# Computing emotions on discourse

**Mauricio Iza**[1]

**Abstract.** Findings on the role that emotion plays on discourse have transformed Artificial Intelligence computations. Modern research explores how to simulate more intelligent and flexible systems. Several studies focus on the role that emotion has in order to establish values for alternative decision and decision outcomes. However, emotional concepts in these theories are generally not defined formally and it is difficult to describe in systematic detail how processes work. In this sense, structures and processes cannot be explicitly implemented. This work tries to unify emotion as interaction in order to explain the dynamic role it plays in action and cognition.[12]

## 1 INTRODUCTION

Our perception of the world, our thoughts, and our social interactions are crucially dependent on our bodies and our embodied interaction with the world. This is the idea of considering cognition as situated. Cognition does not only take place "in the head", based on internal models of the world, but crucially depends on active situated interaction with the world [12,14,33,6].

Within this framework, a main hypothesis is that emotional mechanisms play a critical role in structuring the high-level thought processes of cognitive systems. Some models of these mechanisms can be usefully integrated in artificial cognitive systems architectures, which constitute a significant step towards cognitive systems that reason and behave, externally and internally, in accordance with emotional requirements [19,35,31,32].

Most current human-computer interaction systems do not account for the fact that human-computer communication is always socially situated and that we use emotion to enhance our communication. Some systems can sense the affective states of the human (e.g., stress, inattention, etc.) and are able of adapting and responding to these affective states. Likely, they are perceived as more natural and efficacious. This suggests that it may be beneficial for computers to recognize the user's emotions and other related cognitive states and expressions [15, 16,29,30].

For instance, interpreting the mix of audio-visual signals is essential in human communication. Researchers have to take into account the advances in the development of unimodal techniques (e.g., speech and audio processing, computer vision, etc.). In traditional human-computer interaction, the user faces a computer and interacts with it via a mouse or a keyboard. In the new applications (e.g., multiple agents, intelligent homes) interactions are not explicit commands. Some of the methods include gesture, speech [27], eye movements [10], etc.

The work of [26] suggested several applications where it is beneficial for computers to recognize human emotions. For instance, knowing the user's emotions, the computer can become a more effective tutor. Synthetic speech with emotions in the voice should sound more pleasing than a monotonous voice. Computer agents could learn the user's preferences through the user's emotions. Another application is to help the human users monitor their stress level. In clinical settings, recognizing a person's inability to express certain facial expressions may help diagnose early psychological disorders.

In relation to the problem of affective communication, [2] identified three points in order to develop systems that takes into account affective information: (i) embodiment: experiencing physical reality; (ii) dynamics: mapping the experience and the emotional state onto a temporal process and a particular label; and (iii) adaptive interaction: conveying emotive response, responding to a recognized emotional state.

As we will see, some tutoring systems have explored this potential to inform user models. Likewise, dialogue systems, mixed-initiative planning systems, or systems that learn from observation could also benefit from such an approach [17]. That is, considering emotion as interaction can be relevant in order to explain the dynamic role it plays in action and cognition (see [11]).

In this work, we will provide some neuroscientific and psychological insights into the sensorimotor grounding of conceptualization and language use. For instance, the role of canonical and mirror neurons as underlying the use of nouns and verbs, in order to develop novel approaches to grounding of robotic conceptualization and language use (more precisely, verbal labelling of objects and actions), based on the insights gained under richer computational and robotic models [14,22,26].

## 2 COMPUTING FRAMEWORK

In order to characterize a computational model of emotion, we have to take into account different interdisciplinary uses to which computational models can be put, such as improving human-computer interaction or enhancing general models of intelligence.

Starting from some integrated computational models have tried to incorporate a variety of cognitive functions (e.g., [1]), more recent cognitive systems in AI focus on the role of emotion in order to address control choices by driving cognitive resources on problems of adaptive significance for the agent [3]. For example, human computer interaction attempts to recognize user's emotion including physiological indicators and facial and vocal expressions. Similarly, how we can use of emotion or emotional displays in avatars that interact with the user, for instance, to increase student motivation in a tutoring system.

In this sense, computational models take different frameworks in research and applications. On one hand, psychological models emphasize on fidelity with respect to human emotion processes.

1 Department of Psychology, University of Malaga, Spain. Email: iza@uma.es.
2

On the other hand, AI models evaluate how the modeling of emotion impacts reasoning processes or improves the fitness between agent and its environment. That is, the model improves and makes more effective the human-computer interaction.

Several models have been proposed and developed. However, some fundamental differences arise from their underlying emotional constructs. For instance, as we will see below, some discussions on if emotion precedes or follows cognition disappears if one adopts a dynamic system perspective. Here, we will discuss two main approaches.

Researchers in AI use mainly two different methods to analyze emotions. One method tries to classify emotions into discrete categories (joy, fear, love, surprise, etc.) using different modalities as inputs. The problem is that the stimulus may contain blended emotions and the ontological classification of these categories may be too much restrictive. Another method tries to delineate multiple dimensions or scales to describe emotions. Two normally used scales are valence and arousal. Valence describes pleasantness of the stimulus, with positive or pleasant on one hand (happiness) and negative on the other (disgust). By contrast, arousal or activation goes from low level (sadness) to high level (surprise). The different emotional labels are plotted at various positions on a 2D plane spanned by these two axes to construct a 2D emotion model.

In this sense, expressions such as facial recognition or vocal emotions are very important in this context. The case of facial recognition is mainly based on Ekman's [9] Facial Action Coding System. Here we can distinguish two kinds of classification schemes. Firstly, static classifiers (Bayesian networks) classify each frame in a video to one of the facial expression categories based on the results of a particular video frame. Secondly, dynamic classifiers use several video frames and classify by analyzing the temporal patterns of the region analyzed or features extracted. Dynamic classifiers are very sensitive to appearance changes in the facial expressions of different individuals, that is, they are more suited for person-dependent experiments [5]. However, static classifiers are easier to train and need less training data but when used on a continuous video sequence they can be unreliable.

Something similar happens in the case of vocal aspect of a communicative message. If we only consider the textual content, we probably miss relevant aspects of the utterance and even misunderstand the real meaning of the message (pragmatic inferences). In contrast to much work research dedicated to spoken language processing, the processing of emotional speech has been almost abandoned.

Quantitative studies of vocal emotions use prosodic information (intonation, rhythm, lexical stress and other features in speech). This is extracted by using measures like pitch, duration and intensity of a given utterance. But, [23] pointed out some limitations of vocal-affect analyzers: (i) perform singular classification of input audio signals into a few emotion categories (e.g., irony, happiness); (ii) do not perform a context sensitive analysis (context, agent) of the input audio signal; (iii) do not analyze extracted vocal expressions information on different time scales: proposed inter-audio-frame analyses are used for the detection of supra-segmental features (pitch and intensity over the duration of a syllable or word) and for the detection of phonetic features; (iv) inferences about moods and attitudes (longer time scales) are difficult to make based on the current vocal-affect analyzers; (v) strong assumptions are adopted (e.g., short sentences, delimited by pauses, etc.) and test data sets used are small containing exaggerated vocal expressions of affective states.

In resume, studies on the psychology on the accuracy of predictions from observations of predictive behavior suggest that the combined face and body approaches are the most informative. But, research in facial expression recognition and vocal affect recognition has been independently performed. The developed systems have most of the drawbacks of unimodal affect analyzers. Most works in facial expression recognition use still photographs or video sequences without speech. Likewise, works on vocal emotion detection normally use only audio information. So, we should consider how we can integrate different agents, contexts and processing sources of information for a natural interaction.

## 3 EMOTIONAL APPROACHES

Models in language processing have researched how words are interpreted by humans. Many models presume the ability to correctly interpret the beliefs, motives and intentions underlying words. The interest relies also on how emotion motivates certain words or actions, inferences, and communicates information about mental state (e.g., [13]).

Here, emotion has been often considered as another kind of information. That is, discrete units or internal states that can be transmitted from subjects to robotic systems and back. There exists broad experimental evidence on grounded language comprehension, such as action related speech activates mirror system or action Sentence Compatibility Effect, where verbal description of spatially directional actions facilitates movements in the same direction.

As these experimental data show, activating accessible constructs or attitudes through one set of stimuli can facilitate cognitive processing of other stimuli under certain circumstances, and can interfere with it under other circumstances. Some of the results support and converge on those centered on the constructs of current concern and emotional arousal.

In the informational approach, one subject has an emotion internally. Emotions are characterized as discrete units within the system or as points in a multidimensional space. It has to be transmitted through a process of encoding, transmission and decoding. Here, emotion is considered to be discrete, well-defined and transferable. This requires designing systems to model and transmit emotion.

In the interactional approach, communication of emotion is not only transmission. It is an active process of co-constructing one's emotional state; it does not require decoding process, only active interpretation. Here, emotion is considered to be complex, ambiguous, adaptable and non-formalizable. This requires designing systems that support social settings agents in producing, negotiating, experiencing and interpreting emotional inferences.

Each approach requires different designing systems. Whereas the informational approach seeks to bind truth into discrete and often sequential units, the interactional approach allows for meaning to be enacted within the social setting. The issue of the dynamic, situated interpretation and attribution of emotional state is very important [25]. This also has been recently considered within the idea of technology as sixth sense (e.g., [21, 28]. Emotion is not formalized into the system; the emotional meaning is supplied by the agents. Emotional meaning emerges as in a situated way over the course of the interaction.

For purpose of modeling emotion generation, research has mainly centered on appraisal theories, which are the dominant basis for that type of computational model. Appraisal theories generally argue that people are constantly evaluating their environment, and that evaluations result in emotions such as fear or anger. Each theory differs in its appraisal variables and the way in which appraisals are generated (simultaneously vs. specific order).

Since all cognitive processes work with the associative network and emotional data is embedded within all the nodes, any process can use emotion data to model emotional effect. EmoCog [18] is designed to be flexible, so that further dimensions can be incorporated into both the associative network and mechanisms (arousal and valence). It tries to integrate emotions in a cognitive architecture with an associative network memory, cognitive attention and appraisal following cognition. The associative network allows for concepts to influence each other emotionally, also hold emotional information for cognitive processes and emotion generation. The cognitive attention subcomponent allows for controlled elaboration and emotional rise and decay. The last subcomponent focus on how appraisal and association management follow cognition in the associative network, how cognition influences emotional generation.

As we have previously seen, many design decisions on affective interaction depends on the underlying assumptions and techniques used in the interface. Some of them involve serious limitations and result in a too much rigid system. A possible solution relies on learning [17]. Instead of using a priori rules to interpret words, we should potentially learn rules depending on different applications, agents and contexts. There are algorithms to adapt the models and we can use prior knowledge when learning new models. For instance, a prior model of emotional word recognition trained based on a certain agent can be used as a starting point for learning models for other agents; or for the same agent in different contexts.

# 4 DISCUSSION

Models in language processing have researched how words are interpreted by humans. Many models presume the ability to correctly interpret the beliefs, motives and intentions underlying words. The interest relies also on how emotion motivates certain words or actions, inferences, and communicates information about mental state. For instance, some tutoring systems have explored this potential to inform user models. Likewise, dialogue systems, mixed-initiative planning systems, or systems that learn from observation could also benefit from such an approach.

As these experimental data show, activating accessible constructs or attitudes through one set of stimuli can facilitate cognitive processing of other stimuli under certain circumstances, and can interfere with it under other circumstances. Some of the results support and converge on those centered on the constructs of current concern and emotional arousal.

Future research has to take seriously into account this question: how to develop models where emotion interacts with cognitive processing. One example could be the work of [26] where it is combined speech-based emotion recognition with adaptive human-computer modeling. With the robust recognition of emotions from speech signals as their goal, the authors analyze the effectiveness of using a plain emotion recognizer, a speech-emotion recognizer combining speech and emotion recognition, and multiple speech-emotion recognizers at the same time. The semi-stochastic dialogue model employed relates user emotion management to the corresponding dialogue interaction history and allows the device to adapt itself to the context, including altering the stylistic realization of its speech.

This goal involves system integration architectures such as Open Agent Architecture [20] and Adaptive Agent Architecture [7]. These multi-agent architectures provide essential infrastructure for integrating many complex modules needed to implement emotional word processing and perform it in a distributed way. In a multiagent architecture, the components needed (e.g., speech recognition, word processing, emotional integration) can be written in different programming languages, on different machines/agents, and with different operating systems. Agent communication languages have been developed to take account of asynchronous delivery, triggered responses, multi-casting and other concepts from distributed systems.

In this kind of architecture, speech and gestures can arrive in parallel or asynchronously via individual modality agents, where results are transferred to a facilitator. Next, these results, normally an n-best list of conjectured lexical items and related time-stamp information, are routed to appropriate agents for further language processing. Then, sets of meaning fragments derived from the speech, or other modality, arrive at the multimodal integrator which decides whether and how long (based on system's temporal thresholds) to wait for recognition results from other modalities. It fuses the meaning fragments into a semantically and temporally compatible whole interpretation before passing the results back to the facilitator. Here, the system's final multimodal interpretation is confirmed by the interface, delivered as multimedia feedback to the agent, and executed by the relevant application.

Many studies in the literature analyse each subsystem (e.g., speech recognizers) but few applications take into account these architectures. A clear example is a comprehensive analysis of human-computer interaction [8]. In the informational approach, emotions are discrete and we evaluate how they are contained and transmitted. In the interactional approach, we have to focus on evaluating things such as awareness, different kind of expressions and engagement between agents in a social setting or action.

# 5 FUTURE SOLUTIONS

Future solutions are mainly oriented to robot applications (e.g., [4, 34]). On the one hand, how robots can be used in cognitive rehabilitation as compensatory technology (e.g., autism). On the other hand, how robots can evolve or adapt to new situations or contexts. For instance, [15] presented a cognitive architecture that tries to measure the effects of emotional cognition on learning. That is, it uses emotional intelligence to learn new concepts from previously unknown kind of experiences.

We can interpret the suggested selection mechanism as an information filter. This information filter only selects the measurement for the required features and passes them to the memory system. Features that do not contribute in solving a given task are discarded. This also requires a dynamical and flexible system architecture that allows for a demand-driven combination of processing modules. We have proposed such architecture for the congruent emotion of word processing. To acquire more complex information, the system needs to combine those procedures in a suitable way within memory representation. Beside this, the system has to decide which properties it has to measure for solving the current task. The resulting representation is demand related, as only the pieces of information to solve the task is acquired. This task driven representation can serve as a foundation for learning new relations between words and emotions and for interpreting current interactions.

[34] addressed the problem of the detection and revealing of the relevant "context" to inform affect detection. They implemented a context-based affect detection component embedded in an improvisational virtual platform. The software allows up to five human characters and one intelligent agent to be engaged in one session to conduct creative improvisation within loose scenarios. Some of these conversations reveal personal subjective opinions or feelings about situations, while others are caused by social interactions and show opinions and emotional responses to other participant characters. In order to detect affect from such contexts, first of all a naïve Bayes classifier is used to categorize these two types of conversations based on linguistic cues. A semantic-based analysis is also used to further derive the discussion themes and identify the target audiences for the social interaction inputs. Then, two statistical approaches have been developed to provide affect detection in the social and personal emotion contexts. The emotional history of each individual character is used in interpreting affect relating to the personal contexts, while the social context affect detection takes account of interpersonal relationships, sentence types, emotions implied by the potential target audiences in their most recent interactions and discussion themes. The new development of context-based affect detection is integrated with the intelligent agent.

In this context, a psychological framework of emotional language processing is needed to describe the steps humans take when they interact with other computer systems or agents [24]. This framework can be used to help evaluate the efficiency and naturalness of a user interface (e.g., design principles, emotional inferences, etc.).

So, the key question is to represent, reason, and exploit various models of word processing to more effectively process input, generate output, and manage the dialog and interaction between different agents. The input data (words) should be, cognitive and emotionally, processed in a joint feature space according to a context-dependent model.

# REFERENCES


[1] Anderson, J.R. *Rules of the mind*. Hillsdale, NJ: LEA (1993).

[2] Bicho, E., Louro, L. & Erlhagen, W. Integrating verbal and nonverbal communication in a dynamic neural field architecture for human-robot interaction. *Frontiers in Neurorobotics 4*, article 5 (2010).

[3] Blanchard, A. & Cañamero, L. Developing affect-modulated behaviors: stability, exploration, exploitation or imitation? *6th International Workshop on Epigenetic Robotics*. Paris (2006).

[4] Callejas, Z., Griol, D., López-Cózar, R., Espejo, G. & Abalos, N. Merging intention and emotion to develop adaptive dialogue systems. In C. Delgado & T. Kobayashi (eds.), *Proceedings of the paralinguistic information and its integration in spoken dialogue system workshop,* pp. 165-174. Springer (2011).

[5] Canli, T., Sivers, H., Thomason, M.E., Whitfield-Gabrieli, S., John, D.E. & Gotlib, I.H. Brain activation to emotional words in depressed vs. healthy subjects. *Neuro Report 15* (17): 2585-2588 (2004).

[6] Clark, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences 36*, 181-253 (2013).

[7] Cohen, I., Sebe, N., Cozman, F., Cirelo, M. & Huang, T.S. Semisupervised learning of classifiers: theory, algorithms and their applications to human-computer interaction. *IEEE Trnasactions on PAMI 22*(12), 1553-1567 (2004).

[8] Dix, A., Finlay, J., Abowd, G. & Beale, R. *Human-computer interaction*. Englewood Cliffs, NJ: Prentice Hall (2003).

[9] Ekman, P. Facial expressions. In C. Blakemore & S. Jennett (Eds.), *Oxford Companion to the Body*. London: Oxford University Press (2001).

[10] Grauman, K., Betke, M., Lombardi, J., Gips, J. & Bradski, G. Communication via eye blinks and eyebrow raises: video-based human-computer interfaces. *Universal Access in the Information Society 2*(4), 359-373 (2003).

[11] Gratch J, Marsella S & Petta P. Modeling the cognitive antecedents and con-sequences of emotion. *Journal of Cognitive Systems Research 10*(1), 1-5 (2009).

[12] Hamker, F.H. Modeling feature-based attention as an active top-down inference process. *BioSystems 86*, 91-99 (2006).

[13] Horchak, O.V., Giger, J.C., Cabral, M. & Pochwatko, G. From demonstration to theory in embodied language comprehension: A review. *Cognitive Systems Research 29-30,* 66-85 (2014).

[14] Kisley M.A., Campbell, A.M., Larson, J.M., Naftz, A.E., Regnier, J.T. & Davalos, D.B. The impact of verbal framing on brain activity evoked by emotional images. *Journal of Integrative Neuroscience 10*(4), 513-524 (2011).

[15] Kushiro, K., Harada, Y., & Takeno, J. Robot uses emotions to detect and learn the unknown. *Biologically Inspired Cognitive Architectures, 4*, 69-78 (2013).

[16] Laurue, O., Poirier, P. & Nkambou, R. The emergence of (artificial) emotions from cognitive and neurological processes. *Biologically Inspired Cognitive Architectures 4*, 54-68 (2013).

[17] LeCun, Y, Benjio, Y. & Hinton, G. Deep learning. *Nature 521*, 436-444 (2015).

[18] Lin, J., Spraragen, M., Blythe, J. & Zyda, M. EmoCog: Computational Integration of Emotion and Cognitive Architecture. In *Proceedings of the Twenty-Fourth FLAIRS Conference* (2011).

[19] Lupyan, G. & Clark, A. Words and the world: Predictive coding and the language-perception-cognition interface. *Current Directions in Psychological science 24*(4), 279-284 (2015).

[20] Martin, D., Cheyer, A. & Moran, D. The open agent architecture: a framework for building distributed software systems. *Applied Artificial Intelligence 13*, 91-128 (1999).

[21] McCarthy, J. & Wright, P. *Technology as experience*. Cambridge: MIT Press (2004).

[22] Ochsner, K.N. & Gross, J.J. The cognitive control of emotion. *Trends Cogn. Sci. 9*, 242-249 (2005).

[23] Pantic, M. & Rothkrantz, L.M. Toward and affect sensitive multimodal human-computer interaction. *Proceedings of the IEEE 91* (9), 1370-1390 (2003).

[24] Parkinson, B. What holds emotions together? Meaning and response coordination. *Cognitive System Research 10*, 31-47 (2009).

[25] Picard, R.W. *Affective computing.* MIT Press (1997).

[26] Pittermann, J., Pittermann, A. & Minker, W. *Handling emotions in human-computer dialogues*. Berlin: Springer Verlag (2010).

[27] Potamianos, G., Neti, C., Luettin, J. & Mattehews, I. Audio-visual automatic speech recognition: an overview. In E. Vatikiotis-Bateson & P. Perrier (Eds.), *Issues in visual and audio-visual speech processing.* MIT Press. (2004)

[28] Rebhan, S. & Eggert, J. Dynamic, task-related and demand-driven scene representation. *Cog Comput 3*, 124-145 (2011).

[29] Seth, A.K. Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences 17*(11), 565-573 (2013).

[30] Srinivasa, N. & Chelian, S.E. Executive control of cognitive agents using a biologically inspired model architecture of the prefrontal cortex. *Biologically Inspired Cognitive Architectures 2*, 13-24 (2012).

[31] Sun, R. & Mathews, C. Implicit cognition, emotion, and meta-cognitive control. *Mind Soc 11*, 107-119 (2012).

[32] Sun, R., Wilson, N. & Lynch, M. Emotion: A unified mechanistic interpretation from a cognitive architecture. *Cogn. Comput. 8*, 1-14 (2016).

[33] Watson, C.E., Armstrong, B.C. & Plaut, D.C. Connectionist modeling of neuropsychological deficits in semantics, language, and reading. In M. Faust (ed.), *The handbook of the neuropsychology of language*, pp. 103-124. London: Blackwell (2012).

[34] Zhanj, L. & Barnden, J. Affect sensing using linguistic, semantic and cognitive cues in multi-threaded improvisational dialogue. *Cognitive Computation 4*(4), 436-459 (2012).

[35] Ziemke, T. & Lowe, R. On the role of emotion in embodied cognitive architectures: from organisms to robots. *Cogn. Comput. 1*, 104-117 (2009).