Conceptual Operations with Semantics for a Companion Robot

Artemiy Kotov[1, 2* [0000-0003-3353-5549]], Liudmila Zaidelman[1, 2 [0000-0002-2941-144X]], Anna Zinina[1, 2 [0000-0001-9575-1875]], Nikita Arinkin[1, 2 [0000-0003-2303-2817]], Alexander Filatov[1], Kirill Kivva[1,3]

[1] Russian State University for the Humanities, Moscow, Russia
[2] National Research Center "Kurchatov Institute", Moscow, Russia
[3] Bauman Moscow State Technical University, Moscow, Russia
kotov@harpia.ru

Abstract. We study the features crucial for a companion robot conceptual processing and suggest a practical implementation of a cognitive architecture that support these features while operating a real F-2 companion robot. The robot is designed to react to incoming speech and visual events, guide a person in a problem space and accumulate knowledge from texts and events in memory for further dialogue support. We show how a conceptual representation system designed for a companion robot deals with several types of conceptual representations: text semantics, sets of emotions and reactions, operations in a problem space and in semantic memory. We also suggest a conceptual representation based on linguistic valency structures (semantic predication) that is suitable to link the processing components. The general processing architecture is based on the production approach: it may trigger several scripts and combine speech and behavioral outputs of these scripts on the robot. The system performs conceptual operations with semantics while processing texts on a server in a standalone mode, or while controlling the robot in a dialogue mode, or assisting a user in solving Tangram puzzle.

Keywords: Conceptual Processing, Knowledge Base, Production Logic, Semantic Representation, Emotional Computer Agent.

1 Introduction

Conceptual processing systems designed for companion robots should handle not only traditional linguistic tasks like dialogue support and question answering, but also perform visual recognition and problem-solving. In each of these tasks a system operates with diverse conceptual representations: text semantics, scene setup, own intermediary inferences, and memory. These representations should be shared between the main processing modules, and converted, where required. We develop *F-2* robot with conceptual processing system aimed at the integration of natural communication, problem solving and visual comprehension tasks. The system is designed to demonstrate the following interconnections between the information processing modules:

1. Reactions of the robot should be invoked by incoming events of different nature: utterances, user movements or game actions. All these events depict or denote the real-world events, and thus may be processed in a compatible way.
2. Different reactions should compete and may output diverse and even contradictory behavioral patterns, as suggested by the model of *proto-specialists* [Minsky, 1988] or *CogAff* architecture [Sloman, Chrisley, 2003]. In particular, emotional processing may compete with rational processing, and the compound rational/emotional behavioral patterns (blending reactions) may be executed by the robot.
3. An emotion may influence a conceptual representation (*frame*) of an incoming event that is described as *top-down* emotional processing [Clore, Ortony, 2000]. Further, conceptual representation of a situation may influence a specific concept, as suggested by the *situational conceptualization* theory [Yeh, Barsalou, 2006].
4. The system should combine *reactive* processing scheme, applicable to immediate emotional reactions and speech replies, and *goal-oriented* processing scheme, applied to compound plans and problem solving.

Unlike the majority of neural networks, the system also has to keep a conceptual representation in a readable form for research purposes. This form should also allow the system to store incoming events in a memory base for further knowledge retrieval.

2 Cognitive and dialogue support architectures

Conceptual processing systems are implemented in several areas of cognitive and computational research. In particular, they are required (a) to control interactive artificial organisms – robots and virtual agents, (b) to model human logic, natural or scientific inferences, (c) to simulate human competence in problem solving, and (d) to support natural dialogue regarding problems and actions. In linguistics such systems are used (e) to extract and classify speech semantics, (f) to make inferences basing on text semantics, and (d) to provide speech responses in a dialogue. In this publication it is only possible to give a bird's eye view over this vast scientific area.

Shank et al. have introduced *scripts* as a basic model for natural inferences in his classic works on conceptual processing [Schank, Abelson, 1977]. Scripts allowed the system to model typical sequences of actions (like attending a restaurant), to reconstruct missing facts from a text and thus to support question answering on the missing data. The system, designed by Dorofeev and Martemianov [1969], is another classic example of text comprehension engine: the system extracted semantic predications from a fairy-tale and constructed possible outcomes in each situation. For the action graph of the outcomes it anticipated *good* actions for the protagonist and *bad* actions for the antagonist, thus prognosing agents' actions. The system even had a concept of *soul*, which could be destabilized by external stimuli, forcing the agent to operate on text semantics until the *soul* is finally balanced. One of the first systems of conceptual representation linking text semantics and problem space – SHRDLU – was designed to handle the representation of a real situation (arrangement of blocks) and to suggest the appropriate actions [Winograd, Flores, 1987]. Within the development of F-2 conceptual processing system, we mostly rely on SOAR architecture, designed for operations in problem space and dialogue support tasks [Laird, Newell, Rosenbloom, 1987]. SOAR also relies on scripts (productions) to consider the possible moves in the problem space. It may suggest moves to a user, once the script graph successfully reaches the target state (solution) of the problem.

Minsky has suggested to support emotion processing with a limited set of *proto-specialists* [Minsky, 1988]. *Proto-specialists* of an agent suggest the reactions in case of danger or urgent lucrative opportunity. Sloman has extended this approach in his CogAff architecture: it was suggested that *reactions* and *alarms* (units of the basic *reactive* level) compete with “rational” inferences on the level of *deliberative reasoning* and with the processes of introspection on *meta-management* level [Sloman, Chrisley, 2003]. Sloman has suggested that *rational processing* is more accurate in the recognition and provides better planning, while *reactions* are fast and shallow: they ensure quick response in critical situations. It was suggested that *secondary* or *tertiary* emotions may combine rational end emotional units, like, triggering an emotional response by a rational inference or by a meta-management process, constantly returning the thoughts to the emotional image. The compound nature of emotional responses has also been studied in linguistics. As noted by Sharonov, an emotional event may trigger numerous emotional and etiquette speech reactions, which linearize in time to the series of (a) interjections – *Oh!* (b) emotional evaluations – *God!* (c) emotional classifications – *What a mess!* (d) acquisition of speaker’s responsibility – *What have I done!* and (e) etiquette replies – *I’m so sorry!* [Sharonov, 2008]. The order is defined by the processing difficulty, as the primer segments are more expressive and are generated faster, while latter segments require more resources and time.

Dialogue support systems and automatic question answering is another fast-growing approach to semantic representation and the simulation of reactions (dialogue turns). The classic papers by Jurafsky [2000; Jurafsky, Martin, 2019] present a comprehensive overview of dialogue support and question-answering systems. Dialogue systems are usually divided into rule-based, information-retrieval and statistical systems [Jurafsky, 2000; Jurafsky, Martin, 2019; Cahn, 2017], while modern systems combine the three approaches. It was shown that most of the participants at Amazon Alexa Prize competition [Ram et al., 2018] used rule-based approaches, while boosting the performance with neural networks and machine learning algorithms such as: Hierarchical Latent Variable Encoder-Decoder [Serban et al., 2017], a two-level Long Short-Term Memory network [Adewale et al., 2017] and others. In this sense, a dialogue support system may be generally considered as production architecture, where an input utterance triggers a script and the best response is further selected.

3 F-2 robot conceptual architecture

Robot may process texts or visual events at its input. Texts can be received from a text source (a text file or RSS subscription) or as an oral speech, in this case Yandex Speech API is used to decode the signal to the written form. The audio recognition may return several ambiguous results, which are processed in parallel up to the stage of scripts, where the preferred variant is selected. Robot is also equipped with two cameras (the number can be easily extended) and may also receive the information on the location and movement of different objects, in particular, it processes the location of faces and the location of game elements in Tangram puzzle. The general architecture of the system is represented on Fig. 1.

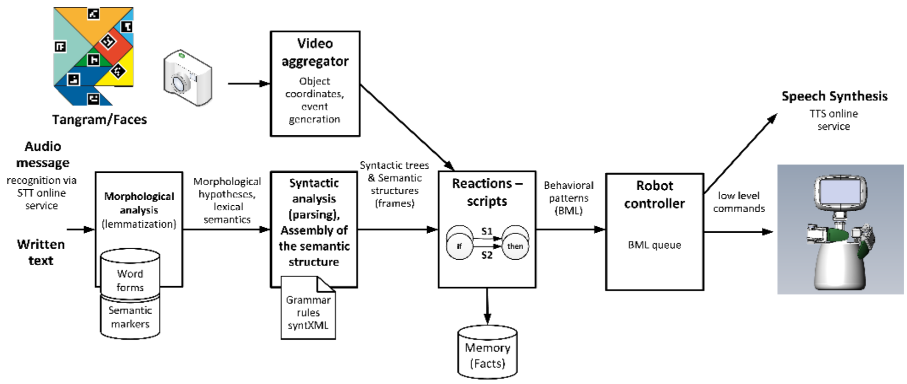

Fig. 1. General architecture of the conceptual processing system.

Similar to classic conceptual processors and dialogue support systems, we use a set of scripts at the core of the system to classify an input (both speech and visual) and to suggest an appropriate reaction. For each input the system selects the corresponding scripts: it calculates the distance between the input and scripts' premises, sorts the scripts following the reduction of the similarity and processes the topmost scripts from the list, e. g. executes the actions attached to the scripts. The scripts are divided into 3 groups: (a) *dominant scripts* or *d-scripts* for the emotional processing, like INADEQ for *They lie to you!* or INADEQ for *They are crazy!* [Kotov, 2003], (b) *behavioral rules* – etiquette and dialogue support routines, which cause certain behavior in a given situation, like an excuse or a speech response, and (c) *rational scripts* or *r-scripts* for rational classification of input stimuli and the simulation of inference (perspective component). Within the proposed architecture, the scripts are used (a) as the dynamic model of emotions and reactions, expressed in verbal and nonverbal behavior of the robot, (b) as the representations of *regular situations* for the resolution of ambiguity, (c) as indexes for judgements in the memory base. The system relies on the semantic representations, constructed by syntactic parser and by visual recognition system. These representations are also saved to memory and may be retrieved for an output. In this sense, semantic representations link the components of speech perception, visual awareness, inference, memory, and performance. Although *production* approach is frequently contrasted to neural networks, we implement a network-like evaluation of semantic markers to calculate the distance between an input and scripts. Further, a neural network is used to evaluate binding within a syntactic tree in order to select "the best" trees in case of ambiguity.

A script may be annotated by a behavioral pattern: a combination of speech, facial expression and/or gestures. These patterns are described in Behavior Markup Language – BML [Kopp et al., 2006]. BML packages from the activated scripts are transmitted to the robot controller, it monitors the activity of the robot and executes BML packages from scripts with higher activation as well as the compatible scripts, e. g. packages *A* and *B* can be executed simultaneously, if *A* applies to *head* and *face*, while *B* describes a *hand gesture*. Several scripts may be simultaneously invoked by a stimulus, and their output may be combined, like an emotional exclamation (*Oh!*) may be followed by a full phrase (*Usually, an intuition deceives a person!*), while anxiety is simultaneously expressed non-verbally through *automanipulation*: robot joins hands or scratches its own body (figure 2).

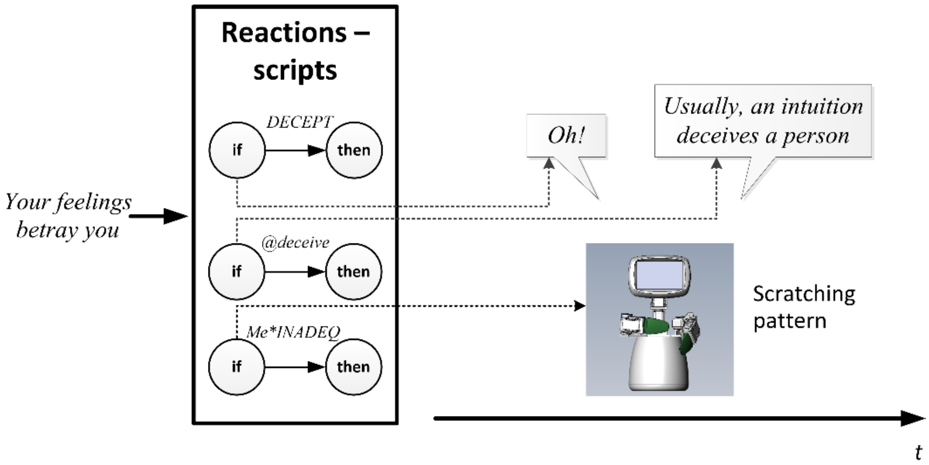

Fig. 2. Generation of behavioral patterns by different scripts in time.

3.1 Video recognition

Video recognition should allow the robot to interact with a user as well as with a problem space. For the interaction with a user we use **face_recognition/dlib** libraries that enable the robot to detect faces in video stream and associate faces with known referents. The component generates events like ‘John is_present’ to invoke the reaction of interest on the robot – robot moves gaze direction towards the person, moves eyebrows, etc. To model the interaction with the user while solving a problem we have chosen *the Tangram puzzle*. We have implemented an interaction scenario, where the robot controls user moves through an automatic video recognition system and suggests optimal moves, guiding the human through the problem space according to a solution script. Robot stores the solution (or several alternative solutions), evaluates each move by the player as a positive or negative action in respect to the closest solution and suggests the next move. In case the user switches to another possible solution, the recognition system switches accordingly. Unlike in SOAR architecture, the solution graph is designed from top to bottom: starting from the solution combination (goal, final state) to the specific moves required in the situation (current state) (Fig. 3).

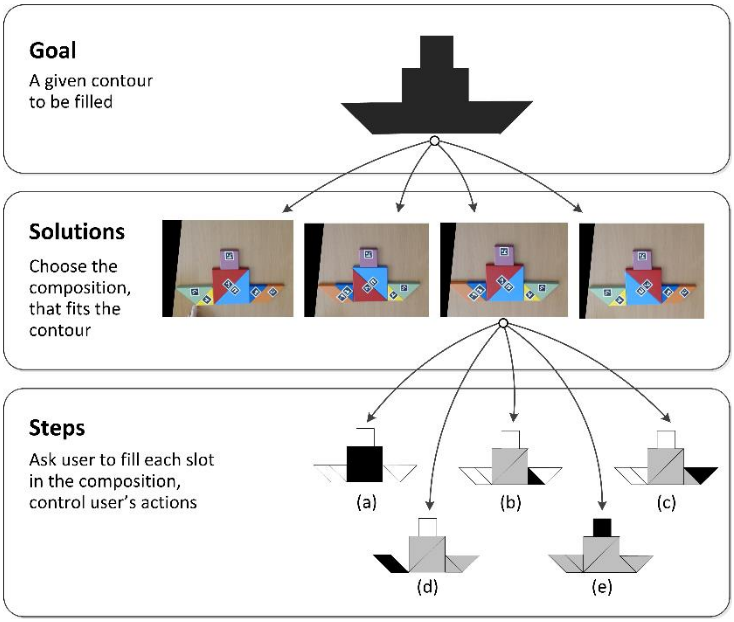

Fig. 3. Tangram solution tree.

The recognition system generates the events in a format similar to text semantic representations: 'user moved element *a*' and 'game elements *a, b, d* are correctly placed for the closest solution *Y*' (see 3.6). This, in turn, allows the robot to construct representations 'user moved element *a* correctly for the solution *Y*' and react to this event. These representations of user actions can be used not only by the reactions, but, potentially, by text synthesis system for flexible discussion on the user actions. The approach may also utilize an external problem-solving component, which constructs a solution for a given puzzle in a form of a script path and guides the user through the path or discusses the process of solution.

3.2 The extraction of valencies from syntactic structures

Each recognition result is processed by syntactic parser with the Russian grammar containing over 600 syntactic rules in syntXML format [Kotov, Zinina, Filatov, 2015], as a result, a dependency syntactic tree is constructed (example in Fig. 4).

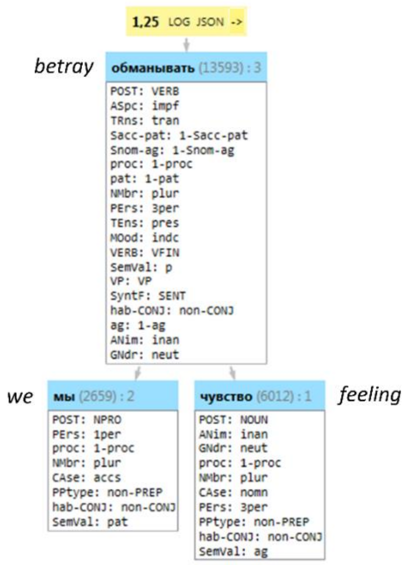

Fig. 4. Syntactic tree for a sentence *Feelings betray us*.

Speech processing module should represent the utterances in a form sufficient to calculate the distances to the premises of scripts. We rely on shallow semantic representations: semantics of a single clause is represented as a *semantic predication* – a set of semantic markers distributed between semantic valencies. A predicate is assigned to *p* (*predicate*) valency, while actants are assigned to *ag* (*agent*), *pat* (*patient*), *instr* (*instrument*) etc. – following an extended list by Fillmore [Fillmore, 1968]. For a compound sentence several semantic predications are constructed with co-reference links. In case of ambiguity, a set of syntactic trees is constructed, all the trees are processed in parallel and evaluated by neural network to select the best trees for further steps of analysis. Semantics of each valency is the aggregation of semantic markers of all the words within the corresponding subtree, e. g. for a *noun phrase* a semantic representation is a sum of markers for the *noun* and all the *adjectives*. Lexical ambiguity is represented as a Cartesian product of semantic sets for ambiguous lexemes.

3.3 Semantic markers in valencies

Words in the semantic dictionary are annotated by semantic markers, so that similar words have the greater number of common markers. We use a set of 4778 markers consisting of (a) *focal markers* of d-scripts (see 3.4), like ‘liar’ – typical agent in the situation of deception (209 markers), (b) markers from a semantic dictionary

[Shvedova, 1998], (669 markers) , and (c) semantic markers assigned to words after clustering of *word2vec* vectors (3900 markers). Markers, based on *word2vec*, are assigned basing on clustering of nouns to 2000 clusters and 600 “superclusters”. In this way we get a two level “ontology” with a “basic level” and a “top level” marker for each word. In a similar way, verbs were divided into 1000 clusters and 300 superclusters. Table 1 shows a semantic representation (predicate structure) for an utterance *Our feelings betray us*. Lexical semantics is distributed into “1 1”/ “1 2” slots (as verb *betray* may be a communicative or behavioral action). Markers assigned after *word2vec* clusters are indicated by “@”.

Table 1. Semantic representation of *Our feelings betray us (/lie to us)*.

Betray: **P** (predicate)	*Our feelings:* **Ag** (agens)	*Us:* **Pat** (patient)
1 1 present tense	1 1 many	1 1 somebody
1 1 assertive	1 1 abstract	1 1 egocentric – *me*
1 1 to communicate	1 1 negative emotions	1 1 includes another person
1 1 DECEPT attribute	1 1 positive emotions	1 1 physical object
1 1 to report	1 1 @feeling_NOUN	1 1 principal – *speaker*
1 1 @to_simulate	1 1 @173_NOUN	1 1 set of people
1 1 @214_VERB	1 1 own	
1 2 present tense		
1 2 assertive		
1 2 social action		
1 2 DECEPT attribute		
1 2 @to_simulate		
1 2 @214_VERB		

Within the research of conceptualization Barsalou has noted that the structure of a notion, e. g. *chair*, depends on the situation where the notion appears, e. g. *kitchen, cinema, hotel* or *move a chair, sit on a chair, buy a chair* [Barsalou, 1992]. This observation has been categorized with a set of rules of conceptualization [Yeh, Barsalou, 2006]. Following these assertions, a set of semantic features within a notion (concept) is affected by the *frame* of the situation. A similar process is described by the psychology of emotions as emotional *top-down* processing: subjective representation of a situation can be changed by an invoked emotion [Clore, Ortony, 2000], e. g. a person tends to overrate his ability to eat when being hungry, and overrates a threat or danger when being frightened. As an incoming event may contain not all the relevant markers, defined in script premise, *focal markers* [Glovinskaya, 2004] are applied to the incoming representation on a *top-down* basis: an input event is attracted by the scripts, e. g. “gets more emotional”.

3.4 Input processing with scripts

Rational scripts (r-scripts) provide rational inference from the input predications and also contribute to the resolution of ambiguity. Each input predication is associated with an r-script, in case of ambiguity, the system selects the closest representation. To design the scripts, we have aggregated predicative structures, where all the words in each valency constitute a coherent set, e.g. {*team, sportsman, champion, player, hockey_player*} *defeated* {*host, guest*} or {*candidate, promotee*} *defeated* {*mayor, governor*}. In this sense, each script searches for a prototype situation in the incoming stimuli. We have applied the following methods to construct the premises of scripts:

1. Following the clustering of verbs, we have selected 1391 verbs with relevant verb frequency > 0,02% (percent of all the verbs in the text corpus), but and at least one verb from each *word2vec* cluster. For each verb we have defined the most frequent words in each valency in the text corpus (over 80 million wordforms), collected by the previous runs of our parser. The results were manually inspected: in case words from two (or more) different superclusters occupied a valency, the whole *frame* was divided into two or more frames.
2. For each verb in the dictionary we have selected facts, where all the words in *agent* valency belong to one supercluster, while words in *patient* valency (for transitive verbs) belong to another supercluster. So, different *frames* were constructed for different verb senses:
   (i) {*finger, palm, hand*} *gripes* {*shoulder, finger, palm, hand*}
   (ii) {*anxiety, fear, depression*} *gripes* {*neck*}

1619 scripts were constructed after the procedure. For r-scripts, a typical speech answer was designed as a speech output reaction as the sequence of words in the valencies. The scripts are associated with *microstates*, which change the sensitivity of scripts and simulate current emotional profile: *nervous, calm, motivated, hypocritical*, etc. An incoming stimulus invokes several scripts, which generate compound reactions, as represented in Table 2. As the sensitivity of each script is proportional to the activation of microstate, one can design (a) an *emotional mode*, when d-scripts DECEPT and Me*INADEQ are preferred, (b) a *rational mode*, where r-scripts are preferred, or (c) *aggressive* or *depressive* agent (DECEPT and Me*INADEQ are preferred respectively).

Table 2. Distribution of script activation for *Our feelings betray us (/lie to us)*

Similarity	Script	Speech output
0,2231	@deceive	[Usually] *An intuition deceives a person*
0,2131	DECEPT (d-script)	*Everybody lies!*
0,1593	@inflate/puff[1]	[Usually] *Girl puffs lips*
0,1561	Me*INADEQ (d-script)	*I do something stupid!*
0,1559	@lie	[Usually] *Man lies to a man*

[1] The script @inflate/puff is activated due to speech homonymy and may be used for a speech game or humorous response.

The order of scripts at the top of activation list is sensitive to the distribution of valencies in a particular communication. In case the phrase addresses *you*, not *us* (e. g. *Feelings betray you*), the activation of scripts is more emotional: DECEPT (d-script) is the leading, as the robot personally interprets incoming phrases with the pronoun *you*. Scripts DECEPT, @deceive and Me*INADEQ can construct multimodal behavior for the robot in time; @inflate/puff and @lie scripts are suppressed, as belonging to the same microstate as @deceive. In this case DECEPT is getting the highest initial activation due to its high sensitivity and generates an interjection thus discharging the activation. @deceive generates a verbal response and Me*INADEQ generates automanipulation as a sign of confusion. Since DECEPT and @deceive both generate speech outputs, they compete in time for the robot's mouth and are expressed one after another (Figure 2). A script may generate many behavioral response patterns for different parts of the robot's body – these patterns are kept in buffer, and compatible patterns (e. g. corresponding to different parts of the body) are selected for execution.

3.5 Memory

The semantic representation of input sentences is indexed and saved to a database to support long-term memory and perspective question answering. A base of 1 million facts is automatically collected for 80 million wordform corpus. Phrases are indexed by the script, used to select the meaning, and can be retrieved by the script index. The base can return all the predications, which correspond to the script premise, or an arbitrary semantic pattern, for example, corresponding to the semantics of an input question. Table 3 represents sentences, returned for the pattern 'feelings deceive'. We suggest that question answering may retrieve utterances from the array of all the analyzed texts, or from a subbase with manually inspected utterances, or with judgements from a trusted source (e. g. a schoolbook). An advantage of the method is that the knowledge of an agent is directly enriched through the automatic analysis of incoming texts, without any retraining. We also expect that scripts can provide more flexible Q&A by training on Q&A pairs, where a script premise corresponds to the question and script inference corresponds to the answer.

Table 3. Judgements for *Our feelings betray us* **semantic pattern in the memory base**

Fact	Sentence
1094963	*anger troubled him, pushed him to insolence.*
2527406	*the fact is that physical sensations deceive them, because…*
8446093	*no, you know that feelings do not deceive you.*
146731	*Is it possible, that the sixth sense deceives him?*

4 Conclusion

We suggest that each object and action within the surrounding of the robot may be represented by a list of markers and the structure of the situation – as a distribution of the referents between the valencies. Text semantics and visual recognition results may thus be represented in a unified way. We consider the main contribution of the work as a practical implementation of proto-specialists classic theory and the extension of CogAff architecture to the area of semantic processing and event representation in semantic form. Further, the interaction between the incoming representations and scripts provides a practical implementation of the theory of situated conceptualization, as the input semantics changes depending on the chosen reaction (script). The general architecture inherits a classic *production* approach and simulates balanced and parallel rational/emotional processing of contradictory reactions. This allows the robot to construct compound behavioral patterns (blending reactions), as suggested within the linguistic analysis of multimodal communication. The whole design of conceptual processing system is implemented and works in a standalone mode on a server, accumulating and processing news and blogs on an everyday basis, as well as on F-2 robot in the Tangram support mode, and a dialogue mode. In sum, the combination of *semantic predications* as the form of representation, and *scripts* as a processing architecture, suggests a cognitive architecture with features, essential for a companion robot.

5 Acknowledgment

The present study has been supported by the Russian Science Foundation, project No 19-18-00547.

References

1. Adewale, O., Beatson, A., Buniatyan, D., Ge, J., Khodak, M., Lee, H., Suo, D.: Pixie: A social chatbot. Alexa Prize Proceedings (2017).
2. Barsalou, L. W.: Frames, concepts, and conceptual fields. Frames, fields, and contrasts: new essays in semantic and lexical organization. Hillsdale, N.J.: L. Erlbaum Associates. pp. 21–74 (1992).
3. Cahn, J.: CHATBOT: Architecture, design, & development. University of Pennsylvania School of Engineering and Applied Science Department of Computer and Information Science (2017).
4. Clore, G. L., Ortony, A.: Cognition in Emotion: Always, Sometimes, or Never? Cognitive Neuroscience of Emotion. Oxford Univ. Press. pp. 24–61 (2000).
5. Dorofeev, G. V., Martemyanov, Yu. C.: The logical conclusion and identification of the relationships between sentences in the text. Machine translation and applied linguistics. vol. 12. pp. 36–59 (1969).
6. Fillmore, C. J.: The Case for Case. Universals in linguistic theory. New York: Holt, Rinehart & Winston. pp. 1–68 (1968).
7. Glovinskaya, M. Ya.: Hidden hyperbole as a manifestation and justification of verbal aggression. Sacred meanings: Word. Text. The culture. M.: Languages of Slavic culture. pp. 69–76 (2004).
8. Jurafsky, D.: Speech & language processing. Pearson Education India (2000).
9. Jurafsky, D., Martin, J. H.: Speech and language processing (draft). Chapter Dialogue Systems and Chatbots (Draft of October 16, 2019).

10. Kopp, S., Krenn, B., Marsella, S., Marshall, A. N., Pelachaud, C., Pirker, H., Vilhjálmsson, H.: Towards a Common Framework for Multimodal Generation: The Behavior Markup Language. Intelligent Virtual Agents. pp. 205–217 (2006).
11. Kotov, A., Zinina, A., Filatov, A.: Semantic Parser for Sentiment Analysis and the Emotional Computer Agents. Proceedings of the AINL-ISMW FRUCT 2015. pp. 167–170 (2015).
12. Kotov, A. A.: Description of speech exposure in a linguistic. Computer Linguistics and Intelligent Technologies. M.: Nauka. pp. 299-304 (2003).
13. Laird, J. E., Newell, A., Rosenbloom, P. S.: SOAR: An architecture for general intelligence. Artif. Intell. vol. 33. № 1. pp. 1–64 (1987).
14. Minsky, M. L.: The Society of Mind. New-York, London: Touchstone Book (1988).
15. Ram, A., Prasad, R., Khatri, C., Venkatesh, A., Gabriel, R., Liu, Q., King, E.: Conversational ai: The science behind the alexa prize. arXiv preprint arXiv:1801.03604 (2018).
16. Schank, R. C., Abelson, R. P.: Scripts, plans, goals, and understanding: an inquiry into human knowledge structures. Hillsdale, N.J., New York: L. Erlbaum Associates (1977).
17. Serban, I. V., Sankar, C., Zhang, S., Lin, Z., Subramanian, S., Kim, T., Sotelo, J. M.: The octopus approach to the Alexa competition: A deep ensemble-based socialbot. Alexa Prize Proceedings (2017).
18. Sharonov, I. A.: Interjection in speech, text, and dictionary. M.: RGGU (2008).
19. Shvedova, N. Yu.: Russian semantic dictionary. Explanatory dictionary systematized by classes of words and meanings. M.: Azbukovnik (1998).
20. Sloman, A., Chrisley, R.: Virtual Machines and Consciousness. J. Conscious. Stud. vol. 10. № 4–5. pp. 133–172 (2003).
21. Winograd, T., Flores, F.: Understanding Computers and Cognition: A New Foundation for Design: Addison-Wesley (1987).
22. Yeh, W., Barsalou, L.W.: The situated nature of concepts. Am. J. Psychol. vol. 119. № 3. pp. 349–384 (2006).