34] present an approach to the detection of users sentiment polarity (positive, neutral or positive) in Facebook, which is presented as a "sentiment analysis tool". In [35], sentiment is considered as the affective part of opinion.

The use of *affect*, *emotion*, *sentiment* and *opinion* in the previous works is more dependant on the implied scientific communities than on a difference between these terms. Munezero et al. [15] offers an in-depth analysis of these differences from a NLP point of view. According to them, affects exist before self-personal awareness and are a predecessor to emotions. Affects have therefore no expression in language. Emotions differ from sentiments by their duration (emotions are briefer) and by the presence of a target (emotions are not always targeted towards an object). According to [15], opinions are not the same as sentiments: they are personal interpretations of information and are not necessarily emotionally charged.

Attitudes, feelings and mental states: Other authors have also tried to define the sentiment-related phenomena in language. Kerbrat-Orecchioni [36] do not use the words sentiment or opinion but instead the terms *attitudes* and *feelings*. They define *attitudes* as "concerned with our feelings, including emotional reactions, judgements of behavior and evaluation of things" (p. 35) and distinguish three kinds of *attitudes*: affect (personal reaction referring to an emotional state), judgment (assigning qualities—e.g., tenacity—to individuals according to normative principles) and appreciation (evaluation of an object—e.g., a product or a process).

Linguists such as [37] or [38] make a similar distinction. Charaudeau [38] relies on discourse theory in the line of [39] to define discourse modalities to spot evaluative features, distinguishing opinion, appreciation, agreement, acceptation and judgment. Kerbrat-Orecchioni [37] discusses that evaluation can be expressed emotionally: speakers may or may not be emotionally involved in the content of their utterance.

Other approaches, as [40] or [41], refer to the Private State Theory, which defines *mental states* as involving opinions, beliefs, judgements, appraisals and affects.

2.1.2 ECAs Perspective

Emotions, feelings and mental states: The ECA community has strong roots in psychology and the computational models that are most frequently used are based on psychological theories of *emotion*. The Emotion Markup Language (EmotionML) from the W3C[2] specifies the various terminologies and theoretical models that have to be used for both the automatic recognition of emotion-related states from user behavior and the generation of emotion-related system behavior.

The literature in psychology concerning theories of *emotions* is plentiful [42] and different theories have led to varying definitions that use different words to describe emotion. More recently, different authors have centered on the distinction of emotions as perceptions of bodily states versus cognitive approaches [43]. The first vision identifies with Jame's theory [44], in which emotions are considered to be *feelings*. In the second approach every emotion is determined by an evaluative judgment [43], and therefore an emotion can be considered to be a *cognitive state*.

Some authors argue that the human mind consists of a variety of *mental states* [45] that, although usually denoted with different terms such as emotions, cognitions or perceptions, cannot be distinguished in practice by neuroimaging, which shows that there is no specific brain activity for each of them and they are merged and processed in the same networks.

Emotions, and affects versus moods, a question of duration: When defining *emotion*, the question of intensity and duration is a key issue tackled by the introduction of a new term:

2. http://www.w3.org/TR/emotionml/

mood. Oatley and Johnson-Laird [46] elaborate on emotion-based states over different time durations, from a few seconds for facial expressions to minutes and hours for experiences of emotion, hours to weeks for *mood*, months for psychological illnesses, and lifetimes for emotion-based traits of personality. Beedie et al. [47] go a step further in the distinction between emotion and mood and use 16 criteria that differentiate emotions and moods, including not only duration but also cause and intensity. Apart from duration and intensity, [14] also discusses the fact that appraisal is not linked in the same way to emotions as to moods, as moods often appear without apparent cause (e.g., being gloomy or cheerful). Sometimes certain terms are considered pathologies or disorders. For example, [48] address loneliness and isolation in older adults by considering mood to investigate the longitudinal interventions of an ECA.

Some authors describe *affect* as comprised of emotions (specific, intense, and short responses to stimuli) and moods (longer, more ambiguous, non-attributable, and lower intensity) [49]. However, it has also been described as a temperamental response to emotions. For example, [50] present a study on the relationship between affect intensity, regulation of emotion and drinking to cope. When the authors talk about affect they use it to describe the intensity of emotions. Thompson et al. [51] also describe affect as intensity and attention to emotion.

Social stances, interpersonal stances: Social interaction studies have a different perspective on emotion and introduce a new concept that is widespread within the ECA community: *social/interpersonal stances*. Social interactions were modelled by Scherer [14] in his definition of interpersonal stances: "an affective style that spontaneously develops or is strategically employed in the interaction with a person or a group of persons (e.g., being polite, distant, cold, warm, supportive, contemptuous)." Ochs et al. [52] make the same distinction between interpersonal stances such as warm and polite and emotions to analyze the expressions of a smile in an ECA. Another reference work in social interaction is Vinciarelli's concept of social signal processing [53] which distinguishes social signals (short-time phenomena) from social behaviors (longer duration, temporal patterns of non-verbal cues including facial expressions, body postures and gestures, and vocal outbursts like laughter).

2.2 Theoretical Frameworks of Sentiment-Related Phenomena

Behind the various terminologies described, there are different frameworks used in a variety of ways by the two communities under study in this paper, and by the ECA community in particular. We propose here a common categorization of these frameworks that arise from both communities: dimensional (representation of sentiment-related phenomena on different axes, such as valence and arousal [54], [55]), discrete (representation of sentiment-related phenomena into distinct categories, such as fear or joy [56], [57], [58]) and appraisal (sentiment-related phenomena are linked to the process of evaluation of events, objects or persons [59][12] [36]). Such categorization has been already used in the affective computing community with EmotionML (see Section 2.1.2) and suits an opinion mining perspective.

2.2.1 Opinion Mining Perspective

Although the underlying theoretical frameworks of the opinion mining and sentiment analysis are not always clear, there exists some background literature in this domain which comes from three main research communities: psychology, sociology and linguistics.

The *dimensional approach*, from psychology, is the most frequently used, even though it is barely mentioned explicitly. Most detection systems focus on the valence axis and propose, as mentioned in Section 2.1.1, a polarity analysis consisting of a positive versus negative classification. This is the case of [60], [61] or [62], [29] in movie reviews, [63] in various products reviews (recommended or not recommended), [33] in micro-blog data, and [34] in facebook data. Other authors also employ the arousal axis, e.g., [64]. Usually the arousal dimension represents the strength of a certain emotion encountered in a text [65].

In sentiment analysis, the studies that focus on emotions rather than on polarity, usually employ the *discrete approach*, in which a number of categories are considered and each unit (expression, sentence, paragraph, or document) is classified into these categories. In this case the first challenge is to determine the categories or labels to consider, which is a core aspect of the domain, given the complexity of sentiment-related phenomena. In contrast to polarity detection, the aim is to detect more categories of emotion or subjectivity in texts, such as in blogs and social networks.

For example, [66] presents a model that distinguishes between nine categories (anger, disgust, fear, guilt, interest, joy, sadness, shame, surprise), [30] recognizes six (anger, disgust, fear, happiness, sadness and surprise), while [28] distinguishes between 22 categories defined in the Ortony Clore and Collins (OCC) model (described in the next paragraph). Other authors focus on specific components of sentiment-related phenomena through categories such as agreement versus disagreement [67], or flame detection in private messages sent to controversial pages [68]. Some authors also employ *hybrid approaches*. For example, [69] proposes an automated emotional markup in text using both categories and emotion dimensions.

Two types of *appraisal theories* are considered in sentiment analysis and opinion mining communities. The first one is the OCC model which focuses on the cognitive structure of emotions [59], [70] and the second one [36] is based on the social constructionism theory [71] and considers not only emotions and affects but also judgment and appreciation (see Section 2.1.1). The use of appraisal theories for opinion mining is especially interesting where it is necessary to reliably determine the target of the opinion, such as for the automatic analysis of reviews (e.g., movies, hotels, products) and opinions about brands and services (e.g., in social media). As discussed by [72], in some cases, it is not only necessary to consider the affect or attitude (evaluative stance about an object), but also to find the targets (the object of the stance) and sources (the person taking the stance) that may be implied.

The *appraisal theory* has been used in this way for the automatic analysis of textual opinions. [73] refers to this definition of appraisal expression and its relation to the target and the source for the analysis of customer opinion in call-center transcripts. [66] proposes an @AM (ATtitude Analysis Model) modeling the previously described appraisal model and distinguishing affect, judgment and appreciation (see Section 2.1.1) on sentences extracted from personal stories. An alternative but similar model to the appraisal theory has been proposed in [74] in order to model the source and the target of the opinions and emotions. It proposes a specific approach in order to model direct versus indirect expressions of opinions and emotions that are not distinguished in the appraisal framework.

2.2.2 ECAs Perspective

The ECA community essentially relies on the field of psychology to build its model for expressing emotional behaviors and interacting with humans, with a focus on the detection of *emotion* from speech or from visual cues (facial expressions, gesture, posture).

In the *dimensional approach*, most studies focus on detecting negative emotional states, as they are detrimental to the performance of the system, either because they can cause the user to give up interacting with the system [75], or because they make it more difficult for the system to understand the input distorted by a deep negative emotion, such as the study of recognition of speech under stress [76]. In other domains of application, avoiding user frustration is not only important for the system, but also for user security (for example [77] detects drivers' frustration and cognitive load in their speech), or for their learning (as in the case of tutoring systems [78]).

Other studies propose a more sophisticated use of the valence-arousal dimensions for artificial agents. For example, [79] proposes the use of several layers that coordinate to obtain a more expressive behavior.

With respect to the *discrete approach*, on the one hand there are studies that employ basic emotional categories, such as Ekman's Big-Six [56], the main four proposed by Picard [80], or the eight basic emotions described by Plutchik [44]. On the other hand, other authors expand the notion of emotion by using a list of emotional terms depending on the applicative contexts [25], [81]. Work integrating user sentiment detection based on linguistic cues for ECAs focuses on selected emotional categories, (e.g., the Big-Six of Ekman) and complex affect labels (such as approving), e.g., the OCC model in EMMA [82], the nine emotions including the Big-Six in Emoheart [31], or other different user emotional states [83] (that are actually a distinction between neutral, polite and frustrated). Also [84] proposes a classification between positive versus negative user sentiment as an input of human-agent interaction. Positive versus negative classifications overlap with the representation of sentiment in the valence axis, as previously explained in Section 2.2.1

The main *appraisal model* used is the OCC model (see Section 2.2.1). As shown in [85], it has been very popular within the ECA community, especially for synthesizing the emotional behaviors of the agents, classifying events, objects and actions and quantifying the emotions generated and how they interact with other emotions to define and express the agent's state. The use of appraisal theories in ECAs makes it possible to compute more complex models on how emotions should drive the ECAs behavior and decision-making processes. For example [86] discuss how emotion drives judgment in these agents, while [87] presents a model based on the OCC theory in which the agent analyzes the context of the interaction and sets its goals and behaviors accordingly. These selection mechanisms has also been smoothed using fuzzy logic [88], [89].

However, according to [90], the focus of OCC on cognitive components results on very narrow computational models, so they describe some alternatives that include other emotional elements such as Scherer's appraisal model based on sequential checking theory[3] [91] and dynamic aspects [92]. Several authors are investigating these approaches within general agent architectures [93], for example with the belief-desire-intention (BDI) and biologically inspired cognitive architectures (BICA) models [94], [95].

2.3 Summary—Overlapping Terminologies, Three Models and Two Communities

The ECA community essentially tends to rely on psychological theories in order to model phenomena focused on *emotion*, *affect* and *mood*. The recent tendency of ECA community is also to integrate the point of view of social interactions. Opinion mining community prefers the terms *opinion*, *sentiment* or *affect* but the major part of existing works does not provide in-depth definitions of the phenomena associated with these three terms. It is interesting to notice that the term *affect* can take very different meanings according to the studies. However, some existing linguistic studies using the term *attitude* go further in the definition of sentiment-related phenomena for NLP.

Table 1 presents a summary of the three theoretical frameworks used by both communities. It shows that polarity recognition—that can be viewed both as a dimensional and discrete approach—is very frequent in the community of opinion mining (positive versus negative classification) and is also used by ECAs to detect user's negative emotions for interaction management. The appraisal OCC model is used by both communities but each community has also their own appraisal model (Martin and White [36] for opinion mining and dynamic appraisal [92] for ECAs).

It is also important to note that the very unfrequent studies integrating the detection of user sentiment in ECAs based on linguistic cues, concern more visualisation than interaction issues [31], [82]. We identify only two studies that integrate a sentiment detection module for human-agent interaction [83], [84].

With the aim to integrate the detection of sentiment-related phenomena in human-agent interaction, the two communities have to be connected. First, the ECA community has to go beyond the model of *emotions* and to consider other sentiment-related phenomena such as attitudes and social stances. Second, the opinion mining community has to improve its definition and its models of the targeted sentiment-related phenomena. Section 4.2 presents guidelines in this sense.

3. According to the sequential checking theory, emotional states are determined according to a sequence of appraisal checks (e.g., novelty and coping potential checks).

TABLE 1
Illustrations of Most Frequent Models of Sentiment-Related Phenomena by the Two Communities

	Opinion mining applications	ECAs applications
Dimensional approaches Sentiment can be characterized as a point in a n-dimensional space. The most widespread dimensions are valence (pleasantness or hedonic value) and arousal (bodily activation) [54], [55].	• Polarity detection (positive versus negative): opinions and reviews, attitude towards brands, preferences and comparatives, agreements [60], [61], [62], [63], [29], [34], [33], [96]. • Computation of the strength of emotions expressed in texts: analysis of blogs, social networks, movie reviews [64].	• Detection of negative emotions in order to optimize ASR [76]. • Detection of negative emotions for interaction management [75], [84]. • Model agent expressive behavior using valence-arousal [79].
Discrete approaches Sentiments are considered separately from one another and assumed to be unique experiential states [56], [57], [58], [80], [44].	• Analysis of positive vs. negative categories (see references of dimensional approaches). • Emotion/affect detection in texts [30], [69] (discrete approach mixed with dimensional approach), [66], [28] (22 OCC categories). • Analysis of other sentiment-related categories: [67] (agreement versus disagreement), [68] (flame detection). • Subjectivity detection in texts [17].	• Recognition of negative emotion categories (see references of dimensional approaches). • Detection and avoidance of user frustration in driving situations [77] or for tutoring systems [78] or for a child conversational computer game [83]. • Detection of various emotions according to the application for dialog systems [25], [81], [83]. • Textual affect visualization in avatars [82], [31].
Appraisal approaches The focus is put on the modeling of the process of evaluation of events, object, or persons [59], [12], [36].	• Detection of the target of the user sentiment in: reviews [72], call-centers [73], personal stories [66], using for example Martin and White's appraisal model [36]. • Textual affect sensing [28], [70] and representation of affective knowledge in lexical resources [97], for example with the OCC appraisal model.	• modeling the system and user decision making and behavioral models as well as the context of the interaction, using different models (e.g., OCC [86], [87], [88], [89], Scherer's Sequential Checking Theory [91], or the dynamic appraisal EMA model [92]) and architectures (e.g., BDI [94] or BICA [95]).

3 Computational Models: Transverse Overview

Sentiment/opinion detection methods used in human-agent interaction are rare and, when they are employed, they are not different from the ones used in opinion mining and not designed for socio-affective interactions. Besides, many works that address affective components in human-computer interaction deal mainly with conceptualizing and detecting emotional states. However, it is also necessary to investigate how they can be used by dialog management systems to build socio-affective interactions. The following sections discuss the existing approaches to address this challenge, starting with an overview of sentiment detection methods, followed by a description of the main approaches to affective dialog management. The challenges for using such methods in human-agent interactions including multimodal sentiment analysis will be developed in Section 4.

3.1 Sentiment Detection Methods

One of the key particular aspects to opinion mining is to process negation and intensifiers to correctly discern the polarity and intensity of the judgment. Anaphora resolution [98] and relation extraction are also crucial in order to identify the target of the evaluation (see appraisal approaches in Table 1), e.g., when the target is indicated by a personal pronoun. Disambiguation of the opinion sense also depends on various contexts such as the persons expressing the opinion and their personality, and their social or political affiliations. Ironic expressions are also very frequent and difficult to detect [99]. Moreover, the ability of the system to analyze metaphors [100] can be useful for opinion analysis (i.e., the use of global warming versus climate change). In order to tackle these aspects, the literature in sentiment analysis presents two types of methods (and hybrid versions of them): methods based on semantic rules and statistical methods.

The first ones concern the design of rules for the extraction and automatic labeling of sentiment-related expressions contained in texts. They rely not only on the occurrence of words belonging to sentiment lexicons [101] but also on linguistic extraction patterns combining different features obtained from morpho-syntactic analysis (word form, lemma, part of speech tags) or from the outputs of other rules [64], [102]. There are a variety of open source sentiment lexicons in English used by sentiment detection systems: SentiWordNet [96], EmotiWord[103], the LIWC dictionary [101], WordNet Affect [97], and Sentiful [104] are some of the most widespread.

The second type of methods is based not only on simple statistic methods but also on supervised machine learning methods. As an example of a simple statistical method, we can cite [63], which uses unsupervised learning techniques based on the mutual information between document phrases and the words "excellent" and "poor", where the mutual information is computed using statistics gathered by a search engine. Supervised machine learning methods such as support vector machines and naive Bayes classifiers are frequently used for this task [60].

Rule-based methods handle the previously listed challenges better as it is possible to build dedicated rules for anaphora resolution, processing negations and intensifiers, target identification [72], source identification, and metaphor disambiguation [105]. However, in order to attain acceptable success rates it is necessary to restrict the domain of application as much as possible. Thus, their main disadvantage is the difficulty of building generic extraction patterns and lexicons to extract all sentiment-related expressions contained in the data and to assign them a relevant label in varying contexts. Some authors have proposed different solutions, such as expanding the affective lexicon with new entries based on semantic similarity [103] or linear programming [106].

Supervised machine learning methods make it possible to generate more interoperable models, but they require the availability of labelled data for training. The quality of the models learned strongly depends on the reliability of sentiment annotation, which is affected by raters' subjectivity as existing annotation guides are rare. Besides, this type of methods allows building models that are sometimes difficult to interpret and control, which is why hybrid (statistical and rule-based methods) are sometimes used [107] [108]. Also some authors propose methods that provide more control, e.g., [109] uses conditional random fields (CRF) tackling opinion source identification as a sequential tagging task, whereas [110] identifies the target of the opinion with CRF. In [67] a Bayesian network is also used to model pragmatic dependencies to improve agreement analysis in meetings.

The tendency is now to handle the drawbacks of each type of method using hybrid methods, which bring both the generalizable nature of machine-learning approaches and the in-depth modeling offered by semantic rules. For instance, [72] uses probabilistic models for disambiguation, and tools like AutoSlog provide supervised pattern trainers that have been used for opinion source identification [109], while [62] proposes merging unsupervised machine learning algorithms with supervised ones.

Different contexts and domains of application may introduce other specific challenges. It is thus important to keep in mind that the performance of such systems is strongly dependent on:

- the type of data to be analyzed: the style and the language register of the writer/speaker, the quality of the syntactic structure of the data (tweets and oral transcription versus newpaper). For example, sentiment detection in social networks or transcriptions of speech must be able to cope with ill-formed linguistic structures [111] and other phenomena such as orthographic mistakes, emoticons and other symbols [112].
- the classes of sentiment that are considered: classification according to ten classes is indeed more difficult than for two classes. Performance also depends on the choice of classes being studied, e.g., discriminating between appreciation and judgment is generally more difficult than between positive and negative.
- the quality of the ground-truth annotations: evaluating performance consists of analyzing differences between ground truth annotations and system decisions. In the field of sentiment, we cannot speak about actual human error, since the situation is more complex (subjectivity of the annotations, and imbalanced proportion of emotional and neutral contents [113]).

3.2 Socio-Affective Dialog Management

Usually, interactive agents follow an "affective loop" [114], a cycle comprised of three steps: recognizing the user's emotion, selecting the most suitable action according to what has been recognized, and synthesizing an appropriate affective response.

These phases can be further divided. For example, [95] decomposes affective modeling into five core generic computational tasks: i) defining a mapping between the elicitor and an emotion, ii) computing the intensity of the resulting emotion, iii) computing the decay of the emotion over time, iv) integrating multiple emotions if necessary, and v) integrating the new emotion with existing emotions or moods.

In traditional conversational systems, the dialog manager receives a semantic representation of the user's previous input, and uses dialog history and different user or interaction models and/or external information sources to decide on the next system response [115]. Affective dialog management implies processing affective information as another input to the dialog manager or the user model used by the manager.

According to [116], emotion can be considered the manipulation of the range of interaction affordances available to the agent (the agent has a set of dialog actions or behaviors from where to choose how to behave, and emotion is part of the function that selects the best one). As the emotional state of the counterparts is affected by their interlocutors, the authors indicate that "emotion becomes a strategy for goal fulfillment" that allows the dialog manager to select the most appropriate action over a more meaningful space of possible utterances.

[117] presents a statistical dialog manager based on partially observable Markov decision processes (POMDPs) that considers an extra input that includes information about the user state (positive or negative). Similarly, [118] suggests the use of Affective Profiles for the users and the system and discuss how to use them to build script-based affective dialogs using the AIML markup language adapted to the emotions recognized in the user's textual input. In [25] a mental-state model is presented that is comprised of the user's recognized emotion and predicted intention, which is used by the dialog manager to select the most suitable system response. An affective dialog manager can be used to modulate the ECA's responses according to the user's affect [25].

3.3 Summary—Affective Interaction Based on Sentiment Detection

A major part of existing sentiment analysis methods focus on the distinction of positive versus negative affect, even though both advanced machine learning and rule-based methods offer a great potential to go further, as they make it possible to distinguish more fine-grained categories and to identify the target and the source of sentiment-related phenomena.

Socio-affective interactions aim to tackle user's socio-emotional behavior in different ways. In order to achieve such complex behaviors, the systems must detect specific aspects of user's behavior not only as an input of interaction strategies but also as an output (both to check the efficiency of interaction strategies and to help the detection). The two following sections present tracks to handle these issues.

4 Merging the Two Communities: Open Questions and Prospects

There are many opportunities for the communities under study to benefit from each other. In the following sections we will discuss the issues that should be addressed and provide some guidelines on how they could be integrated, especially from the perspective of integrating sentiment analysis in ECA.

4.1 Emphasizing Socio-Affective Interaction Strategies and Goals

The type of sentiment-related phenomena that are relevant to human-agent interaction varies with the application and thus with the socio-affective interaction goals. Although interaction goals may be domain specific, there also exist common goals, such as the acceptability, social competence and believability of the ECA, or the user engagement in the interaction. To fulfil such goals, socio-affective interaction strategies are set up focusing on phenomena that are rather related to *emotions*. The following sections describe the main interaction goals and strategies pointing out possibilities for enhanced human-agent interaction by integrating relevant sentiment-related phenomena.

4.1.1 Dealing with Users' Emotions: From Awareness to Self-Awareness

In most approaches found in the literature, agents have been designed for the sake of avoiding user *frustration*. A clear example are tutoring systems in which affective information can be used to choose the best pedagogic strategy to avoid frustration during learning.

However, there are also applications in which it is interesting to provoke negative emotional responses in the users in order to improve the believability of agents. Several authors have attempted to frustrate users and then assess how they respond to agents with different degrees of empathetic responses [119]. Also [120] generated a model of impoliteness for believable non-player characters in video games. Uncomfortable user experiences in human-computer interaction may also have benefits [121]. For example, FearNot! [122] developed during the EU projects Victec and eCircus explored the use of negative ECAs for different purposes including the treatment of phobias. Similarly, Campos et al. [123] developed a serious game for helping children to learn conflict resolution strategies.

We believe that additional efforts should be made to analyze the impact of socio-affective strategies on users' negative emotions, not only at particular moments of the interaction, but also on how they develop over time and how they can be tackled using different dialog management strategies. To achieve this goal, sentiment analysis can play a relevant role helping to understand the progress of the user emotional states. First, it can provide additional information sources that intricate with emotions but also have different time spans and thus offer complementary knowledge, such as opinions and attitudes. Second, it can provide accurate information about the source and target of sentiment-related phenomena that can help to discern whether the changes in the users' states are due to the interaction policies or to other issues related to the agent or the context of the interaction.

Boehner et al. suggest a different perspective of affective interaction in which the aim "is not making systems more aware of emotions, but making people more aware of emotions through system use and design" [124]. Different studies indicate that sustained affective interactions can help users examine their emotional experiences over time, which has been proven to be helpful for emotion regulation [51]. Thus, the emotion management and synthesis in affective agents can be fine-tuned to provide a more satisfying user experience [125]. However, this field has still not been sufficiently addressed. The deeper understanding of the user through sentiment analysis described earlier, can help to further advance this area offering improved possibilities for emotional feedback.

4.1.2 Building Long-Term Relations Based on Empathy and Trust

In application domains that require long-term relations between the agents and the users [126]—this is the case of relational agents such as coaches, tutors [127], healthcare assistants and social companions [128]—, it is crucial to be able to detect and manage phenomena that are more permanent than emotions. Thus, sentiment analysis methods can contribute to consider the elements that have a longer duration, such as *mood*, trust, or different types of *empathy* (see Section 2.1.2).

As a matter of fact, one of the main socio-affective interaction goals in these contexts is to emphatically engage with human users over long periods of time [129], for which some affect-related constructs such as empathy, trust, persuasion or comfort are particularly relevant.[4] Empathy has been characterized attending to different dimensions. Looije et al. [131] build on the concepts of empathy and trust, and divide empathy into three dimensions: complimentary, attentive and compassionate. [132] studied the implications of parallel and reactive empathy in tutoring agents. Parallel empathy describes a person displaying an emotional state similar to that of another individual, which expresses the ability to identify with the emotions of others. Reactive empathy provides insight about recovering from that emotional state. The authors showed that combining both types of empathy by displaying adequate emotional expressions changed the emotional states of student from fear to neutral states. Similarly, [133] presents the CARE system, which supports empathetic assessment and interpretation. With respect to trust, different aspects have been studied, such as the importance of the ability of the system to provide effective explanations of its own behavior [134], [135].

4.1.3 Increasing the Agent Likeability and the Engagement of the User

The modulation of the ECA's responses according to the user's affect helps the agent to be more effective and to establish richer relations, fostering likeability [25]. Sentiment analysis can help to build enhanced interactions to make ECAs more likable or to improve the user's opinion about their agency.

Currently, when authors address these challenges, they use opinion surveys with users in order to find out whether they were successful to produce adequate strategies. This way, they can optimize the interaction policies to create a new version of the ECAs, which are iteratively enhanced through these perceptive studies [136]. However, this is an offline process in which the ECA is not learning from the interactions, and the reason why it is prevalent is because it is complex to find out whether the user likes the agent on the fly. Sentiment analysis can help to solve this challenge by incorporating affective cues coupled with the target of the affect, which can be used to track the perceived likeability of the ECA and the user engagement levels during the interaction. Furthermore, it can contribute to provide new inputs to particular socio-affective strategies specifically envisaged to improve agents likeability and foster user engagement. This is the case of mirroring strategies.

As discussed in [137], people tend to align their way of speaking when they talk to each other and to machines so a basic strategy is to use mirroring. Mirroring has been proved as a sign of human empathy and interpersonal competence [138], and can therefore be considered an appropriate behavior for an agent that should be perceived as socially competent. Different studies have addressed emotional mirroring in ECAs, for example [139] mimicked user emotions through the agent's facial expressions, and [140] used facial expressions, gestures and head positions. Other studies address more complex mirroring strategies including more subtle mechanisms, as it is the case of the social dynamics of laughter [141], in which the user state is not copied, but interpreted in different ways that affect the system behavior at several levels. Agents which adequately employ mirroring strategies are perceived as more empathetic, and one of the consequences is an increased rapport [142], [143].

However, [144] found that the likeability of their ECA depended mainly on the agent's ability to answer the users requests appropriately, which shows the importance of building appropriate interaction strategies integrating deep analysis of user sentiment-related phenomena for which mirroring may not be sufficiently complex. Besides, engagement policies that are coupled with user models [79], [145] are required to vary their behavior in a more subtle way. Indeed, a sustained interaction with an agent that always behaves in the same way would drastically decrease user satisfaction over time [115], but they are also useful in short-term relations to maintain the interaction with the user [146].

4.2 Putting Theoretical Models into Practice

The large variety of terminologies (Section 2.1) and theoretical frameworks (Section 2.2) shows the complexity of sentiment analysis. As explained in Section 2.2, the majority of computational methods of sentiment analysis focus on the positive/negative distinction although they also consider arousal sometimes [64]. This model is well suited to simplify the opinion classification problem for machine-learning approaches and can be satisfactory for ECAs in some cases.

However, these two dimensions are not sufficient to individuate the existing sentiment-related phenomena in language [97]. One of the key challenges for sentiment analysis is to go beyond the positive versus negative classes. Regarding human-agent interaction, this breakthrough is crucial to fulfil the socio-affective goals presented in Section 4.1. With this in mind, works such as [72], distinguishing affect from judgment and linking sentiment to its target, are promising and make this challenge reachable for human-agent interaction.

We believe that merging the perspectives of opinion mining and ECA communities can be of reciprocal help. On the one hand, the appraisal models used by the ECA community can be used for opinion mining to define the cognitive effects of a past emotion in the expression of the user's current subjectivity. For instance, [147] use the OCC model to link words such as "victory" to later expressions of pride, hope or disappointment expressed by winners and losers [148]. On the other hand, clearly identifying the targets and the sources of sentiment-related phenomena is a highly useful input for the agent's socio-affective interaction strategies described in Section 4.1. Opinion mining is based to a great extent on spotting entities, objects and aspects for which the user expresses an opinion [149]. It would allow ECAs to perform a deeper analysis of the target of the users' sentiment-related phenomenon. For example, in the OCC theory, each user judges the desirability of an event, the approbation of an action, and the attraction of an object [59]. The attraction of an object denotes the correspondence of its characteristics with the user's appreciations.

4. Prolonged interaction with agents may influence user's affect and behavior in many different ways. The scientific community is involved in establishing ethical guidelines for such agents. For example, [130] discusses the ETICA (Ethical Issues of Emerging ICT Applications) project and its implications for empathetic care robots.

Thus, being able to identify those characteristics and evaluate how they relate to user's affect is of great importance for managing appraisal.

More generally, the target and the source of sentiment-related phenomena are information that could enable an ECA both to adapt its behavior and to try to lead the user to an optimal state for interaction. For instance, if the ECA detects that the user does not like another agent or object, it can decide to no longer speak about this object [150]. The detection of user interest and preferences in the interaction and in the topic of the conversation is of great value to human-agent interactions. Such aspect is also crucial if the socio-affective interaction goal is to increase the user liking towards the agent (see Section 4.1).

Another benefit concerns the type of distinction introduced in Section 2.1.1 considering affect, judgment and appreciation, which will enable the ECA community to consider other phenomena than the user's emotions. It allows the ECA to determine the stances of the user with regard to the content of its utterance and to the ECA. The ECA can also lead the user to feel/express other stances thanks to its dialog strategy. For instance, if the user expresses a negative appreciation towards the ECA, it can try to improve its relationship with the user.

The appraisal approach of [36] used by the opinion mining community could be very interesting to answer the previously described issues as it provides the possibility to:

- Model the sentiment-related phenomena with its source and target.
- Distinguish the various types of sentiment-related phenomena, e.g., judgement and affect.

Another interesting issue for modeling users' sentiment-related behavior is to be able to recognize and synthesize transitory emotional states. ECAs work with constant and transient affective aspects. For example, [151] conceptualizes events and actions as user acts that create transient affective meaning. In some cases, this information is complemented with constant or less ephemeral aspects such as moods (see Section 2.1.2) or personality traits [25], that influence the transient expression of emotion. Detecting such aspects from user utterances would therefore be a great issue for NLP methods described in Section 3.1.

4.3 Adapting Computational Methods to Conversational Speech

The integration of the previously described opinion mining methods into the domain of ECAs is not direct, as the conversational nature of human-agent interactions imposes certain restrictions that are not present in textual NLP. In the following sections we describe some of the most relevant issues that must be considered.

4.3.1 Integration of the Multimodal Context

The targeted human-agent interaction is face-to-face, which means that verbal content takes place in a context of multimodal speech [152]. The non-verbal signal provides complementary information to the linguistic content for sentiment analysis [29] through features which are dedicated to characterizing more the affective component of sentiment than the component related to judgment and appreciation (see Section 2.2). There are a large number of studies focusing on the acoustic features such as prosody, voice quality or spectral features for emotion detection [76], [153], [25] and more generally on non-verbal features (posture, gesture, gaze, or facial expressions [154]). The field of social signal processing contains a bank of literature on this issue [53]. The direct integration of non-verbal emotional content is widespread in human-agent interaction [84], [105], [155], [156].

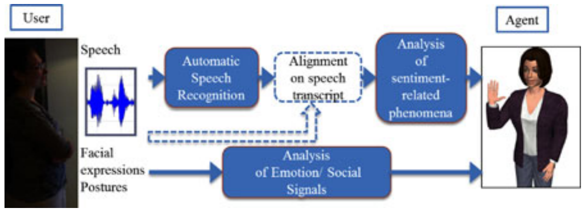

Fig. 2. Integrating sentiment analysis in a multimodal context in a human-agent interaction.

Fig. 2 illustrates the two ways of integrating a sentiment analysis module with the analysis of non-verbal content of speech. A first step towards multimodality is therefore to consider sentiment analysis from the point of view of the speech signal and to rely on the combination of linguistic and prosodic features. The first approach is to carry out emotion/sentiment detection using each modality (i.e., prosodic features and linguistic features) independently. The second (and less common) approach is to combine linguistic and acoustic features by aligning the transcript with the extracted prosodic features for the analysis of sentiment, as suggested by [157]. In particular, such approaches can be useful to characterize paralinguistic events such as laughs [158] or sighs [159], which are key cues of sentiment-related phenomena [160] or to disambiguate ironic expressions that are not detectable through the linguistic content.

4.3.2 Automatic Speech Recognition (ASR) Outputs and Spontaneous Speech Features

The linguistic-based sentiment analysis carried out on automatic speech transcripts has to deal with speech variability: inter-speaker and intra-speaker variability (emotion, speech style, linguistic variation, grammatical construction, badly pronounced words etc.). This variability causes two issues for oral data analysis, rarely addressed in the context of sentiment analysis.

Firstly, the performance of the ASR systems strongly depends on the background noise and quality of the recording systems. Thus, the confidence score and the various hypotheses of the ASR outputs have to be taken into account. This issue is scarcely handled in the sentiment analysis/opinion mining community [18], [73], [161] but is already handled by other communities: for example, [162] investigated the effectiveness of using more ASR hypotheses (n-best lists) to extract lexical features for the automatic analysis of the level of interest in spoken dialogs.

Secondly, even correctly transcribed, spontaneous speech features contained in the user utterance such as disfluencies, backchannels and interruptions introduce some noise into the text from the point of view of a text-based detection system. However, it should be noted that the disfluency rate in user utterances depends on contextual factors, such as the role of the speaker in the conversation [163] or the interaction context (users tend to simplify utterances and use less disfluencies in human-machine dialogs than in human-human dialogs [164]). Nevertheless, the presence of disfluency phenomena (e.g., repetitions, hesitations, etc.) disrupts the syntactical structure of the message [165] and should be taken into account in sentiment analysis methods, without ignoring the fact that disfluencies can have a strategic function in the conversation [166], [167]. In particular, [29] uses disfluencies (filled pauses, fillers, stuttering, laughter, breathing and sigh) to improve multimodal emotion recognition. The automatic detection of disfluencies can be difficult in particular in the case of edit disfluencies where the speaker corrects or alters the utterance, or abandons it entirely and starts over [168]. Existing disfluency detection systems make use of acoustic and/or lexical cues using sequence detection methods [169], such as the work we carried out on call-center data [170].

4.3.3 Turn-Taking Behavior

The interaction between a human and an ECA also corresponds to *conversational speech* and the user's turn-taking behavior can be a powerful indicator of affect and can be integrated as an additional cue for sentiment detection (e.g., what is the user's dynamic in terms of turn-taking? does the user employ short turns? does the user try to frequently interrupt the agent?). [171] provides a state of the art of social signals in turn-taking and presents some studies analyzing the role of turn-taking as an indicator of speaker's attitude: "a clash of opinions also means a clash of turn-taking" [172]. In [113], we demonstrated that dialog length and width, understood as the total number of turns and the number of turns required to provide a single information item (e.g., repetitions, asking for help, etc.), improves emotion recognition accuracy. [173] also demonstrate that conversational features can be used as a single source to reliably model user affect by predicting satisfaction ratings.

Detecting the users' turn-taking behavior can be done locally with the detection of interruptions such as in [84], which describes an acoustic turn-taking module, which can use features (speech rate or speaker turn duration) and structural features [67] (used as cues for agreement analysis in meetings). For the analysis of human-human interactions, [174] uses turn-taking features (steady conversational period) and [175] calls on an echo state network to predict timing activity.

4.4 Integrating the Temporal Constraints

Timing plays a major role in the relevance of ECA's emotional reaction to users' expression of sentiment, raising the issue of the processing time for sentiment detection methods; a relevant answer to user sentiment that occurs too late because of too long processing time in the interaction is not conceivable.

The processing time of sentiment analysis is addressed in a quite different manner in the context of opinion mining. The community—joining forces with Big Data movement–puts a slant on the development of scalable methods, that is, on methods that are able to deal with large databases and that can incrementally adapt to new incoming data [17]. From the perspective of human-agent interaction, the question is not how to deal with a large amount of data, but how to provide a quick reaction to users' sentiment through affective dialog strategies (see Section 3.2). In human-agent interaction the sentiment detection method takes as input an incoming audio stream and the unit of analysis or decision frame has to be determined according to the temporal constraints.

4.4.1 Methods with Various Levels of Complexity

Affective dialog strategies can be designed with full knowledge of the sentiment detection time, such as in [84]. In this paper, the complex rule-based approach for sentiment analysis of [102] (compositional approach) is used as an input of a long-term interaction strategy. Noting this, affective dialog strategies can use computational methods with various processing times: interaction strategies such as backchannels for short processing time detection methods, and long-term interaction strategies integrating the in-depth understanding of the interaction context for the most complex detection methods.

Methods eligible for a quick reaction time are the same as those used for large-scale databases: rule-based approaches (spotting words belonging to a sentiment lexicon) and statistical methods (see Section 3.1). Moreover, in the case of speech processing, methods relying on word spotting have the advantage of not requiring an enriched automatic transcript and can directly rely on keyword spotting, which can also reduce the processing time [176].

Machine learning methods remain interesting in terms of computational complexity for the decision step (as their cost is mainly the preliminary off-line training phase). However, machine learning methods using a complex representation of textual features can have a better performance but also a longer processing time, even at the decision step. This is also the case of rule-based methods relying on complex semantic rules.

It is worth noting that the processing time of such methods strongly depends on both the device on which the ECA is developed (smartphone, laptop, features of the computer), and the complexity of the methods.

4.4.2 Definition of the Decision Frame

Another aspect related to the ECA's reaction time is the decision frame used for sentiment analysis. Indeed, sentiment analysis can be expressed in two ways: the classification of text utterances with a predefined length or the assignment of sentiment linked expressions to sentiment-related concepts [177]. In both cases, the units of text from which the system takes the decision have to be defined. In opinion mining inputs are generally written texts of various lengths (e.g., shorter in the case of tweets and longer in the case of news articles). For longer texts various linguistic units are used: phrases, sentences, paragraphs or even documents.

When the input is the speech stream of the user, such as in human-agent interactions, the units of analysis are quite different. The choice of the units of analysis and of the decision frames have a strong impact on the reaction time of the agent. Such concern is scarcely addressed in the literature of sentiment analysis, because it is a specific issue associated with the handling of conversational speech. For example, [67] uses *spurts* (periods of speech without pauses greater than 0.5s) as the unit of analysis instead of sentences arguing that such a segmentation is easier to obtain; [178] uses speech segments corresponding to speaker turns in political debate transcripts.

The agent's reaction can be triggered by the occurrence of a word (for methods such as keyword spotting), a clause, a sentence, a speaker turn, or even several consecutive speaker turn segments.

4.5 Contextualizing the Interaction

Context is defined in very different ways in the sentiment analysis and ECA communities, and sometimes also within the communities themselves.

In sentiment analysis, context is generally understood as the domain in which an opinion is being expressed. The domain deeply influences polarity assessment, as for example the word "small" can be either positive or negative depending on the entity or the aspect of the entity being analyzed [179]. Other authors describe it as the linguistic context, that is, the syntactic structures or the collocations of words [180], or even the words and linguistic constructs that influence the analysis. For example, [149] describe contextual sentiment influencers such as negations and contrary words (e.g., however), and in [181] there is a discussion on contextual valence shifters. This context is dealt with using the sentiment analysis methods described in Section 3.1

In the ECA community, context may refer to external aspects such as the user's location or the environment in which the communication takes place (e.g., noisy versus quiet places), or it can refer to aspects internal to the user, either stable aspects such as the user's gender, age, personality and language, or varying aspects such as emotions, moods, or temporal preferences [182].

In addition, especially in works centred on social networks, sentiment analysis studies the context of the group and the social dynamics. As mentioned in [183], the opinion expressed by a user in a forum can be influenced by the opinions previously expressed by others. This may be included in ECA systems to complement the external context representation with information not only about the environment and physical location, but also about collective behavior.

Traditionally the sentiment analysis community has not paid much attention to the internal aspects that are widespread in the ECA community. For instance, [184] claimed that little research had been conducted to understand opinions in the social and geopolitical context and claim that such information would shed light on public opinions and concerns. In [185] the authors present a framework to add emotional tags to existing social network services based on recognizing emotion from different sources. This idea can be extended by incorporating the knowledge of the ECA community in the development of multimodal emotion recognizers.

To develop ECAs that incorporate affective and social features, the community has developed models of affective conversational behavior [186] and social skills that can be used to develop sophisticated user models that help to

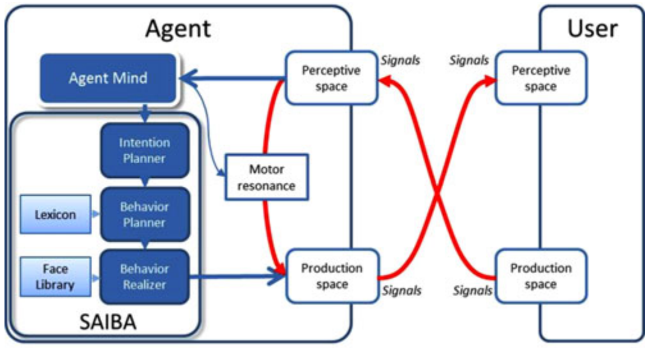

Fig. 3. Global architecture of the Greta platform [52].

disambiguate textual sentiment analysis. In Section 5 we propose additional guidelines to this aspect.

4.6 Facilitating Long-Term Relations

Although the focus of ECAs is to provide an immediate response, when the aim is to sustain long-term relations with the user, the ability of the system to learn from a long history of previous interactions gains importance. The conversational nature of human-ECA interaction is also beneficial to obtain more information from the user, who is in some cases more keen to self-disclose with an artificial interlocutor [187]. There is therefore more space for using sentiment analysis approaches to extract knowledge from large amounts of data.

To build such complex behaviors it could be beneficial to build long-term user models that include predictions of future user actions and affect. Some areas of sentiment analysis have addressed the problem of long-term predictions, specially the works related to the effect of sentiment in investments. For example, [188] studied the influence of investor sentiment's orientation on future exchange rates considering different time scopes, ranging from 1 to 60 months.

These aspects could be used by the sentiment analysis community to address the study of the volatility of positive and negative sentiments. Also ECA models for interactions sustained over long periods of time could enrich opinion mining studies with user personality models that contextualize the interpretation of their opinions and reactions.

5 Guidelines: The Case of the Greta Platform

In the previous section, we have presented the various open questions and prospects for the integration of a sentiment analysis module in an affective dialog system. The goal of this section is to make the discussion more concrete through the prism of the Greta platform. Two use cases are also discussed: a virtual agent that interacts with the visitors in a museum and a virtual recruiter training youngsters in preparation for job interviews.

5.1 The Greta Platform

Fig. 3 presents the global architecture of the Greta ECA platform [52], [189]. It is a SAIBA (situation, agent, intention, behavior, animation) architecture (an international common multimodal behavior generation framework [190]). It relies on the *Agent Mind*, which includes a representation of the agent's cognitive and socio-emotional states (its emotions, its social relationships, its goals and beliefs) and reasoning models to compute and update these features during the interaction. This module provides the communicative intentions of the agent.

In this architecture, the sentiment analysis module is dedicated to enrich agent's perceptive space as an input of Agent Mind in order to define interaction strategies. With the aim to present our guidelines for the integration of sentiment analysis in Greta platform, we propose to define different levels of interaction strategies: from *Tit for tat* interaction strategies to strategies for long-term relations.

5.1.1 Tit for Tat *Interaction Strategies*

As previously explained, a correct timing of the agent feedback is paramount and implies a trade-off between the depth of the sentiment analysis and the processing time of the module. This is especially the case for what we name here *Tit for tat* strategies, which are dedicated to trigger immediate actions such as backchanneling [191], turn-taking, and quick selection of the next agent's utterance.

The use of audio signal variations linked to emotions for *Tit for tat* interaction strategies has been already explored in the literature, e.g., [84] for backchannel triggering, and in the Greta platform during the SEMAINE project [192] (for both turn-taking management and bakchannel triggering). In the SEMAINE project, keyword spotting on user input has been also implemented through the Content Module and used to select agent's sentences. Sentiment analysis is still missing in existing human-agent dialog systems integrating *Tit for tat* interaction strategies. However, as seen in Section 4.4, using keyword spotting techniques from audio streams can provide an answer to the issue of timing by detecting keyword markers of sentiment-related phenomena from a predefined list and by using them as an input of the *Tit for tat* interaction strategies.

5.1.2 Strategies with Long-Term Goals

The agent's strategy can include not only immediate actions such as backchanneling and turn-taking but also long-term goals for the interaction such as the ones presented in Section 4.1. Section 4.2 shows that the analysis of sentiment-related phenomena can have a key role in the definition and the adaptation of the system goals, by enriching the modeling of the user's socio-emotional behavior, interest and preferences. It requires to work on a higher level that can be combined with keyword spotting to provide a richer analysis of user's behavior. Opining mining techniques for syntactic and semantic analysis make it possible to carry out a more detailed analysis of the user input (e.g., detect the source and the target of the user's attitude).

Deep semantic analysis can be carried out considering each utterance independently from the previous ones or considering the whole conversation as a significant context to disambiguate the new inputs (analysis at the pragmatic level). Langlet and Clavel [193] takes a first step towards taking into account the interaction context by analyzing the SEMAINE database of human-agent interactions [194]. In particular, we have investigated how the interaction context can be considered by modeling the speech acts and attitudes of the agent in order to interpret user's sentiment. modeling the interaction can lead to the prediction of user sentiment-related behavior. In that case, the expected behavior can be compared to the actual output of the detection system and the agent strategy can be defined according to the difference encountered.

5.1.3 Strategies to Build Long-Term Relations

As described in Section 4.6, long-term relations must be built by integrating the in-depth understanding of the interaction through the analysis of users' behavior over a prolonged usage of the system. With this aim, two main aspects have to be studied: first, how to build an incremental user model and, second, how to manage the interaction considering the model.

In the Greta platform, the information obtained from new interactions in order to build a user model is an input to the Agent mind. NLP can contribute to the user model by providing not only a cue for user emotional state detection, but also a cue for user engagement [36] or for the user's social stance (e.g., friendliness of the interpersonal circumplex axis). In order to make the user model incremental, the sentiment detection system has to integrate the various decision frames presented in Section 4.4 (clause, sentence, speaker turn, history of speaker turns) and to be merged with the outputs of the sentiment analysis module with audio and video analysis of socio-emotional behaviors (see Fig. 2)

With respect to interaction management, the incremental user model is dedicated to adjust the ECA's interaction strategy (e.g., mirroring strategy, or lead the user to more positive states, see Section 4.6). In [195], we have built a dialog manager within the Agent Mind that selects the most appropriate system objectives and responses according to the user model. In order to deal with this more informative user model, we have still to study how to establish accurate connections between the intent planner and the behavior planner.

5.2 Grounding Prospects on Use Cases

5.2.1 Talking with Museum Visitors: A First Proof of Concept

A first integration of sentiment analysis in human-agent interaction is provided in the A1:1 French project, which aims at developing an interaction between a virtual human-sized agent and museum visitors. The agent plays the role of another visitor discussing the exhibit. The focus is put on fostering user *engagement* as defined by [196]: "the value that a participant in an interaction attributes to the goal of being together with the other participant(s) and of continuing the interaction". We consider two aspects:

- to define the agent's interaction strategies considering user's sentiment-related phenomena (here, user's appreciations) in order to foster users engagement,
- to detect user's sentiment-related phenomena by relying on the interaction context (here, the model of agent's utterance).

The first aspect follows the socio-affective interaction goals presented in Section 4.1 by investigating an alignment strategy. In particular, we focus on the role of *verbal* alignment in the engagement process [197]. We are developing agent's strategies consisting on the agent's alignment or disalignment to user's attitudes. The definition of attitudes provided by Martin and White [36] and the underlying appraisal model (see Section 2.1.1) is especially relevant in this context. It allows us to focus on the detection of user's *appreciation* (in contrast to the judgment and affect components of the model) of museum artwork and on the alignment of the agent's *appreciation* to the user's appreciation [198]. We have carried out a user perceptive study and first results show that sharing appreciations with the users improves their perception of the agent's engagement and believability.

The second aspect specifies the use of the interaction context in order to improve the detection of sentiment-related phenomena (see Section 4.5). The model investigated in [193] and introduced in Section 5.1.2 has been implemented in [199] to analyze a specific sentiment-related phenomenon: the user's likes and dislikes. Based on Martin and White theoretical framework [36], we build a computational model of user's like and dislike.

The system illustrates how the detection of user's likes and dislikes can be enhanced by modeling the agent's dialog acts and displayed attitudes. We used rules and local grammars that have proven to be especially relevant for the integration of a formal model of the agents in the detection system. The rules also allow identifying the target of user's like and dislike as advised by Martin and White's model [36]. As a first step, the identification of the target type focuses on determining if the target of user's like is the agent—which could be linked to the analysis of agent likeability—or another entity. This pioneering version of the system shows encouraging results for the different tasks performed by the system that concern the detection of relevant like/dislike expressions (substantial agreement with a Fleiss kappa at 0.61), the polarity assignment (almost perfect agreement with a Fleiss kappa at 0.84) and the identification of the target type (53 percent agreement between the reference and the system output).

Besides, the potential targets of appreciation within the museum application could be predefined according to the list of museum artworks. For example, if we know that the agent is questioning the user on their appreciation of a given artwork, this information can be modelled and used by the rules designed to detect user's appreciation and its target to perform a more accurate detection of the user's appreciation and its target. This example underlines the importance of considering jointly sentiment analysis and dialog modeling, an approach that is not commonly addressed in the human-agent interaction community.

5.2.2 Training Young People for Job Interviews: Prospects

Another interesting application is the one provided by the EU TARDIS project[5] which aims to develop a serious game to train youngsters for job interviews [200]. Each training session takes place between the human interviewee and a virtual recruiter. We have developed a first version of the ECA that plays the role of the recruiter using the Greta platform [201]. The multimodal response of the agent is modulated to achieve an adaptive dialog strategy (to challenge or comfort the user). This strategy is tailored to the desired difficulty of the interview and the anxiety experienced by the interviewee after each question. The agent therefore tries to change the user's emotion through behavioral choices and expressive manifestations [195]. For behavioral choices, it selects a degree of difficulty of the question to be posed in two dimensions: complexity of the question and openness of the expected response. With regard to expressive manifestation, it renders different interpersonal attitudes through its facial expressions, gestures, poses, and the phrasing of the question.

Thus, as suggested in Section 4.1.3, we have envisioned socio-emotional strategies not only as a way of engaging the user or making the agent more likeable, but also as a key element to drive the ECA's behavior.

As a matter of fact, socio-emotional strategies are used in our system both to guide the interview and to give feedback to the users about their skills. Also, we have considered different time spans for the duration of the negative emotionality experienced by the interviewees (Section 4.4), and studied the overall tendency of anxiety levels during the interview.

We have conducted a perceptive study with 110 individuals and a virtual recruiter ECA endowed with our model. The results show that the agents multimodal behavior successfully conveyed the expected social attitudes including different socio-affective objectives such as wanting to hire, fail or destabilize the interviewee [201], [195].

We believe that this model can be improved by considering the opinion mining perspective, following the proposals indicated in the previous sections in order to build more sophisticated policies. We present here some ideas for the future integration of sentiment analysis for this use case.

An interesting improvement would be to include a cognitive appraisal approach (see Section 2.2.1) to model emotions. In the current version of the system, user anxiety is computed using paralinguistic cues [200]. It could be interesting to use linguistic cues and NLP methods and to rely on the modeling of the context (see Section 4.5) in order to improve the detection of the anxiety of the user. We consider so far that the cause of the change in the user's anxiety is the agent's previous question (colored with a hostile/friendly attitude), but we could go further by defining a richer cause that would also include aspects inner to the interviewee by considering, for example, relevant psychological issues identified by the human resources literature [202].

NLP methods presented in Section 3.1 can be used to analyze user utterances and to build user models, as it is described in Section 4.6. Additionaly, we propose to use knowledge about the user experience in doing interviews (real and previous interactions with the game), the user curriculum (whether he/she is better prepared, has more experience in the area, etc.) and demographic data. Using an appraisal detection system based on NLP, it would be possible to detect coping mechanisms that are not only related to emotionality, but also to strategies specifically related to the application domain. For example, to identify variations on possible behaviors for impression management [203]. This would allow the ECA to provide more accurate responses depending on the adequacy of the strategies employed by the user in each situation.

5. http://www.tardis-project.eu

With respect to the multimodal integration, we propose to consider aligning the lexical information with the prosodic features employed (see Fig. 2). Currently the system takes into account information about the tone of voice and other acoustic features [200], but disfluencies can be particularly useful for emotion detection [204], especially in the job interview domain [205].

Finally, to facilitate long-term relations we propose to exploit the ludic aspects of serious games. User-adaptation and the dynamic nature of the socio-affective interaction in Tardis facilities its usage for long periods of time while obtaining different user experiences in each session.

6 Conclusion

Studies on sentiment analysis in the literature are focused on the perspective of a single community, and usually based on optimizing the features, algorithms and methods used for the distinction between positive and negative sentiment-related phenomena, without accurately defining the different sentiment-related terms. The literature on human-computer interaction has been called up and compared with the literature on sentiment analysis with the aim of developing interaction models that consider sentiment analysis for socio-emotional ECAs.

The paper has proposed various avenues for the integration of sentiment analysis in face-to-face human-agent interactions. First, it is necessary to choose an adequate psycho-linguistic model to characterize human-agent affective dialogs. Second we have to juggle with the use of semantic rules and machine learning methods in order to integrate: the multimodal nature of sentiment-related phenomena, the variability of temporal and decision frames, varying levels of complexity required by the timing constraint of the interaction, and the heterogeneity of the contextual information. Another big issue is to deal with ASR outputs. Third, we have to investigate sentiment analysis methods as an input and an output of both short and long-term strategies.

We have addressed all these points providing guidelines that cover the new insights that the studied communities can provide to each other. Finally, to show the practicability of the proposed guidelines, we have described how to particularize them to two different target applications within the Greta ECA platform: a virtual recruiter for job interview training and a virtual agent in a museum.

Acknowledgments

The authors would like to thank Catherine Pelachaud, Caroline Langlet, Magalie Ochs and Sabrina Campano for valuable insights and suggestions and Rachel Bawden for sharing her expertise on English language. This work has been supported by the French collaborative project A1:1, the european project ARIA-VALUSPA, and performed within the Labex SMART (ANR-11-LABX-65) supported by French state funds managed by the ANR within the Investissements d'Avenir programme under reference ANR-11-IDEX-0004-02.

References

[1] M. El-Bèze, A. Jackiewicz, and S. Hunston, "Opinions, sentiments et jugements d'évaluation," *Traitement Automatique des Langues*, vol. 51, no. 3, pp. 7–17, 2010.

[2] H.-J. Jang, J. Sim, Y. Lee, and O. Kwon, "Deep sentiment analysis: Mining the causality between Personality-value- attitude for analyzing business ads in social media," *Expert Syst. Appl.*, vol. 40, pp. 7492–7503, 2013.

[3] S. Sommer, A. Schieber, K. Heinrich, and A. Hilbert, "What is the conversation about? A Topic-model-based approach for analyzing customer sentiments in twitter," *Int. J. Intell. Inf. Technol.*, vol. 8, pp. 10–25, 2012.

[4] M. Mostafa, "More than words: Social networks' text mining for consumer brand sentiments," *Expert Syst. Appl.*, vol. 40, pp. 4241–4251, 2013.

[5] J. Smailovic, M. Grcar, N. Lavrac, and M. Znidarsic, "Predictive sentiment analysis of tweets: A stock market application," in *Proc. 3rd Int. Conf. Human-Comput. Interaction Knowl. Discovery Complex, Unstructured, Big Data*, 2013, vol. 7947, p. 442.

[6] P. Suignard, "NaviQuest: Un outil pour naviguer dans une base de questions posées à un Agent Conversationnel," in *Proc. Workshop sur les Agents Conversationnels Animés*, Lille, 2010.

[7] J. Klatt, S. Marsella, and N. Krämer, "Negotiations in the context of AIDS prevention: An agent-based model using theory of mind," in *Proc. 11th Int. Conf. Intell. Virtual Agents*, 2011, pp. 209–215.

[8] M. Turunen, J. Hakulinen, O. Sthl, B. Gamback, P. Hansen, M. C. Rodríguez Gancedo, R. S. de la Cámara, C. Smith, D. Charlton, and M. Cavazza, "Multimodal and mobile conversational health and fitness companions," *Comput. Speech Language*, vol. 25, pp. 192–209, 2011.

[9] Schlegel, D. Grandjean, K. R. Scherer, and K. R. Scherer, "Emotion recognition: Unidimensional ability or a set of modality-and emotion-specific skills?" *Personality Individual Differences*, vol. 53, pp. 16–21, 2012.

[10] K. R. Scherer, "The nature and dynamics of relevance and valence appraisals: Theoretical advances and recent evidence," *Emotion Rev.*, vol. 5, no. 2, pp. 150–162, 2013.

[11] K. R. Scherer, M. Mortillaro, and M. Mehu, "Understanding the mechanisms underlying the production of facial expression of emotion: A componential perspective," *Emotion Rev.*, vol. 5, no. 1, pp. 47–53, 2013.

[12] A. Moors, P. Ellsworth, K. R. Scherer, and N. H. Frijda, "Appraisal theories of emotion: State of the art and future development," *Emotion Rev.*, vol. 5, pp. 119–124, 2013.

[13] N. Dael, M. Mortillaro, K. R. Scherer, and K. R. Scherer, "Emotion expression in body action and posture," *Emotion*, vol. 12, no. 5, pp. 1085–1101, 2012.

[14] K. Scherer, "What are emotions? And how can they be measured?" *Soc. Sci. Inf.*, vol. 44, no. 4, pp. 695–729, 2005.

[15] M. D. Munezero, C. Suero Montero, E. Sutinen, and J. Pajunen, "Are they different? affect, feeling, emotion, sentiment, and opinion detection in text," *IEEE Trans. Affective Comput.*, vol. 5, no. 2, pp. 101–111, Apr.–Jun. 2014.

[16] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 2, nos. 1/2, pp. 1–135, 2008.

[17] M. Tsytsarau and T. Palpanas, "Survey on mining subjective data on the web," *Data Mining Knowl. Discovery*, vol. 24, no. 3, pp. 478–514, 2012.

[18] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," *IEEE Intell. Syst.*, vol. 28, no. 2, pp. 15–21, Mar.–Apr. 2013.

[19] A. Montoyo, P. Martínez-Barco, and A. Balahur, "Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments," *Decision Support Syst.*, vol. 53, no. 4, pp. 675–679, Mar.–Apr. 2012.

[20] B. Schuller and A. Batliner, *Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing*. Hoboken, NJ, USA: Wiley, 2013.

[21] B. Schuller, S. Steidl, and A. Batliner, "Introduction to the special issue on paralinguistics in naturalistic speech and language," *Comput. Speech Language*, vol. 27, no. 1, pp. 1–3, 2013.

[22] K. Karpouzis, E. Andre, and A. Batliner, "Special issue on Emotion-aware natural interaction," *Adv. Human-Computer Interaction*, vol. 2010, 2010.

[23] J. Pittermann, A. Pittermann, and W. Minker, "Emotion recognition and adaptation in spoken dialogue systems, " *Int. J. Speech Technol.*, vol. 13, pp. 49–60, 2010.

[24] T. H. Bui, M. Poel, A. Nijholt, and J. Zwiers, "A tractable hybrid DDNPOMDP approach to affective dialogue modeling for probabilistic frame-based dialogue systems," *Natural Language Eng.*, vol. 15, no. 2, pp. 273–307, 2009.

[25] Z. Callejas, D. Griol, and R. López-Cózar, "Predicting user mental states in spoken dialogue systems," *EURASIP J. Adv. Signal Process.*, vol. 2011, no. 1, p. 6, 2011.
[26] S. D'Mello and A. Graesser, "AutoTutor and affective autotutor: Learning by talking with cognitively and emotionally intelligent computers that talk back," *ACM Trans. Interact. Intell. Syst.*, vol. 2, no. 4, pp. 1–39, 2013.
[27] U. Krcadinac, P. Pasquier, J. Jovanovic, and V. Devedzic, "Synesketch: An open source library for sentence-based emotion recognition," *IEEE Trans. Affective Comput.*, vol. 4, no. 3, pp. 312–325, Jul.–Sep. 2013.
[28] M. Ishizuka, "Textual affect sensing and affective communication," in *Proc. IEEE 11th Int. Conf. Cognitive Informat. Cognitive Comput.*, 2012, pp. 2–3.
[29] M. Wollmer, F. Weninger, T. Knaup, B. Schuller, S. Congkai, K. Sagae, and L. Morency, "Youtube movie reviews: Sentiment analysis in an audio-visual context," *IEEE Intell. Syst.*, vol. 28, no. 3, pp. 46–53, May/Jun. 2013.
[30] I. Perikos and I. Hatzilygeroudis, *Recognizing Emotion Presence in Natural Language Sentences*, series Communications in Computer and Information Science. Berlin, Germany: Springer, 2013, no. 384, pp. 30–39.
[31] A. Neviarouskaya, H. Prendinger, and M. Ishizuka, "User study on AffectIM, an avatar-based Instant Messaging system employing rule-based affect sensing from text," *Int. J. Human-Comput. Stud.*, vol. 68, no. 7, pp. 432–450, 2010.
[32] C. Strapparava and R. Mihalcea, "Semeval-2007 task 14: Affective text," in *Proc. 4th Int. Workshop Semantic Eval.*, 2007, pp. 70–74.
[33] C.-C. Chan and K. J. Liszka, *Application of Rough Set Theory to Sentiment Analysis of Microblog Data*, series Intelligent Systems Reference Library. Berlin, Germany: Springer, 2013, no. 43, pp. 185–202.
[34] A. Ortigosa, J. M. Martín, and R. M. Carro, "Sentiment analysis in facebook and its application to E-learning," *Comput. Human Behavior*, vol. 31, pp. 527–541, 2014.
[35] S.-M. Kim and E. Hovy, "Determining the sentiment of opinions," in *Proc. 20th Int. Conf. Comput. Linguistics*, 2004, pp. 1367.
[36] J. R. Martin and P. R. White, *The Language of Evaluation. Appraisal in English*. New York, NY, USA: Palgrave Macmillan, 2005.
[37] C. Kerbrat-Orecchioni, *L'énonciation: De La Subjectivité Dans Le Langage*. Paris, France: Armand colin, 2009.
[38] P. Charaudeau, *Grammaire Du Sens Et De L'expression*. New York, NY, USA: Hachette, 1992.
[39] E. Benveniste, "L'appareil formel de l'énonciation," in *Problèmes De Linguistique Générale II*. Paris, France: Gallimard, 1974, pp. 79–88.
[40] J. Wiebe, T. Wilson, and C. Cardie, "Annotating expressions of opinions and emotions in language," *Language Resources Eval.*, vol. 39, nos. 2/3, pp. 165–210, 2005.
[41] E. Breck, Y. Choi, and C. Cardie, "Identifying expressions of opinion in context," in *Proc. Int. Joint Conf. Artif. Intell.*, 2007, pp. 2683–2688.
[42] G. Johnson, "Theories of emotion. Internet Encyclopaedia of Philosophy: A Peer-Reviewed Academic Resource," 2009.
[43] J. Deigh, *Concepts of Emotions in Modern Philosophy and Psychology*. London, U.K.: Oxford Univ. Press, 2009.
[44] R. Plutchik, *Emotions and Life: Perspectives from Psychology, Biology, and Evolution*. Washington, DC, USA: American Psychological Association, 2003.
[45] S. Oosterwijk, K. A. Lindquist, E. Anderson, R. Dautoff, Y. Moriguchi, and L. F. Barrett, "States of mind: Emotions, body feelings, and thoughts share distributed neural networks," *NeuroImage*, vol. 62, pp. 2110–2128, 2012.
[46] K. Oatley and P. N. Johnson-Laird, "Cognitive approaches to emotions," *Trends Cognitive Sci.*, vol. 18, no. 3, pp. 134–140, 2014.
[47] C. Beedie, P. Terry, and A. Lane, "Distinctions between emotion and mood," *Cognition Emotion*, vol. 19, no. 6, pp. 847–878, 2005.
[48] L. Ring, B. Barry, K. Totzke, and T. Bickmore, "Addressing loneliness and isolation in older adults: Proactive affective agents provide better support," in *Proc. Humaine Assoc. Conf. Affective Comput. Intell. Interaction*, 2013, pp. 61–66.
[49] S. Schnall, *Affect, Mood and Emotions*. Amsterdam, The Netherlands: Elsevier, 2010, pp. 544–548.
[50] J. C. Veilleux, K. D. Skinner, E. D. Reese, and J. A. Shaver, "Negative affect intensity influences drinking to cope through facets of emotion dysregulation," *Personality Individual Differences*, vol. 59, pp. 96–101, 2014.
[51] R. J. Thompson, J. Mata, S. M. Jaeggi, M. Buschkuehl, J. Jonides, and I. H. Gotlib, "Concurrent and prospective relations between attention to emotion and affect intensity: An experience sampling study," *Emotion*, vol. 11, pp. 1489–1494, 2011.
[52] M. Ochs, K. Prepin, and C. Pelachaud, "From emotions to interpersonal stances: Multi-levels analysis of smiling virtual characters," in *Proc. Affective Comput. Intell. Interaction Conf.*, 2013, pp. 258–263.
[53] A. Vinciarelli, M. Pantic, and H. Bourlard, "Social signal processing: Survey of an emerging domain," *Image Vis. Comput.*, vol. 27, no. 12, pp. 1743–1759, 2009.
[54] O. C. S. Tzeng, R. Hoosain, and C. E. Osgood, "Cross-cultural componential analysis on affect attribution of emotion terms," *J. Psycholinguistic Res.*, vol. 16, no. 5, pp. 443–465, 1987.
[55] J. A. Russell, "A circumplex model of affect," *J. Personality Social Psychology*, vol. 39, pp. 1161–1178, 1980.
[56] P. Ekman, "Basic emotions," in *Handbook of Cognition and Emotion*. New York, NY, USA: Wiley, 1999.
[57] C. E. Izard, "The face of emotion," New York: Appleton-Century-Crofts, 1971.
[58] C. M. Whissel, *The Dictionary of Affect in Language. Emotion: Theory, Research, and Experience*, R. Plutchik and H. Kellerman, Eds. New York, NY, USA: Academic, 1989.
[59] A. Ortony, *The Cognitive Structure of Emotions*. Cambridge, U.K.: Cambridge Univ. Press, 1990.
[60] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: Sentiment classification using machine learning techniques," in *Proc. Conf. Empirical Methods Natural Language Process.*, 2002, vol. 10, pp. 79–86.
[61] R. Lavalley, C. Clavel, M. El-Bèze, and P. Bellot, "Finding topic-specific strings in text categorization and opinion mining contexts," in *Proc. Int. Conf. Data Mining*, 2010, pp. 38–44.
[62] M.-T. Martín-Valdivia, E. Martínez-Cámara, J.-M. Perea-Ortega, and L. A. Ureña López, "Sentiment polarity detection in Spanish reviews combining supervised and unsupervised approaches," *Expert Syst. Appl.*, vol. 40, pp. 3934–3942, 2013.
[63] P. Turney, "Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews," in *Proc. 40th Annu. Meeting Comput. Linguistics*, 2002, pp. 417–424.
[64] A. Osherenko and E. André, "Differentiated semantic analysis in lexical affect sensing," in *Proc. Conf. Affective Comput. Intell. Interaction Workshops*, 2009, pp. 1–6.
[65] M. Thelwall, K. Buckley, and G. Paltoglou, "Sentiment strength detection for the social web," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 63, no. 1, pp. 163–173, Jan. 2012.
[66] A. Neviarouskaya, H. Prendinger, and M. Ishizuka, "Recognition of affect, judgment, and appreciation in text," in *Proc. 23rd Int. Conf. Comput. Linguistics*, 2010, pp. 806–814.
[67] M. Galley and K. McKeown, "Identifying agreement and disagreement in conversational speech: Use of Bayesian networks to model pragmatic dependencies," in *Proc. 42nd Annu. Meeting Assoc. Comput. Linguistics*, 2004, p. 669.
[68] E. Spertus, "Smokey: Automatic recognition of hostile messages," in *Proc. 14th Nat. Conf. Artif. Intell.*, 1997, pp. 1058–1065.
[69] V. Francisco and P. Gervás, "Emotag: An approach to automated markup of emotions in texts," *Comput. Intell.*, vol. 29, no. 4, pp. 680–721, 2013.
[70] M. A. M. Shaikh, H. Prendinger, and M. Ishizuka, "A linguistic interpretation of the occ emotion model for affect sensing from text," in *Affective Information Processing*. New York, NY, USA: Springer, 2009, pp. 45–73.
[71] R. Harré and W. G. Parrott, *The Emotions: Social, Cultural and Biological Dimensions*. Newbury Park, CA, USA: Sage, 1996.
[72] K. Bloom, N. Garg, and S. Argamon, "Extracting appraisal expressions," in *Proc. Human Language Technol./North Amer. Assoc. Comput. Linguists*, Apr. 2007, pp. 308–315.
[73] C. Clavel, G. Adda, F. Cailliau, M. Garnier-Rizet, A. Cavet, G. Chapuis, S. Courcinous, C. Danesi, A.-L. Daquo, M. Deldossi, S. Guillemin-Lanne, M. Seizou, and P. Suignard, "Spontaneous speech and opinion detection: Mining call-centre transcripts," *Language Resources Eval.*, vol. 47, no. 4, pp. 1089–1125, 2013.
[74] J. Wiebe, T. Wilson, and C. Cardie, "Annotating expressions of opinions and emotions in language," *Language Resources Eval.*, vol. 39, nos. 2/3, pp. 165–210, 2005.
[75] A. Schmitt, T. Polzehl, and W. Minker, "Facing reality: Simulating deployment of anger recognition in IVR systems," in *Spoken Dialogue Systems for Ambient Environments*, ser. Lecture Notes in Computer Science, G. G. Lee, J. Mariani, W. Minker, and S. Nakamura, Eds. New York, NY, USA: Springer, 2010, no. 6392, pp. 122–131.

[76] B. Schuller, A. Batliner, S. Steidl, and D. Seppi, “Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge,” *Speech Commun.*, vol. 53, no. 9, pp. 1062–1087, 2011.
[77] H. Boril, S. Omid Sadjadi, T. Kleinschmidt, and J. H. L. Hansen, “Analysis and detection of cognitive load and frustration in drivers’ speech,” in *Proc. 11th Annu. Conf. Int. Speech Commun.*, 2010, pp. 502–505.
[78] D. J. Litman and K. Forbes-Riley, “Recognizing student emotions and attitudes on the basis of utterances in spoken tutoring dialogues with both human and computer tutors,” *Speech Commun.*, vol. 48, no. 5, pp. 559–590, 2006.
[79] M. Sellers, “Toward a comprehensive theory of emotion for biological and artificial agents,” *Biol. Inspired Cognitive Archit.*, vol. 4, pp. 3–26, 2013.
[80] R. W. Picard, *Affective Computing*. Cambridge, MA, USA: MIT Press, 2000.
[81] A. Batliner, S. Steidl, C. Hacker, and E. Nöth, “Private emotions versus social interaction: A data-driven approach towards analysing emotion in speech,” *User Model. User-Adapted Interaction*, vol. 18, nos. 1/2, pp. 175–206, 2007.
[82] L. Zhang, J. A. Barnden, R. J. Hendley, M. G. Lee, A. M. Wallington, and Z. Wen, “Affect detection and metaphor in E-drama,” *Int. J. Continuing Eng. Educ. Life-Long Learn.*, vol. 18, no. 2, p. 234, 2008.
[83] S. Yildirim, S. Narayanan, and A. Potamianos, “Detecting emotional state of a child in a conversational computer game,” *Comput. Speech Language*, vol. 25, no. 1, pp. 29–44, 2011.
[84] C. Smith, N. Crook, S. Dobnik, and D. Charlton, “Interaction strategies for an affective conversational agent,” *Presence: Teleoperators Virtual Environments*, vol. 20, no. 5, pp. 395–411, 2011.
[85] J. Vallverdú and D. Casacuberta, Eds., *Handbook of Research on Synthetic Emotions and Sociable Robotics: New Applications in Affective Computing and Artificial Intelligence*. Hershey, PA, USA: IGI Global, 2009.
[86] G. L. Clore and J. Palmer, “Affective guidance of intelligent agents: How emotion controls cognition,” *Cognitive Syst. Res.*, vol. 10, no. 1, pp. 21–30, 2009.
[87] J.-W. Yoon and S.-B. Cho, “An intelligent synthetic character for smartphone with Bayesian networks and behavior selection networks,” *Expert Syst. Appl.*, vol. 39, no. 12, pp. 11284–11292, 2012.
[88] M. S. El-Nasr, J. Yen, and T. R. Ioerger, “FLAMEFuzzy logic adaptive model of emotions,” *Autonomous Agents Multi-Agent Syst.*, vol. 3, no. 3, pp. 219–257, 2000.
[89] E. M. Eisman, V. López, and J. L. Castro, “Controlling the emotional state of an embodied conversational agent with a dynamic probabilistic fuzzy rules based system,” *Expert Syst. Appl.*, vol. 36, no. 6, pp. 9698–9708, 2009.
[90] J. Gratch, S. Marsella, and P. Petta, “Modeling the cognitive antecedents and consequences of emotion,” *Cognitive Syst. Res.*, vol. 10, no. 1, pp. 1–5, 2009.
[91] R. P. Marinier III, J. E. Laird, and R. L. Lewis, “A computational unification of cognitive behavior and emotion,” *Cognitive Syst. Res.*, vol. 10, no. 1, pp. 48–69, 2009.
[92] S. C. Marsella and J. Gratch, “EMA: A process model of appraisal dynamics,” *Cognitive Syst. Res.*, vol. 10, no. 1, pp. 70–90, 2009.
[93] R. Reisenzein, E. Hudlicka, M. Dastani, J. Gratch, K. V. Hindriks, E. Lorini, and J.-J. C. Meyer, “Computational modeling of emotion: Toward improving the inter- and intradisciplinary exchange,” *IEEE Trans. Affective Comput.*, vol. 4, no. 3, pp. 246–266, Jul.–Sep. 2013.
[94] P. A. Jaques and R. M. Vicari, “A BDI approach to infer student’s emotions in an intelligent learning environment,” *Comput. Educ.*, vol. 49, no. 2, pp. 360–384, 2007.
[95] E. Hudlicka, “Affective bica: Challenges and open questions,” *Biol. Inspired Cognitive Archit.*, vol. 7, pp. 98–125, 2014.
[96] S. Baccianella, A. Esuli, and F. Sebastiani, “Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining,” in *Proc. Language Resources Eval. Conf.*, Valletta, Malta, 2010, pp. 2200–2204.
[97] C. Strapparava and A. Valitutti, “WordNet affect: An affective extension of WordNet,” in *Proc. 4th Int. Conf. Language Resources Eval.*, 2004, pp. 1083–1086.
[98] A. Moreo, J. L. Castro, and J. M. Zurita, “Handling context in lexicon-based sentiment analysis,” in *Advances in Computational Intelligence*, S. Greco, B. Bouchon-Meunier, G. Coletti, M. Fedrizzi, B. Matarazzo, and R. R. Yager, Eds. New York, NY, USA: Springer, 2012, no. 298, pp. 245–254.
[99] A. Reyes and P. Rosso, “On the difficulty of automatically detecting irony: beyond a simple case of negation,” *Knowl. Inf. Syst.*, vol. 40, no. 3, pp. 595–614, 2014.
[100] K. Ahmad, G. Budin, U. Wien, A. Devitt, S. Glucksberg, G. Heyer, U. Leipzig, M. T. Musacchio, M. T. Pazienza, M. Rogers, C. Vogel, and Y. Wilks, *Affective Computing and Sentiment Analysis: Emotion, Metaphor and Terminology*, vol. 45. New York, NY, USA: Springer, 2011.
[101] J. W. Pennebaker, M. E. Francis, and R. J. Booth, *Linguistic Inquiry and Word Count LIWC2001*. Hillsdale, NJ, USA: Erlbaum, 2001.
[102] K. Moilanen and S. Pulman, “Sentiment composition,” in *Proc. Int. Conf. Recent Adv. Natural Language Process.*, 2007, pp. 378–382.
[103] N. Malandrakis, A. Potamianos, E. Iosif, and S. Narayanan, “Emotiword: Affective lexicon creation with application to interaction and multimedia data,” in *Computational Intelligence for Multimedia Understanding*. New York, NY, USA: Springer, 2012, pp. 30–41.
[104] A. Neviarouskaya, H. Prendinger, and M. Ishizuka, “SentiFul: A lexicon for sentiment analysis,” *IEEE Trans. Affective Comput.*, vol. 2, no. 1, pp. 22–36, Jan. 2011.
[105] L. Zhang, “Exploration of affect sensing from speech and metaphorical text,” in *Learning by Playing. Game-Based Education System Design and Development*. New York, NY, USA: Springer, 2009, pp. 251–262.
[106] Y. Choi and C. Cardie, “Adapting a polarity lexicon using integer linear programming for domain-specific sentiment classification,” in *Proc. Conf. Empirical Methods Natural Language Process.*, 2009, pp. 590–598.
[107] A. Kaur and V. Gupta, “A survey on sentiment analysis and opinion mining techniques,” *J. Emerging Technol. Web Intell.*, vol. 5, no. 4, pp. 367–371, 2013.
[108] W. Medhat, A. Hassan, and H. Korashy, “Sentiment analysis algorithms and applications: A survey,” *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, 2014.
[109] Y. Choi, C. Cardie, E. Riloff, and S. Patwardhan, “Identifying sources of opinions with conditional random fields and extraction patterns,” in *Proc. Conf. Human Language Technol. Empirical Methods Natural Language Process.*, 2005, pp. 355–362.
[110] N. Jakob and I. Gurevych, “Extracting opinion targets in a single- and cross-domain setting with conditional random fields,” in *Proc. Conf. Empirical Methods Natural Language Process.*, 2010, pp. 1035–1045.
[111] K. Khan, B. Baharudin, A. Khan, and A. Ullah, “Mining opinion components from unstructured reviews: A review,” *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 26, no. 11, pp. 258–275, 2014.
[112] M. A. Russell, *Mining the Social Web*. Newton, MA, USA: O’Reilly, 2011.
[113] Z. Callejas and R. López-Cózar, “Influence of contextual information in emotion annotation for spoken dialogue systems,” *Speech Commun.*, vol. 50, no. 5, pp. 416–433, 2008.
[114] K. Hook, “Affective loop experiences: Designing for interactional embodiment,” *Philosophical Trans. Roy. Soc. B: Biol. Sci.*, vol. 364, no. 1535, pp. 3585–3595, 2009.
[115] D. Griol, J. M. Molina, and Z. Callejas, “Modeling the user state for context-aware spoken interaction in ambient assisted living,” *Appl. Intell.*, vol. 40, pp. 1–23, 2014.
[116] Y. Wilks, R. Catizone, S. Worgan, and M. Turunen, “Some background on dialogue management and conversational speech for dialogue systems,” *Comput. Speech Language*, vol. 25, no. 2, pp. 128–139, 2011.
[117] T. H. Bui, J. Zwiers, M. Poel, and A. Nijholt, “Affective dialogue management using factored POMDPs,” in *Interactive Collaborative Information Systems*, series Studies in Computational Intelligence, R. Babuska and F. C. A. Groen, Eds. New York, NY, USA: Springer, 2010, no. 281, pp. 207–236.
[118] M. Skowron, S. Rank, M. Theunis, and J. Sienkiewicz, “The good, the bad and the neutral: Affective profile in dialog System-user communication,” in *Affective Computing and Intelligent Interaction*, series Lecture Notes in Computer Science, S. D’Mello, A. Graesser, B. Schuller, and J.-C. Martin, Eds. New York, NY, USA: Springer, 2011, no. 6974, pp. 337–346.
[119] R. Beale and C. Creed, “Affective interaction: How emotional agents affect users,” *Int. J. Human-Comput. Stud.*, vol. 67, pp. 755–776, 2009.
[120] S. Campano and N. Sabouret, “A socio-emotional model of impoliteness for non-player characters,” in *Proc. 3rd Int. Conf. Affective Comput. Intell. Interaction Workshops*, 2009, pp. 1–7.

[121] S. Benford, C. Greenhalgh, G. Giannachi, B. Walker, J. Marshall, and T. Rodden, "Uncomfortable interactions," in *Proc. ACM Annu. Conf. Human Factors Comput. Syst.*, 2012, pp. 2005–2014.

[122] N. Vannini, S. Enz, M. Sapouna, D. Wolke, S. Watson, S. Woods, K. Dautenhahn, L. Hall, A. Paiva, E. André, R. Aylett, and W. Schneider, "FearNot!: A computer-based anti-bullying-programme designed to foster peer intervention," *Eur. J. Psychol. Educ.*, vol. 26, no. 1, pp. 21–44, 2011.

[123] H. Campos, J. Campos, C. Martinho, and A. Paiva, "Virtual agents in conflict," in *Intelligent Virtual Agents*, series Lecture Notes in Computer Science, Y. Nakano, M. Neff, A. Paiva, and M. Walker, Eds. New York, NY, USA: Springer, 2012, no. 7502, pp. 105–111.

[124] K. Boehner, R. DePaula, P. Dourish, and P. Sengers, "How emotion is made and measured," *Int. J. Human-Comput. Stud.*, vol. 65, no. 4, pp. 275–291, 2007.

[125] C. Schlett and R. Ziegler, "Job emotions and job cognitions as determinants of job satisfaction: The moderating role of individual differences in need for affect," *J. Vocational Behavior*, vol. 84, pp. 74–89, 2014.

[126] T. W. Bickmore and R. W. Picard, "Establishing and maintaining long-term human-computer relationships," *ACM Trans. Comput.-Human Interaction*, vol. 12, no. 2, pp. 293–327, 2005.

[127] K. Forbes-Riley, M. Rotaru, and D. J. Litman, "The relative impact of student affect on performance models in a spoken dialogue tutoring system," *User Model. User-Adapted Interaction*, vol. 18, no. 1, pp. 11–43, 2008.

[128] N. C. Kramer, S. Eimler, A. von der Ptten, and S. Payr, "Theory of companions: What can theoretical models contribute to applications and understanding of human-robot interaction?" *Appl. Artif. Intell.*, vol. 25, no. 6, pp. 474–502, 2011.

[129] T. Bickmore, D. Schulman, and L. Yin, "Maintaining engagement in long-term interventions with relational agents," *Appl. Artif. Intell.*, vol. 24, no. 6, pp. 648–666, 2010.

[130] B. C. Stahl, N. McBride, K. Wakunuma, and C. Flick, "The empathic care robot: A prototype of responsible research and innovation," *Technol. Forecasting Social Change*, vol. 84, pp. 74–85, 2014.

[131] R. Looije, M. Neerincx, and F. Cnossen, "Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors," *Int. J. Human Comput. Stud.*, vol. 68, pp. 386–397, 2010.

[132] C. N. Moridis and A. A. Economides, "Affective learning: Empathetic agents with emotional facial and tone of voice expressions," *IEEE Trans. Affective Comput.*, vol. 3, no. 3, pp. 260–272, Jul.–Sep. 2012.

[133] S. W. McQuiggan and J. C. Lester, "Modeling and evaluating empathy in embodied companion agents," *Int. J. Human-Comput. Stud.*, vol. 65, pp. 348–360, 2007-04.

[134] F. Nothdurft, G. Bertrand, H. Lang, and W. Minker, "Adaptive explanation architecture for maintaining human-computer trust," in *Proc. IEEE 36th Annu. Comput. Softw. Appl. Conf.*, 2012, pp. 176–184.

[135] F. Nothdurft, H. Lang, M. Klepsch, and W. Minker, "Bases of human-computer trust and explanations," in *Proc. Extended Abstracts Human Factors Comput. Syst.*, 2013, pp. 1737–1742.

[136] D. G. Zoraida Callejas, "A framework for the assessment of synthetic personalities according to user perception," *Int. J. Human-Comput. Stud.*, vol. 72, no. 7, pp. 567–583, 2014.

[137] H. P. Branigan, M. J. Pickering, J. Pearson, and J. F. McLean, "Linguistic alignment between people and computers," *J. Pragmatics*, vol. 42, no. 9, pp. 2355–2368, 2010.

[138] J. H. Pfeifer, M. Iacoboni, J. C. Mazziotta, and M. Dapretto, "Mirroring others' emotions relates to empathy and interpersonal competence in children," *NeuroImage*, vol. 39, no. 4, pp. 2076–2085, 2008.

[139] F. Nasoz and C. L. Lisetti, "MAUI avatars: Mirroring the user's sensed emotions via expressive multi-ethnic facial avatars," *J. Vis. Lang. Comput.*, vol. 17, no. 5, pp. 430–444, 2006.

[140] G. Caridakis, A. Raouzaiou, E. Bevacqua, M. Mancini, K. Karpouzis, L. Malatesta, and C. Pelachaud, "Virtual agent multimodal mimicry of humans," *Language Resources Eval.*, vol. 41, nos. 3/4, pp. 367–388, 2007.

[141] F. Bonin, N. Campbell, and C. Vogel, "Time for laughter," *Knowl.-Based Syst.*, pp. 15–24, 2014.

[142] J. C. Acosta and N. G. Ward, "Achieving rapport with turn-by-turn, user-responsive emotional coloring," *Speech Commun.*, vol. 53, no. 910, pp. 1137–1148, 2011.

[143] J. Gratch, S.-H. Kang, and N. Wang, "Using social agents to explore theories of rapport and emotional resonance," *Social Emotions Nat. Artif.*, vol. 181, p. 181, 2013.

[144] N. Novielli, F. de Rosis, and I. Mazzotta, "User attitude towards an embodied conversational agent: Effects of the interaction mode," *J. Pragmatics*, vol. 42, no. 9, pp. 2385–2397, 2010.

[145] R. Pekrun, A. Cusack, K. Murayama, A. J. Elliot, and K. Thomas, "The power of anticipated feedback: Effects on students' achievement goals and achievement emotions," *Learning Instruction*, vol. 29, pp. 115–124, 2014.

[146] C. Peters, C. Pelachaud, E. Bevacqua, M. Mancini, I. Poggi, and U. R. Tre, "Engagement capabilities for ECAS," in *Proc. Autonomous Agents Multi-Agent Syst. Workshop Creating Bonds ECAs*, 2005.

[147] I. Maks and P. Vossen, "A lexicon model for deep sentiment analysis and opinion mining applications," *Decision Support Syst.*, vol. 53, no. 4, pp. 680–688, 2012.

[148] A. Valitutti and C. Strapparava, "Interfacing WordNet-affect with OCC model of emotions," in *Proc. 3rd Int. Workshop -Corpora Res. Emotion Affect*, 2010, pp. 16–19.

[149] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in *Mining Text Data*, C. C. Aggarwal and C. Zhai, Eds. New York, NY, USA: Springer, 2012, pp. 415–463.

[150] M. Ochs, J. Ollivier, B. Coic, T. Brien, and F. Majeric, "AFFIMO: Toward an open-source system to detect AFFinities and eMOtions in user's sentences," in *Proc. Workshop Affect, Compagnon Artificiel, Interaction*, Grenoble, France, 2012, pp. 183–190.

[151] D. R. Heise, *Expressive Order Confirming Sentiments in Social Actions*. Berlin, Germany: Springer, 2007.

[152] M. Dohen, J.-L. Schwartz, and G. Bailly, "Speech and face-to-face communication: An introduction," *Speech Commun.*, vol. 52, no. 6, pp. 477–480, Jun. 2010.

[153] C. Clavel, I. Vasilescu, L. Devillers, G. Richard, and T. Ehrette, "Fear-type emotions recognition for future audio-based surveillance systems," *Speech Commun.*, vol. 50, pp. 487–503, 2008.

[154] J. Nicolle, V. Rapp, K. Bailly, L. Prevost, and M. Chetouani, "Robust continuous prediction of human emotions using multiscale dynamic cues," in *Proc. 14th ACM Int. Conf. Multimodal Interaction*, 2012, pp. 501–508.

[155] A. Osherenko, E. Andre, and T. Vogt, "Affect sensing in speech: Studying fusion of linguistic and acoustic features," *Proc. 3rd Int. Conf. Affective Comput. Intell. Interaction Workshops*, 2009, pp. 1–6.

[156] B. Schuller, G. Rigoll, and M. Lang, "Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture," in *Proc. Int. Conf. Acoust., Speech Signal Process.*, 2004, vol. 1, pp. 577–580.

[157] F. Mairesse, J. Polifroni, and G. Di Fabbrizio, "Can prosody inform sentiment analysis? experiments on short spoken reviews," in *Proc. Int. Conf. Acoust., Speech Signal Process.*, 2012, pp. 5093–5096.

[158] K. P. Truong and D. A. van Leeuwen, "Automatic discrimination between laughter and speech," *Speech Commun.*, vol. 49, no. 2, pp. 144–158, 2007.

[159] J. D. Moore, L. Tian, and C. Lai, "Word-level emotion recognition using high-level features," in *Computational Linguistics and Intelligent Text Processing*. New York, NY, USA: Springer, 2014, pp. 17–31.

[160] J. A. Russell, J.-A. Bachorowski, and J.-M. Fernandez-Dols, "Facial and vocal expressions of emotion," *Annu. Rev. Psychol.*, vol. 54, pp. 329–349, 2003.

[161] S. Ezzat, N. El Gayar, and M. M. Ghanem, "Sentiment analysis of call centre audio conversations using text classification," *Int. J. Comput. Inf. Syst. Ind. Manage. Appl.*, vol. 4, pp. 619–627, 2012.

[162] J. H. Jeon, R. Xia, and Y. Liu, "Level of interest sensing in spoken dialog using decision-level fusion of acoustic and lexical evidence," *Comput. Speech Language*, vol. 28, no. 2, pp. 420–433, 2014.

[163] H. Branigan, R. J. Lickley, and D. McKelvie, "Non-linguistic influences on rates of disfluency in spontaneous speech," in *Proc. 14th Int. Congr. Phonetic Sci.*, Aug. 1999, pp. 387–390.

[164] R. Amalberti, N. Carbonell, and P. Falzon, "User representations of computer systems in human-computer speech interaction," *Int. J. Man-Mach. Stud.*, vol. 38, no. 4, pp. 547–566, 1993.

[165] M. Adda-Decker, B. Habert, C. Barras, G. Adda, P. Boulade Mareuil, and P. Paroubek, "A disfluency study for cleaning spontaneous speech automatic transcripts and improving speech language models," in *Proc. ISCA Tutorial Res. Workshop Disfluency Spontaneous Speech*, Sep. 2003, pp. 67–70.

[166] I. Vasilescu, S. Rosset, and M. Adda-Decker, "On the functions of the vocalic hesitation euh in interactive man-machine question answering dialogs in French," in *Proc. 2nd Int. Symp. Linguistic Patterns Spontaneous Speech Joint Workshop*, 2010, pp. 111–114.

[167] H. H. Clark and J. E. Fox Tree, "Using uh and um in spontaneous speaking," *Cognition*, vol. 84, pp. 73–111, 2002.

[168] J. S. Strassel, *Simple Metadata Annotation Specification*, Linguistic Data Consortium. [Online]. Available: http://www.ldc.upenn.edu/Projects/MDE/Guidelines/SimpleMDE V6.2.pd, 2004.

[169] K. Georgila, "Using integer linear programming for detecting speech disfluencies," in *Proc. Annu. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Companion*, 2009, pp. 109–112.

[170] C. Dutrey, C. Clavel, S. Rosset, I. Vasilescu, and M. Adda-Decker, "A CRF-based approach to automatic disfluency detection in a French call-centre corpus," in *Proc. 15th Annu. Conf. Int. Speech Commun. Assoc.*, 2014, pp. 2897–2901.

[171] M. Maat, *Response Selection and Turn-Taking for a Sensitive Artificial Listening Agent*. Enschede, Netherlands: Univ. Twente, 2011.

[172] B. Oreström, *Turn-Taking in English Conversation*, vol. 66. Melbourne, FL, USA: Krieger, 1983.

[173] S. Lebai Lutfi, F. Fernández-Martínez, J. M. Lucas-Cuesta, L. López-Lebón, and J. M. Montero, "A satisfaction-based model for affect recognition from conversational features in spoken dialog systems," *Speech Commun.*, vol. 55, no. 7, pp. 825–840, 2013.

[174] M. Cristani, A. Pesarin, C. Drioli, A. Tavano, A. Perina, and V. Murino, "Generative modeling and classification of dialogs by a low-level turn-taking feature," *Pattern Recognit.*, vol. 44, no. 8, pp. 1785–1800, 2011.

[175] N. Campbell and S. Scherer, "Comparing measures of synchrony and alignment in dialogue speech timing with respect to turn-taking activity," in *Proc. 11th Annu. Conf. Int. Speech Commun. Assoc.*, Sep. 2010, pp. 2546–2549.

[176] M. W ollmer, B. Schuller, and G. Rigoll, "Keyword spotting exploiting long short-term memory," *Speech Commun.*, vol. 55, no. 2, pp. 252–265, 2013.

[177] P. Nakov, S. Rosenthal, Z. Kozareva, V. Stoyanov, A. Ritter, and T. Wilson, "Semeval-2013 task 2: Sentiment analysis in Twitter," in *Proc. 2nd Joint Conf. Lexical Comput. Semantics, Volume 2: Proc. 7th Int. Workshop Semantic Eval.*, Jun. 2013, pp. 312–320.

[178] M. Thomas, B. Pang, and L. Lee, "Get out the vote: Determining support or opposition from congressional floor-debate transcripts," in *Proc. Conf. Empirical Methods Natural Language Process.*, 2006, pp. 327–335.

[179] X. Ding, B. Liu, and P. S. Yu, "A holistic lexicon-based approach to opinion mining," in *Proc. Int. Conf. Web Search Data Mining*, 2008, pp. 231–240.

[180] R. Lau, C. Lai, and Y. Li, "Leveraging the web context for context-sensitive opinion mining," in *Proc. 2nd IEEE Int. Conf. Comput. Sci. Inf. Technol.*, 2009, pp. 467–471.

[181] L. Polanyi and A. Zaenen, "Contextual valence shifters," in *Computing Attitude and Affect in Text: Theory and Applications*, series. The Information Retrieval Series, J. G. Shanahan, Y. Qu, and J. Wiebe, Eds. New York, NY, USA: Springer, 2006, no. 20, pp. 1–10.

[182] R. López-Cózar and Z. Callejas, "Multimodal dialogue for ambient intelligence and smart environments," in *Handbook of Ambient Intelligence and Smart Environments*, H. Nakashima, H. Aghajan, and J. C. Augusto, Eds. New York, NY, USA: Springer, 2010, pp. 559–579.

[183] H. Liang, "Coevolution of political discussion and common ground in web discussion forum," *Soc. Sci. Comput. Rev.*, vol. 32, no. 2, pp. 155–169, 2014.

[184] H. Chen and D. Zimbra, "AI and opinion mining," *IEEE Intell. Syst.*, vol. 25, no. 3, pp. 74–80, May/Jun. 2010.

[185] K. Alam, M. Rahman, A. Saddik, and W. Gueaieb, "Adding emotional tag to augment context-awareness in social network services," in *Proc. IEEE Instrum. Meas. Technol. Conf.*, 2011, pp. 1–6.

[186] B. Straalen, D. Heylen, M. Theune, and A. Nijholt, "Enhancing embodied conversational agents with social and emotional capabilities," in *Agents for Games and Simulations*, vol. 5920, D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, F. Dignum, J. Bradshaw, B. Silverman, and W. Doesburg, Eds. Berlin, Germany: Springer, 2009, pp. 95–106.

[187] S.-H. Kang and J. Gratch, "Exploring users' social responses to computer counseling interviewers' behavior," *Comput. Human Behavior*, vol. 34, pp. 120–130, 2014.

[188] L. Menkhoff and R. R. Rebitzky, "Investor sentiment in the US-dollar: Longer-term, non-linear orientation on PPP," *J. Empirical Finance*, vol. 15, no. 3, pp. 455–467, 2008.

[189] R. Niewiadomski, S. J. Hyniewska, and C. Pelachaud, "Constraint-based model for synthesis of multimodal sequential expressions of emotions," *IEEE Trans. Affective Comput.*, vol. 2, no. 3, pp. 134–146, Oct.-Dec. 2011.

[190] S. Kopp, B. Krenn, S. Marsella, A. N. Marshall, C. Pelachaud, H. Pirker, K. R. Thorisson, and H. Vilhjalmsson, "Towards a common framework for multimodal generation: The behavior markup language," in *Proc. 6th Int. Conf. Intell. Virtual Agents*, 2006, pp. 205–217.

[191] M. Schroder, E. Bevacqua, R. Cowie, F. Eyben, H. Gunes, D. Heylen, M. Ter Maat, G. McKeown, S. Pammi, M. Pantic, C. Pelachaud, B. Schuller, E. de Sevin, M. Valstar, and M. Wollmer, "Building autonomous sensitive artificial listeners," *IEEE Trans. Affective Comput.*, vol. 3, no. 2, pp. 165–183, Apr.–Jun. 2012.

[192] D. Heylen, E. Bevacqua, M. ter Maat, C. Pelachaud, and E. de Sevin, "Updated demonstrator of the dialogue manager. Project Deliverable D4a, SEMAINE," 2009.

[193] C. Langlet and C. Clavel, "Modelling user's attitudinal reactions to the agent utterances: Focus on the verbal content," in *Proc. 5th Int. Workshop Corpora Res. Emotion, Sentiment Social Signals*, Reykjavik, Iceland, 2014, pp. 90–95.

[194] G. McKeown, M. Valstar, R. Cowie, M. Pantic, and M. Schroder, "The semaine database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent," *IEEE Trans. Affective Comput.*, vol. 3, no. 1, pp. 5–17, Jan. 2012.

[195] Z. Callejas, B. Ravenet, M. Ochs, and C. Pelachaud, "A computational model of social attitudes for a virtual recruiter," in *Proc. Int. Conf. Autonomous Agents Multi-Agent Syst.*, 2014, pp. 93–100.

[196] I. Poggi, *Mind, Hands, Face and Body: A Goal and Belief View of Multimodal Communication*. London, U.K.: Weidler, 2007.

[197] S. Campano, J. Durand, and C. Clavel, "Comparative analysis of verbal alignment in human-human and human-agent interactions," in *Proc. 9th Int. Conf. Language Resources Eval.*, Reykjavik, Iceland, 2014, pp. 4415–4422.

[198] S. Campano, C. Clavel, and C. Pelachaud, "I like this painting too": When an eca shares appreciations to engage users," in *Proc. 14th Int. Joint Conf. Autonomous Agents Multiagent Syst.*, 2015, pp. 1649–1650.

[199] C. Langlet and C. Clavel, "Improving social relationships in face-to-face human-agent interactions: When the agent wants to know users likes and dislikes," in *Proc. Annu. Meeting Assoc. Comput. Linguistic*, 2015, to appear.

[200] K. Anderson, E. André, T. Baur, S. Bernardini, M. Chollet, E. Chryssafidou, I. Damian, C. Ennis, A. Egges, P. Gebhard, H. Jones, M. Ochs, C. Pelachaud, K. Porayska-Pomsta, P. Rizzo, and N. Sabouret, "The TARDIS framework: Intelligent virtual agents for social coaching in job interviews," in *Proc. 10th Int. Conf. Adv. Comput. Entertainment*, 2013, pp. 476–491.

[201] Z. Callejas, B. Ravenet, M. Ochs, and C. Pelachaud, "A model to generate adaptive multimodal job interviews with a virtual recruiter," in *Proc. 9th Int. Conf. Language Resources Eval.*, 2014, pp. 3615–3619.

[202] A. I. Huffcutt, C. H. Van Iddekinge, and P. L. Roth, "Understanding applicant behavior in employment interviews: A theoretical model of interviewee performance," *Role Personality Human Resource Manage.*, vol. 21, no. 4, pp. 353–367, 2011.

[203] T. Macan, "The employment interview: A review of current studies and directions for future research," *Employee Selection Beginning 21st Century*, vol. 19, no. 3, pp. 203–218, 2009.

[204] T. Baur, I. Damian, P. Gebhard, K. Porayska-Pomsta, and E. André, "A job interview simulation: Social cue-based interaction with a virtual character," in *Proc. Int. Conf. Social Comput.*, 2013, pp. 220–227.

[205] S. B. Brundage, K. Graap, K. F. Gibbons, M. Ferrer, and J. Brooks, "Frequency of stuttering during challenging and supportive virtual reality job interviews," *J. Fluency Disorders*, vol. 31, no. 4, pp. 325–339, 2006.

Chloé Clavel is an associate professor in affective computing in the Greta Team belonging to the multimedia (MM) group of the Signal and Image Processing Department of Telecom-ParisTech. Her research focuses on two issues: acoustic analysis of emotional speech and opinion mining through natural language processing. After her PhD degree, she worked in the laboratories of two big French companies that are Thales Research and Technology and EDF R&D where she developed her research around audio and text mining applications. At Telecom-ParisTech, she is currently working on interactions between humans and virtual agents, from users socio-emotional behavior analysis to socio-affective interaction strategies.

Zoraida Callejas is an assistant professor at the University of Granada, where she was awarded a PhD degree in 2008, and is a member of the Research Centre for Information and Communications Technologies (CITIC-UGR). She has been a visiting professor in the Technical University of Liberec, Czech Republic (2007-14), University of Trento, Italy (2008), University of Ulster, Northern Ireland (2009,2013), Technical University of Berlin, Germany (2010), University of Ulm, Germany (2012,2014), and Telecom ParisTech, France (2013). She focuses her research on speech technology and in particular, on affective spoken and multimodal dialog systems. She has made presentations at the main conferences in the area and has published her research in several international journals and books.