Accounting for irony and emotional oscillation in computer architectures

Artemy Kotov
Institute of Linguistics, Russian State University for the Humanities
Russia, Moscow, Miusskaya pl. 6
kotov@harpia.ru

Abstract

We demonstrate computer architecture, operating on semantic structures (sentence meanings or representations of events) and simulating several emotional phenomena: top-down emotional processing, hypocrisy, emotional oscillation, sarcasm and irony. The phenomena can be simulated through the interaction between emotional processing and operations with semantics. We rely on a multimodal corpus of oral exams to observe the usage of emotional expressive cues in situations of strong conflict between internal motivation and external social limitations. We apply the observations to make the computer model simulate the observed cases of combined emotional expressions.

1. Introduction

While psychology and linguistics observe numerous examples of emotional expression, studies in the field of cognitive psychology and computer simulation develop cognitive models and their computer realizations in order to simulate certain subsets of the observed phenomena. Most of the studies concentrate on recognition and simulation of surface emotional phenomena (gestures, movements, phonetics), while operations with text semantics are often considered to be a subject for future studies.

At the same time, interaction between text semantics and emotional processing appears in a row of emotional phenomena, very attractive from the point of view of computer simulation. These are: text influence (computer understanding and synthesis of advertising and propaganda), machine comprehension and synthesis of humor, construction of emotional monologues with apparent changes in the speakers emotions, and indirect emotional expression in speech – emotional hints, sarcasm and irony. These phenomena can be simulated by a computer architecture which includes both: operations with text semantics and a model of emotional processing – and manages the interaction between the two components. In this article we demonstrate our approach towards the construction of such model.

We are developing a computer architecture which operates with limited semantic representations: meanings of incoming phrases or structures of the observed events. The model consists of a number or rules (scripts), which are activated by the incoming events and from speech output of the model.

Internal dynamics and conflicts inside the model help to simulate human reactions with combined semantic processing and emotional dynamics. In this article we address some forms of sophisticated human emotional expressions, like hypocrisy, emotional oscillation, irony, sarcasm and laughter.

The model applies to animate the behavior of simple cartoon agents (Fig 1.). They receive on their input incoming phrases (parsed to simple semantic structures), external or internal events. The agent may be surrounded by many counterparts – he receives phrases or events from one of them, and may react to the event, as if passing through several emotional states and addressing different parties. In Fig. 1 the agent *B* is hit (or simply "touched") by a counteragent *C*. He simulates irritation (M_1): replies with emotional phrases and shows aggressive gestures. Then he passes to another state like "blaming himself" or "looking for a solution" (M_2) and finally "calms down", shows etiquette replies and justifies the opponent (M_3). He distinguishes "words" and "thoughts" and may address different parties: his counteragent *C* or some other presenting agent *A* – taking into account the social distance with each

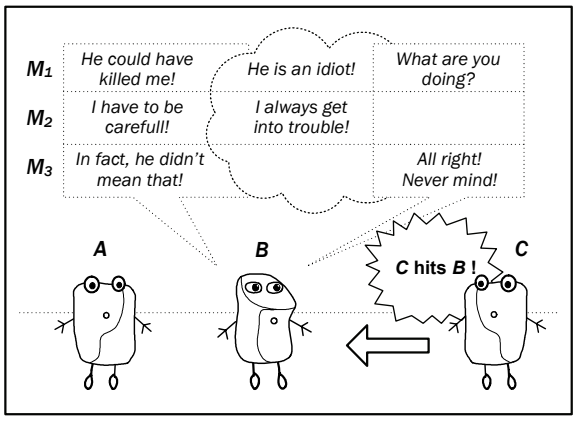

Figure 1: Computer agent B shows speech emotional reactions in a situation, where it was hit by a counteragent C. B passes through several short emotional states – microstates (M_1-M_3) and switches in communication, addressing himself (B), the opponent (C) or some other presenting party (A).

978-1-4244-4799-2/09/$25.00 ©2009 IEEE

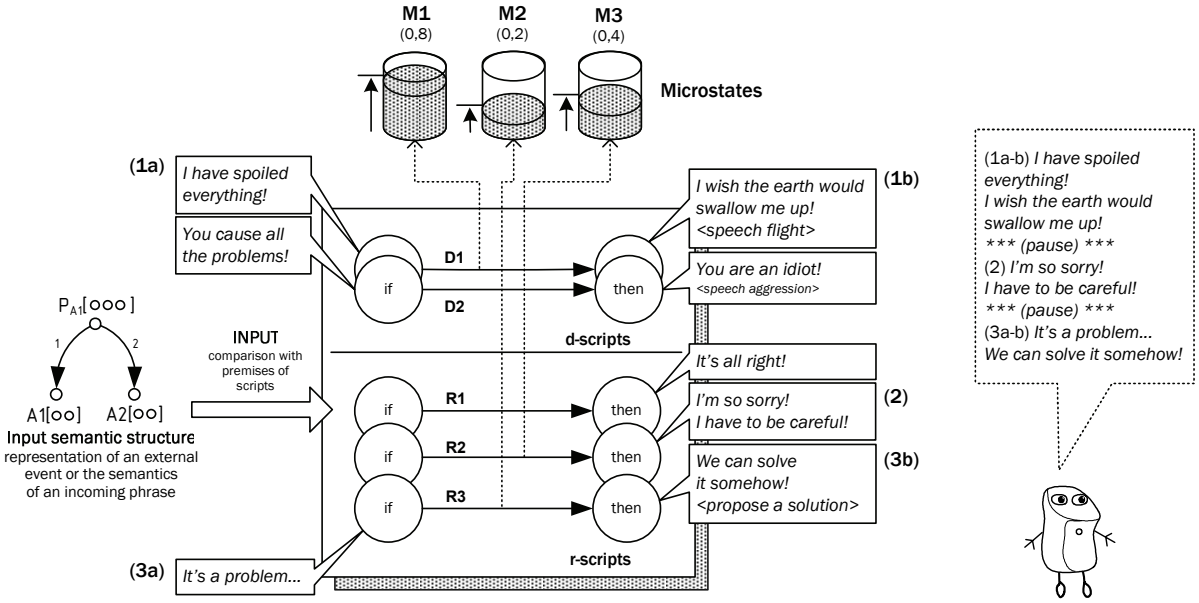

Figure 2: Input semantic structure is compared with a list of scripts, which forces their activation and subsequent speech reactions: *d-scripts* simulate emotional speech reactions and *r-scripts* are responsible for "rational" speech reactions: problem solving, etiquette, conciliation. In the case of a single-sentence answer the winning script forms the speech output of the model. Groups of scripts are linked to microstates (M1-M3), which also receive different level of activation during input processing (imaginary values of activation shown here). In case of an emotional monologue (emotional oscillation) microstates consecutively discharge (from highest to lowest – here M1-M3-M2), resulting in speech output from the connected scripts (here indicated as 1-2-3).

counteragent. This speech behavior is controlled by a computer model.

2. Control of emotional speech dynamics

The idea to account emotions in computer architectures in a form of competing *if-then* operators ("proto-specialists") was suggested by M. Minsky [1] and further elaborated in a number of computer architectures: for example CogAff [2] and the architecture of agent Max [3]. These architectures distinguish units for emotional processing (in our case – *d-scripts*) and units for rational processing (*r-scripts*). The units compete during processing of input, and the winner may depend on both (a) structure of the incoming event and (b) current emotional state of the agent.

The model consists of 63 d-scripts and 40 r-scripts, selected after the analysis of emotional texts in mass media/advertising and basing on the results of a psychological survey in [4]. Each script has *if* and *then* conditions, represented as semantic graphs (or masks of semantic graphs) and in the current version is linked with a number of expressive cues: gestures, ready utterances or utterance templates. At the same time, as the model operates on semantic structures, it can be extended to «full» synthesis of utterances from semantic representations in future versions.

Scripts are grouped into *microstates* – short emotional states with unified expressive patterns. If a microstate is active, its scripts are preferred during processing of input: if we 'touch' an irritated agent, it will emotionally consider this as an 'aggressive strike'. On the other side, during processing of input scripts, corresponding microstates are activated to different degrees. If several microstates are activated by the input event – they may further start to discharge one after another, forming the emotional monologue of the agent (Fig. 2).

Utterances are checked by a filtering component. Aggressive or very sincere utterances are suppressed for the output, and appear only as the "thoughts" of the agent in a special "cloud". The level of filtering depends (a) on the current microstate – as a very nervous agent may feel free to abuse, and (b) on the representation of the addressee – as the agent may be open to different degrees with different counterparts.

The dynamics of microstates and filtering allows us to simulate long emotional monologues for the agent, combined with possible "thoughts" and switches in communication, where the agent addresses different parties.

3. Operations with semantic markers

Representation of an on object differs from situation to situation [5], which is revealed in nomination. Depending on a situation we can call one and the same object as *human*, *voter*, *body*, *man*, *student* or *teacher*. In cognitive semantics the mechanism of this variation is defined as changing of *frames* [1] or *profiling* [6].

Similar procedures with nominative semantics appear during emotional processing. Some speech formulae may activate human emotional processing, while others – do not. This is described as a bottom-up emotional process [7]. As the emotional response is sensitive to

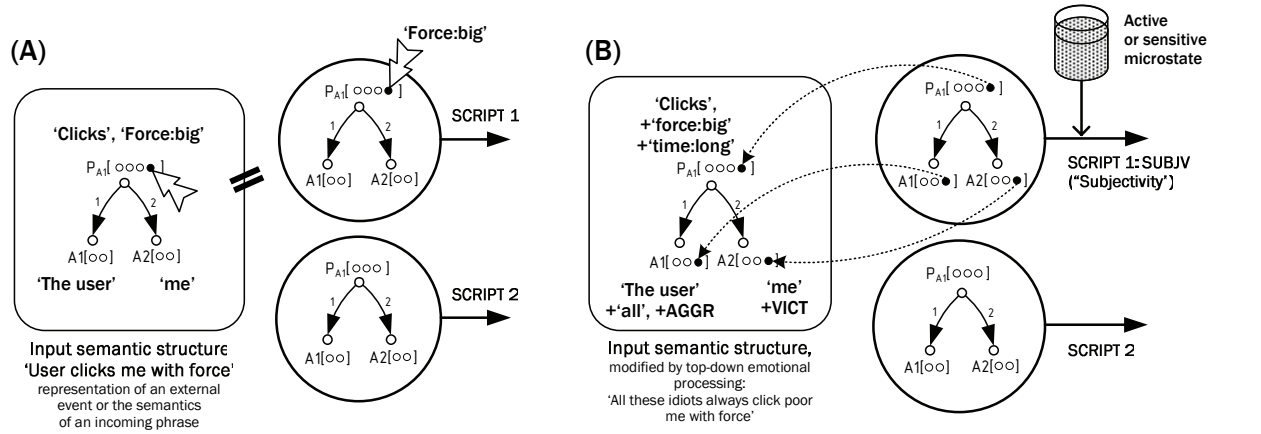

Figure 3: Input semantic structures may contain markers, relevant for a specific script (*key markers*) and stimulating emotional or rational processing: if someone does something to me 'with force' – it is likely to provoke a negative response; this simulates bottom-up emotional processing (A). At the same time, scripts with active microstates receive higher preferences during comparison – they are preferred for neutral incoming events and enrich the event with their key markers, changing the representation of the situation and simulating top-down emotional processing (B).

some specific semantic markers, we can manipulate the nomination in text in order to facilitate the emotional processing of the addressee (when calling some disagreement – *an aggression*) or to reduce and inhibit possible emotional response (when calling a murder – *friendly fire*) [8]. Although from a logical point of view the variation in labeling doesn't affect the truth value of the utterance – humans are quite sensitive to such semantic shifts, they can be manipulated and even enjoy journalistic articles, rich with rhetorical figures. From this point of view – the design of a computer agent, sensitive to semantic shifts, or at least to variations in nomination and grammar constructions – is quite promising.

On the other hand – emotional processing may affect our representation of an event or an incoming text. When we are hungry, we overestimate our appetite, and when we walk in the dark, we may overestimate the danger of an approaching pedestrian. Clore and Ortony describe this as top-down emotional processing [7]. The phenomena is also revealed in nomination: when we are irritated by the opponent, we overestimate in speech the intensity for most of his actions and such nominations may serve as an indirect cue of the speakers aggression [9].

We are designing our model to make it simulate some bottom-top and top-down emotional phenomena. The model operates with semantic structures, where each node (verb or an actant) is represented as a set of semantic markers, each marker having a variable value. Some scripts are sensitive to specific markers – they are defined as *key markers* in script structures. If a *key marker* appears in an input, it contributes to the specific script – simulating bottom-up process.

On the other hand, if a specific script is chosen to process the input, it applies its key markers to the semantic structure and the set of referents. For example, if we 'touch' the agent, who is in bad mood – he may choose a negative d-script (DANGER) for processing, it will apply the marker 'force' to the input semantic structure (or increase its value, if it is already present). For the agent it will seem, that we've just 'hit' him. When choosing the nomination for the event, the agent may choose the words *hit, strike* or to express this marker as an adverb.

As the agent operates with a set of referents (representations of all surrounding relevant objects), it will append the 'malefactor' marker AGGR to the party, 'touching him with force', and a victim marker VICT – to himself. These markers are used to name the objects and to orientate the utterance in a communication with many parties. Each utterance in the system has a pragmatic model – it defines, for example, that a phrase *You are an idiot!* should be addressed to AGGR, and *I always make people suffer!* should be produced by VICT. These markers are assigned by top-down emotional processing, and allow the agent to correctly address the utterances in communication.

The present version of the agent doesn't have any "memory" structure: a person is considered as AGGR until he does something good and AGGR is replaced with an antonymic marker BON. At the same time, the markers contribute the processing of input. If an AGGR-person performs some action (with no significant appraisal) that agent is likely to consider this as a bad action (as AGGR in the corresponding referent matches AGGR key marker in the definitions of negative d-scripts). So for short periods of time the agent simulates *a computed route to appraisal* in bottom-up processing [7]. It allows the agent to react not only to prototypically emotional situations, but also to situations with objects, which appeared in past emotional situations.

4. Combined emotional speech reactions

Interaction between (a) scripts, (b) microstates and (c) filtering – allows us to simulate several emotional phenomena, linked with processing and expression of semantics: in addition to the emotional oscillation (as shown earlier) these are *hypocrisy* and *sarcasm*. In all the following cases we consider that an input semantic structure strongly activates one negative d-script (which represents negative emotional arousal) and weakly activates other scripts: r-scripts and even some positive d-scripts (Fig. 4).

Hypocrisy. In the case of hypocrisy – filtering plays a major role in processing an output. It completely suppresses the negative d-script (its output moves to the "thoughts" of the agent) and raises the activation of the best "polite" r-script. This creates a great contradiction between the "thoughts" and "words" of the agent: it thinks '*You are an idiot!*' and says *It's not your fault*.

The cases of hypocrisy may be used to construct combined emotional cues where superficial expression of one emotion is colored by another (even contrary) internal affective state [10].

Expressive emotional oscillation. In this case the input activates the same two scripts, they are not suppressed and appear in speech one after another; the order is controlled by microstates. The agent seems to blow up – and then calm down: it says *You – idiot!* and in a moment continues – *It's all right! It's not your fault!*

Irony and sarcasm. In the case of sarcasm the agent completely suppresses the activated negative d-script and has to find a way to canalize cumulative activation. As the system activates all the scripts to different degrees, when processing an input – we can look for the best activated script with the opposite sign (a script – member of the opposing microstate). If the agent is irritated, because he believes 'someone beat him' (DANGER negative d-script), then the opposite reaction is to be pleased, because someone pays attention to the agent and makes some social action towards him. This reaction corresponds to ATTENTION positive d-script, which is the best activated positive d-script during processing of the input. In normal conditions this script is activated by input phrases like *You are so nice!* and may force the output *It's so nice, you've paid attention!*

In this situation this script is used to express the opposite activation. The agent says *It's good you have paid attention to me!* to express the opposite meaning (*'you beat me! you are an idiot!'*). During the expression the output may be modified by different irony markers, like gestures, smiles and a grammar marker like "at least".

The advantage of the model is that "emotional oppositions" are defined on microstates, which have no semantics. Microstates are linked with scripts, which have semantics and activate, depending on the situation. Different input situations may activate different scripts to be used for the expression of irony – the script shall depend on the structure of the situation, while the opposition of microstates shall remain the same.

5. Corpus studies

Corpora offer reliable and diverse material to observe emotional cues and account them in computer architectures [11]. There are several approaches to collect and organize corpora data.

On the one hand – we can record real interaction or make people interact with the interface, considering all the incoming data as a corpus and applying different formal computer methods to annotate the corpus data and extract the classes of annotations (three very different examples of this approach include [12, 13, 14]). On the other hand – we can define the annotation classes (like a list of emotions), and ask actors to perform and express the defined units. This approach is implemented in several emotional corpora, for example in GEMEP [15]. The intermediate approach is to record people interacting in real or experimental situations – with further manual annotation of data.

Our main object of study is the expression of emotions, caused by the competition of opposing scripts (oscillation) or by contradiction between activation and filtering. So in our case it was promising to study real situations, combining high motivation (activation of scripts) with strong social limitations (filtering). Corpus data bring additional observations to be accounted in our computer model.

We have collected 236 video records of oral university exams with a duration from 2 to 30 minutes: the corpus covers tests on 3 courses at 5 humanitarian faculties. In each record a student gives a form of oral answer: upholds his written work, or gives definitions to the terms listed in an examination card.

Although the exam situations are not frequent in real life, it represents a prototype interaction for the future use of computer agents. Students have to execute the exam task or leave the impression of competence and achieve the exam mark in another way – through the usage of some communication strategy and through influence on the examiner. In the same way, a mobile robot, facing an impracticable request, may simulate different forms or combined emotional reactions and try to influence the user to receive the appreciation not through successful performance, but through some cute attempts.

The behavior of students and/or examiners was annotated with a help of ELAN video annotation tool. We have annotated utterances of the participants, gestures and movements, performed by (a) eyes, (b) mouth, (c) head, (d) body and (e) hands. During the annotation we have paid attention to the cases, when the performance differs from some "neutral" behavior, thus conveying some emotional expression. For the purpose of the current work we have annotated the cases, when a

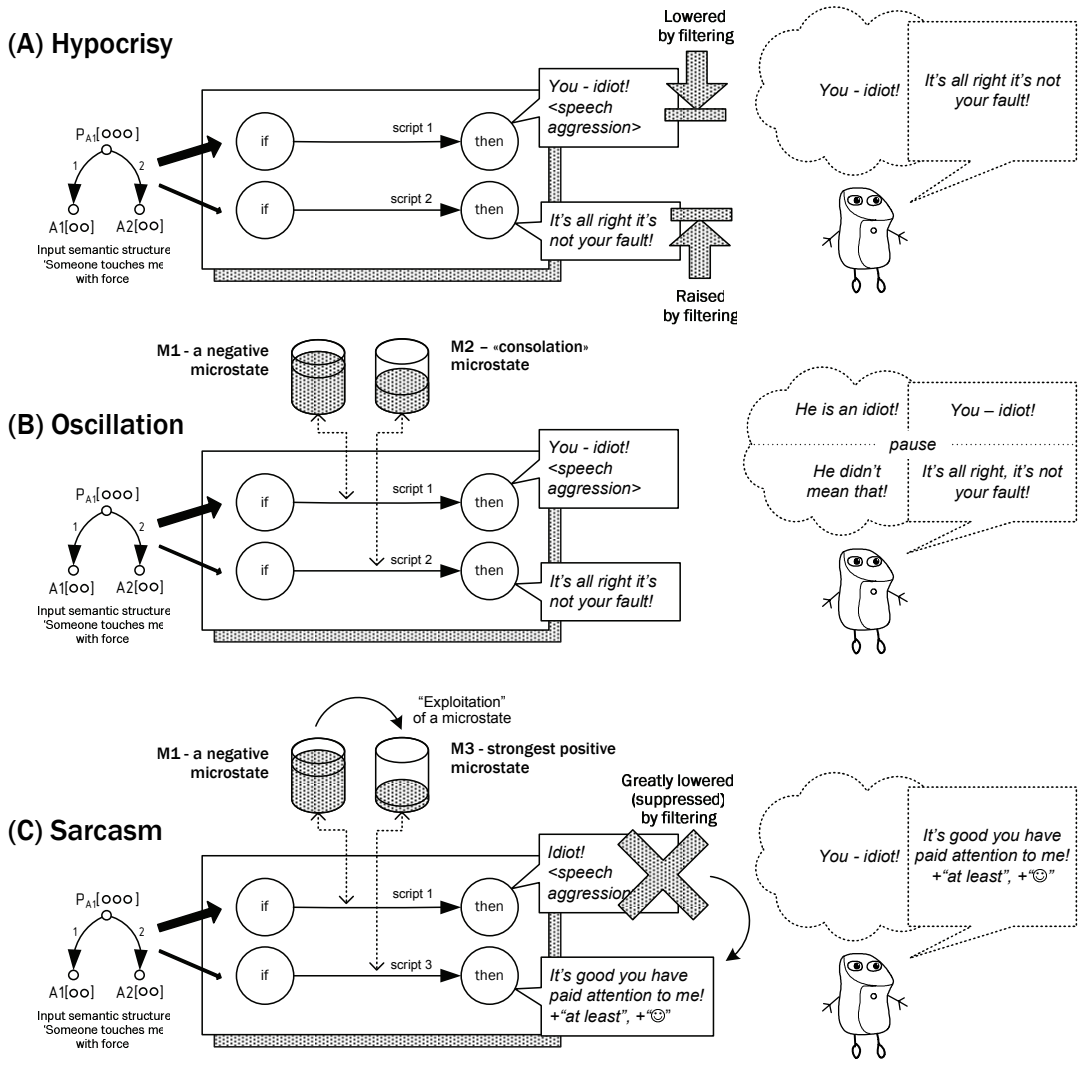

Figure 4: Some types of *hypocrisy*, *emotional oscillation* and *sarcasm* may be simulated as the interaction between the activation of scripts, changes in microstates and filtering. (A) In the case of *hypocrisy,* filtering suppresses inappropriate scripts (pushes their output to "thoughts" of the agent) and rises "polite" scripts, which crates the contradiction between "words" and "thoughts" of the agent. (B) In the case of *emotional oscillation,* the order of scripts is controlled by microstates, scripts appear one after another, simulating changes in moods or emotions – both in "words" and "thoughts". (C) In the case of *sarcasm,* the negative script/microstate pair is completely suppressed by filtering and the system "exploits" the best positive microstate in order to express the suppressed activation – output may be modified by the ironical cues.

certain gesture or phrase is not completely performed (not finished or interrupted) or when several emotional cues form a sequence (possible case of emotional oscillation). We have annotated the cases of laughter and irony, taking into account both – external cues (smiles, "breath spasms", aspiration and pitch variation) and text pragmatics (violation of Grice maxims). The following extensions are the subject of our current work.

(a) **Expressive oscillation and irony.** Laughter and ironical cues frequently appear at the end of a sentence, risky from the point of view of the speaker [16]. A student may smile and change the pitch, while suggesting a questionable answer or asking for another attempt. Students may show cues, usual for much stronger emotions – like growling, twitching, whining etc. – combing the cues with irony markers. This way of expression allows us to reveal the internal arousal, while refraining from being criticized for doing so.

The examples show, that the agent may quite overtly express the activated d-scripts, combining them with irony markers – without the necessity to exploit the opposite microstate and violate Grice maxims.

(b) **Expression of confusion.** In a situation of hesitation or confusion a student may consecutively perform up to 7 different actions, like scratching himself, adjusting hair, clothes, manipulating objects on the table, etc. From the point of view of the computer architecture the confusion is expressed here in a way, similar to the mechanism of irony. Since confusion has quite a few expressive cues, its direct expression here is substituted

by a number of reactions, usual for quite different input conditions: like itch, discomfort, etc.

To simulate the phenomena we define a special microstate, responsible to such minor reactions – 'discomfort'. When activated, the 'confusion' microstate (which has no specific output) starts to exploit the 'discomfort' microstate – the agent pays attention to the existing (even minor) irritators, and substitutes the absence of required performance with scratching, adjusting and void manipulations.

(c) **Role-taking.** Irony and laughter are frequently combined in corpus with role taking, where a student ironically commands the examiner or an examiner undertakes to solve the problems from an experimental paper by the student.

This suggests that the pragmatic model of each utterance should include not only the emotional markers (like AGGR, VICT and BON), but also some role-markers. Through violation of role-markers the agent shall be able to construct ironical utterances, in normal conditions corresponding to another role situation.

(d) **You-goals and script coordination.** Many emotional reactions are the expressions of internal arousal, while others are simulated to influence the addressee (achieve you-goal, contrary to me-goal [17]). In a situation of tension students may quickly jump between several strategies of emotional influence: (a) showing self-inadequacy and asking for indulgence, (b) watching the reactions of the examiner and trying to guess the correct answer, (b) categorically demanding accuracy from the examiner – where each strategy may last for only 3-4 seconds, expressing nervousness through poor coordination and frequent alternation.

While the emotional oscillation of the computer agent, expressing me-goals, may show attractive emotional dynamics (as shown earlier), the oscillation on you-goals may express nervousness through insufficient coordination of expressive strategies.

(d) **Jumps to emotional interaction.** Students may suggest to the examiner some mode of emotional interaction (leniency, jokes, laughter, etc.), when they fail to perform formally within the rules of the suggested exam task.

For each type of rational interaction (interaction within r-scripts) the agent may constantly check the closest d-scripts. In case of failure in performance the agent may suggest jumping to the closest emotional interaction – where it can still expect to achieve some benefits in communication.

6. Conclusion

The framework of activation and filtering applied to operations with semantic structures may help to simulate some examples of sophisticated emotional phenomena – like changes in representation of events (top-down emotional processing), emotional oscillation, sarcasm, irony and indirect expression of confusion.

The accurate design of the indicated functions may serve as the core technology to create believable and attractive computer agents, implemented in computer interfaces and mobile robots.

References

[1] M. Minsky. The Society of Mind, New-York, London: Touchstone Book, 1988.

[2] A. Sloman and R. Chrisley. Virtual Machines and Consciousness. Journal of Consciousness Studies. 10(4-5): 133-172, 2003.

[3] C. Becker, S. Kopp, and I. Wachsmuth. Simulating the Emotion Dynamics of a Multimodal Conversational Agent. ADS 2004, LNAI 3068, Springer-Verlag: Berlin, Heidelberg. 154–165, 2004.

[4] A. Kotov. Application of Psychological Characteristics to D-Script Model for Emotional Speech Processing. ACII 2005, LNCS 3784. Springer-Verlag: Berlin, Heidelberg. 294-302, 2005.

[5] W. Yeh and L.W. Barsalou. The situated nature of concepts. American Journal of Psychology. 119(3): 349-384, 2006.

[6] R.W. Langacker. Concept, Image and Symbol: The Cognitive Basis of Grammar, Berlin: Mouton de Gruyter, 1991.

[7] G.L. Clore and A. Ortony. Cognition in Emotion: Always, Sometimes, or Never? Cognitive Neuroscience of Emotion, Oxford Univ. Press. 24-61, 2000.

[8] R.M. Blakar. Language as a means of social power. Pragmalinguistics, Mouton. 131-169, 1979.

[9] M.Y. Glovinskaya. Hidden hyperbola as a means to express and conceal speech aggression. Sokrovennyje smysly (rus): Moscow. 69-76, 2004.

[10] M. Ochs, et al. Intelligent Expressions of Emotions. ACII 2005, LNCS 3784, Springer-Verlag: Berlin, Heidelberg. 707-714, 2005.

[11] M. Rehm and E. Andre. From Annotated Multimodal Corpora to Simulated Human-Like Behaviors. Modeling Communication, LNAI 4930, Springer-Verlag: Berlin, Heidelberg. 1-17, 2008.

[12] N. Campbell. Technology and Techniques for Talking Together. The Third International Conference on Cognitive Science: Moscow. Vol. 2, 533-534, 2008.

[13] L. Aryananda. Out in the World: What did the Robot Hear and See? Proceedings of the Fifth International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems, Lund University. 131-132, 2005.

[14] M.S. Magnusson. Discovery of T-Templates and Their Real-Time Interpretation Using Theme. Probing Experience. J. H. D. M. Westerink et al. (eds). Philips Research, Vol. 8. Springer. 119-126, 2008.

[15] T. Bänziger and K.R. Scherer. Using Actor Portrayals to Systematically Study Multimodal Emotion Expression: The GEMEP Corpus. ACII 2007, LNCS 4738, Springer-Verlag: Berlin, Heidelberg. 476-487, 2007.

[16] R.R. Provine. Laughter punctuates speech: Linguistic, social and gender contexts of laughter. Ethology. 95: 291-298, 1993.

[17] R.C. Schank. Tell me a story: narrative and intelligence, Evanston, Illinois: Northwestern University Press, 2000.