

А.А. Котов, к. филол. наук, ведущий научный сотрудник, kotov@harpia.ru
З.А. Носовец, лаборант-исследователь, nosovets@phystech.edu
НИЦ «Курчатовский институт», РГГУ, г. Москва, Россия

КОНСТРУИРОВАНИЕ РОБОТОМ РЕПРЕЗЕНТАЦИЙ ПРОИЗВОДНЫХ СИТУАЦИЙ ПРИ ОБРАБОТКЕ СМЫСЛА ТЕКСТА

Аннотация: В проекте робота Ф-2 мы разрабатываем систему автоматического анализа текста (семантический парсер), которая строит семантическую репрезентацию полученных сообщений. В качестве развития этой системы мы разрабатываем систему сценариев – элементов естественно языкового вывода, которые используются для эмоциональной оценки поступивших сообщений (300 доминантных сценариев) и для конструирования репрезентаций производных ситуаций: возможных причин, следствий и интерпретаций для исходной ситуации (4500 рациональных сценариев). Комбинация доминантных и рациональных сценариев может использоваться для моделирования сложных когнитивных функций: активации эмоций по результатам естественно-языкового вывода (т. н. «вторичных эмоций»), моделирования воображения и мнений других участников коммуникации (theory of mind).

Ключевые слова: робот-компаньон, эмоциональная робототехника, когнитивные модели, поддержание диалога, человеко-машинное взаимодействие

A.A. Kotov, Ph. D., Senior researcher, kotov@harpia.ru
Z.A. Nosovets, laboratory researcher, nosovets@phystech.edu
NRC «Kurchatov Institute», RSUH, Moscow, Russia

CONSTRUCTION OF REPRESENTATIONS OF DERIVED SITUATIONS WHEN PROCESSING THE TEXT MEANING BY THE ROBOT

Abstract: Within the F-2 robotic project, we are developing an automatic text analysis system (semantic parser) that builds a semantic representation of the received utterances. As an extension of this system, we are developing a system of scenarios – elements of natural language inference, used for emotional evaluation of received messages (300 dominant scenarios) and for constructing representations of derived situations: possible causes, effects and interpretations for the initial situation (4500 rational scenarios). A combination of dominant and rational scenario systems can be used to model complex cognitive functions: activation of emotions based on the results of natural language inference (the so-called “secondary emotions”), modeling of imagination and opinions of other participants in communication (theory of mind).

Key words: companion robot, emotional robotics, cognitive models, dialogue support systems, human-machine interaction

Робот Ф-2 создаётся как исследовательский робот-компаньон для изучения коммуникации робота с человеком с помощью естественных средств: речи, мимики и жестов. Робот обрабатывает высказывания на русском языке, отвечает на прикосновения, а также распознаёт лица людей, мимику, направление и перемещение взгляда, движение игровых фишек головоломки танграм. Распознанное событие получает семантическую репрезентацию, аналогичную семантическому представлению предложения. Эти репрезентации активизируют сценарии робота, отвечающие за поведение и ответные реакции: речь и движения, описанные на языке BML [1] (см. рисунок 1). Данная архитектура управления роботом основывается на модели CogAff – Cognition and Affect Project [2], где каждое входящее событие обрабатывается инвентарём конкурирующих единиц – продукций или сценариев (см. блок сценариев на рисунке 1).

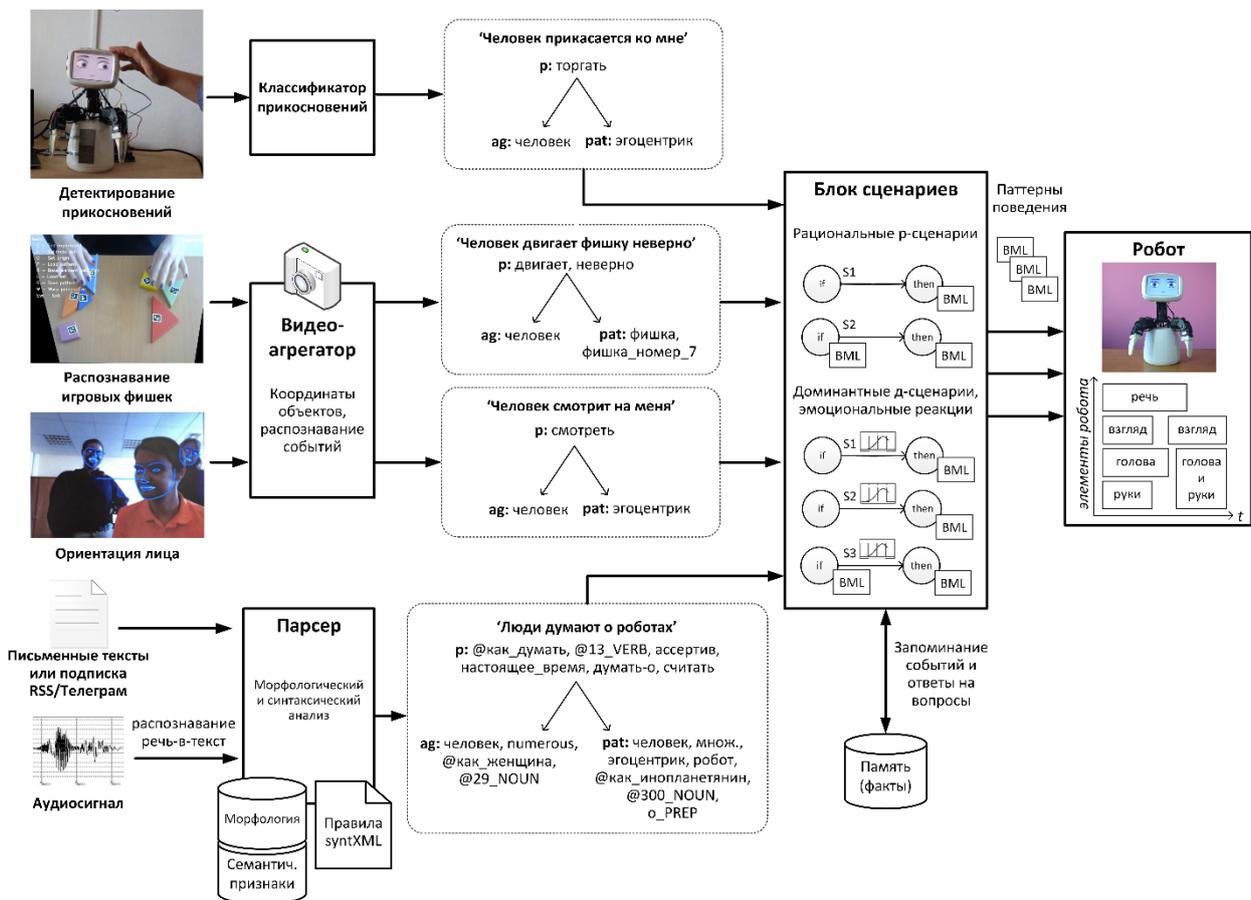


Рис. 1. Система управления роботом Ф-2

Следуя архитектуре CogAff, мы противопоставляем в блоке сценариев (а) базовые сценарии реагирования, аналоги эмоций – *доминантные сцена-*

рии или *d*-сценарии, а также (б) рациональные сценарии реагирования – *p*-сценарии. *D*-сценарии отвечают за распознавание в речи или событиях окружающего мира «негативных» для робота ситуаций: ‘мне что-то угрожает’, ‘против меня что-то замышляют’, ‘я никому не нужен’ и т. д. – или позитивных ситуаций: ‘на меня обращают внимание’, ‘кто-то воплощает мою мечту’ и т. д. Для 34 теоретических *d*-сценариев, описанных в работе [3], мы используем около 300 практических *d*-сценариев в базе данных робота. Рациональные сценарии, напротив, созданы для моделирования естественно-языкового вывода без эмоциональной оценки. Как указано в CogAff, уровень рационального рассуждения призван реконструировать возможные причины и следствия наблюдаемых событий, отвечать на вопросы типа «что будет, если *X*?» (*what-if reasoning*). *P*-сценарии распознают во входящих событиях типичные ситуации (например, ‘человек выдернул вилку из розетки’, ‘человек выдвинул ящик’, ‘исследователь выдвинул гипотезу’) и достраивают репрезентации производных ситуаций: возможные причины, следствия и интерпретации (если ‘человек выдернул вилку из розетки’, то возможно ‘человек куда-то уходит’, ранее ‘некоторый прибор работал’, потом ‘прибор перестал работать’ и т. д.). *P*-сценарии получены нами на основе кластеризации около 10 млн. семантических репрезентаций отдельных клауз из текстов на русском языке, разобранных парсером робота Ф-2. Из 8000 кластеров, где каждый кластер должен был содержать описания близких ситуаций, после ручной проверки было сформировано около 4500 *p*-сценариев. Связанные ситуации для сценариев (причины, следствия и интерпретации) были получены с помощью опроса информантов, которым предлагалось исходное высказывание сценария с просьбой написать его возможные причины, следствия и интерпретации (*что было до, что было после и что это значит?*).

В традиционных системах управления для выполнения обычно выбирается одно действие, получившее наивысшую оценку. Остальные варианты действий отбрасываются. В архитектуре робота Ф-2, напротив, активизация и конкуренция нескольких сценариев при обработке входящего события могут использоваться для моделирования сложных когнитивных эффектов: для имитации иронии [4], модели психического (*theory of mind*), моделирования нескольких точек зрения на ситуацию и для «воображения». На множестве *d*-сценариев этот процесс выглядит следующим образом. Если, к примеру, мы смотрим на робота или говорим ему *На тебя все обращают внимание*, он может «радоваться» (активизируется позитивный *d*-сценарий «внимание» *d-scr1* на рисунке 2) или «беспокоиться», что ‘против него что-то замышляют’ (активизируется негативный *d*-сценарий «планирование» – *d-scr3*). На множестве *d*-сценариев могут быть заданы противопоставления (признаки *a–b* в посылках сценариев *d-scr1* и *d-scr3*). Если при восприятии высказывания *На тебя все обращают внимание* робот активизирует позитивный *d*-сценарий «внимание» (*d-scr1*), то на основании противопоставления «хорошо-плохо» (признаки *a–b*) он подавит негативный *d*-сценарий «планирование» (*d-scr3*).

То есть в позитивной ситуации робот не будет подозревать злонамеренность. Вместе с тем, вытесненное понимание (злонамеренное «планирование») может использоваться роботом в другом когнитивном домене: в области иронии, воображения или модели психического.

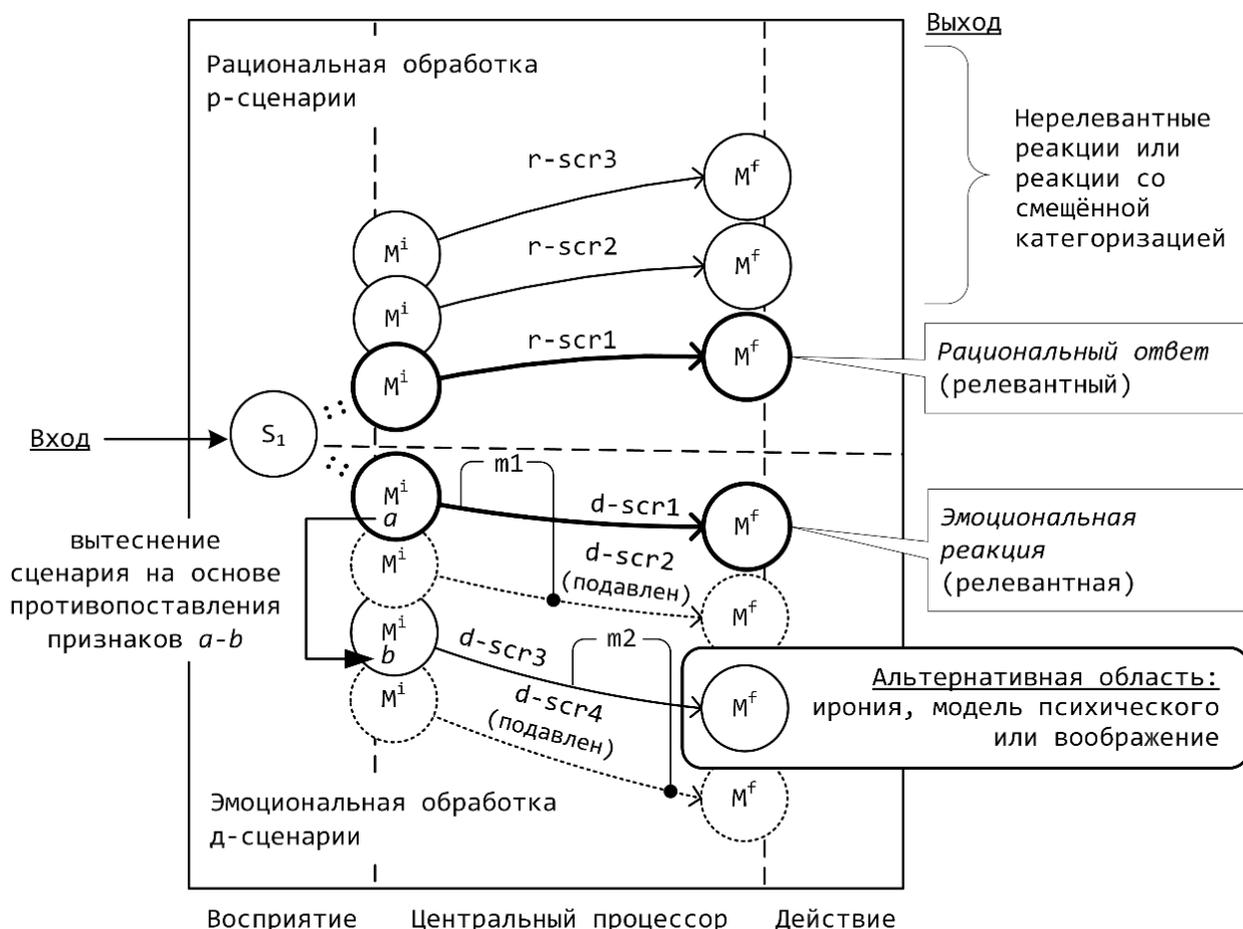


Рис. 2. Операции в блоке сценариев: активизация и подавление сценариев при обработке входящего стимула S_1

Негативный д-сценарий «планирование» может быть использован для иронии («на меня все обращают внимание, наверно они против меня что-то замышляют») или приписан другому лицу (в этой ситуации «другой человек может испугаться: что против него что-то замышляют»). Такое сложное поведение проявляется в диалоге в следующем виде:

Человек: Ты всех бесишь!

Робот: ¹Я расстраиваю людей. ²Люди думают: плохой человек нервнует других людей. ³Было бы здорово: я воодушевляю людей. ⁴Они издают назойливое жужжание, они раздражают вас, когда начинают ползать по вашему телу.

В этом примере робот последовательно: (1) отвечает релевантным высказыванием из негативного д-сценария «я-неадекватен», (2) использует высказывание из негативного д-сценария «он-эмоционален» в качестве взгляда

на себя другого человека (модель психического), этот д-сценарий был подавлен на основе противопоставления «я-другие», (3) использует высказывание из позитивного д-сценария «воодушевление» для материала «воображения» – для конструирования желаемой ситуации, этот сценарий был подавлен на основе противопоставления «плохо-хорошо», (4) вспоминает аналогичную ситуацию из прочитанной книги, это может помочь роботу использовать в диалоге «культурные аналогии». При активизации д-сценария, а также при последующем выборе ответного высказывания робот отождествляет себя с валентностью *экспериенцер* – тот, кто испытывает эмоции в данной ситуации (подчёркнутые слова в примерах). То есть в позитивных д-сценариях робот может отождествлять себя с различными героями текста – протагонистами, а в негативных д-сценариях робот может как противопоставлять себя антагонистам, так и «переживать», что он сам занимает валентность антагониста, как в примере (2) – *плохой человек*.

На множестве р-сценариев конкуренция выглядит следующим образом. Если высказывание *Мама мыла раму* оценивается по близости к р-сценариям, то самым релевантным р-сценарием *r-scr1* (см. рисунок 2) становится ‘инженер менял трубу’: ‘старая труба работала плохо’ – причина, но ‘новая будет работать лучше’ – следствие. Этот сценарий создаёт репрезентации ‘старая рама работала плохо’ и ‘новая рама будет работать лучше’. За неимением более релевантного сценария (‘*X* был грязным’ – ‘человек помыл *X*’ – ‘*X* стал чистым’) сценарий ‘инженер менял трубу’ предлагает удовлетворительное понимание исходной ситуации. Если же мы посмотрим на менее релевантные р-сценарии, то они подскажут репрезентации ‘мама нервничала, поэтому она мыла раму’ (сценарий *r-scr2* ‘человек мят перчатки’ включает причину ‘человек нервничал’), и ‘мама приготовит раму’/‘гости будут есть раму’ (сценарий *r-scr3* ‘хозяйка мыла картошку’ включает следствие ‘хозяйка приготовит картошку’/‘гости будут есть картошку’). Хотя сценарий *r-scr2* ‘человек мят перчатки’ представляется по крайней мере странным, а сценарий *r-scr3* ‘хозяйка мыла картошку’ – вообще ошибочным, нужно учитывать, что эти сценарии всё же проиграли более релевантному сценарию *r-scr1* ‘инженер менял трубу’, то есть для робота они не будут отражать «наиболее адекватную репрезентацию ситуации». Вместе с тем, данные проигравшие сценарии могут использоваться для иронии и воображения. Изменение категоризации (‘рама’ становится ‘едой’) открывает здесь возможности для автоматического синтеза ироничных высказываний и для поддержания эмоциональной коммуникации с человеком. Взаимодействие рациональных и доминантных сценариев также перспективно для имитации сложных «вторичных» эмоций, при которых естественно-языковой вывод (элемент уровня рассуждений в CogAff или р-сценарий) строит производный семантический компонент, который вызывает эмоциональную реакцию (элемент уровня реакций в CogAff или д-сценарий) [5]. Например, сообщение *хозяйка мыла картошку* будет оцени-

ваться как позитивное эмоциональное событие, потому что оно приводит к следствию ‘мы едим картошку’, активизирующему д-сценарий ВКУС.

Таким образом, множественная активизация сценариев может быть ключевым механизмом для обогащения коммуникативного поведения робота, а механизм вытеснения сценариев может стать основной для моделирования на работе ряда сложных когнитивных функций: иронии, воображения или модели психического.

Работа А. А. Котова в данном проекте поддержана грантом Российского научного фонда № 19-18-00547, <https://rscf.ru/project/19-18-00547/>

Список литературы

1. Kopp S., Krenn B., Marsella, S., et al. Towards a Common Framework for Multimodal Generation: The Behavior Markup Language // *Intelligent Virtual Agents*. – 2006. – pp. 205–217.
2. Sloman A. Varieties of Affect and the CogAff Architecture Schema // C. Johnson (Ed.), *Proceedings Symposium on Emotion, Cognition, and Affective Computing AISB’01 Convention*. 2001. – Vol. 10, Issues 4–5. – pp. 39–48.
3. Котов А. А. Механизмы речевого воздействия. – М.: РГГУ, 2021.
4. Kotov A. Accounting for irony and emotional oscillation in computer architectures // *Proceedings of International Conference on Affective Computing and Intelligent Interaction ACII 2009*. – IEEE, 2009. – pp. 506–511.
5. Allen S. R. Concern Processing in Autonomous Agents, Ph.D thesis. – 2001. – 213 p.