

Adaptatividade em Robôs Sociáveis: uma Proposta de um Gerenciador de Diálogos

D. A. Alfnas, M. R. Pereira-Barretto

Abstract— Within sociable robots, a Dialog Manager is responsible for conducting a conversation. This paper proposes architecture to convey natural conversation about things in general, lying on the use of Adaptivity combined with Episodic and Semantic memories to produce human-like behavior.

Keywords— Sociable Robots, Dialog Manager, adaptive technology, episodic memory, semantic memory.

I. INTRODUÇÃO

Robôs sociáveis são robôs autônomos que interagem e se comunicam com humanos e outros agentes físicos autônomos seguindo as regras de comportamento social e as regras associadas ao seu papel na sociedade. Parte essencial de sua construção, portanto, refere-se à sua capacidade de comunicação.

A comunicação entre humanos dá-se de forma verbal e não verbal. Assim, deve-se aliar a capacidade de expressão em linguagem natural à habilidade de compreensão e expressão não-verbal, com especial destaque às emoções transportadas na voz, na face e no gestual.

Avanços na compreensão e expressão verbal têm sido demonstrados; de fato, interfaces por comando de voz (SR, *speech recognition*) estão disponíveis até em telefones celulares e bibliotecas para conversão de texto-para-fala (TTS, *text-to-speech*) de boa qualidade são encontráveis. Entretanto, embora expressivos, tais bibliotecas permitem apenas a execução da entrada-saída: reconhecimento e fonação de palavras. Aspectos prosódicos, como frases interrogativas, ainda não são tratados de forma ampla pelas soluções disponíveis. Os aspectos gramaticais, ligados à compreensão do contexto, ainda são objeto de pesquisas.

No que se relaciona à compreensão e expressão de aspectos não-verbais, particularmente de emoções, os avanços tem ocorrido principalmente nos últimos anos. A detecção de emoções na face e na voz já encontra algoritmos com grau de acerto que se aproxima, em situações controladas, do obtido por seres humanos.

Este artigo propõe uma arquitetura de referência para a construção de um sistema de conversação para robôs sociáveis com a utilização de adaptatividade no gerenciamento de diálogo. A conversação será voltada a “coisas em geral”, como aquelas que uma pessoa gosta (ou desgosta), que possui (ou

não possui), que aconteceram recentemente: o que se chama, em forma coloquial, como “papo furado” ou “bater papo”. Robôs sociáveis com tal habilidade poderão ser utilizados, por exemplo, como acompanhantes para doentes, idosos ou pessoas solitárias em geral. Tal habilidade, entretanto, poderá estar disponível em robôs sociáveis em outros tipos de situação, já que conhecer o interlocutor, no que se relaciona ao que este faz, gosta ou possui, é uma das demonstrações mais fortes de sociabilidade.

As próximas seções deste artigo estarão organizadas da seguinte maneira: na seção 2, faz-se uma breve descrição de trabalhos mais diretamente correlacionados à presente proposta. Na seção 3, o conceito de adaptatividade é sucintamente apresentado. Já na seção 4, apresenta-se a arquitetura proposta. Na seção 5, detalha-se o Gerenciador de Diálogo (*Dialog Manager*), elemento da arquitetura em que a adaptatividade é diretamente empregada. Por fim, na seção 6 estão as considerações finais.

II. TRABALHOS CORRELACIONADOS

Foram considerados como trabalhos correlacionados aqueles que (i) abordam o tema de geração de diálogos em linguagem natural e (ii) baseiam-se em modelos com memória episódica e semântica para a geração do diálogo. Assim, trabalhos baseados em AIML (*Artificial Intelligence Markup Language*), como A.L.I.C.E [3], não foram incluídos, principalmente por não se utilizarem de memória na forma de memória episódica ou semântica, em que pese seu resultado impressionante.

Desde o trabalho pioneiro de Tulving [4], a função de memória nos seres humanos foi entendida na forma mostrada na Fig. 1 [5].

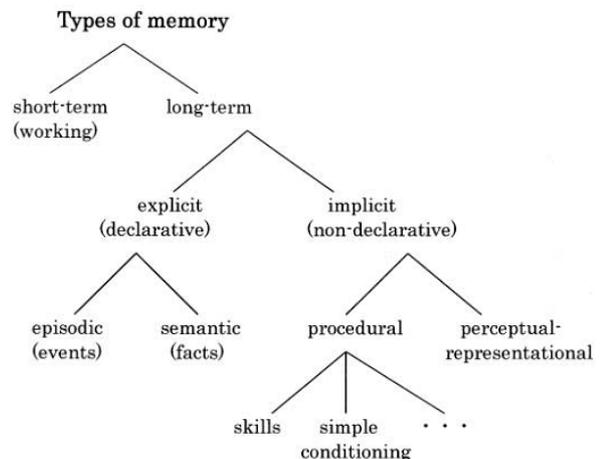


Figura 1. Tipos de memória [5]

M. R. Pereira-Barretto, Escola Politécnica da Universidade de São Paulo (EP-USP), marcos.barretto@poli.usp.br

D. A. Alfnas, Fundação para o Desenvolvimento Tecnológico da Engenharia (FDTE), d.alfnas@fdte.org.br

A classificação diferencia os seguintes tipos de memória:

- Memória de curto prazo (*short-term, working memory*), responsável pela memória nos acontecimentos das últimas horas;
- Memória explícita, que se divide em memória episódica, responsável pela memorização dos eventos (autobiográfica), e memória semântica, responsável pela memorização de fatos;
- Memória implícita, dividindo-se em memória procedural (responsável por procedimentos) e perceptual-representacional (relacionada a conceitos e abstrações).

Para os propósitos do presente trabalho, têm relevância apenas as memórias episódica e semântica.

Em [6], um agente conversacional (CA, *conversational agent*) foi proposto na forma da Fig. 2, voltado a aconselhar estudantes em débito com as mensalidades da Universidade.

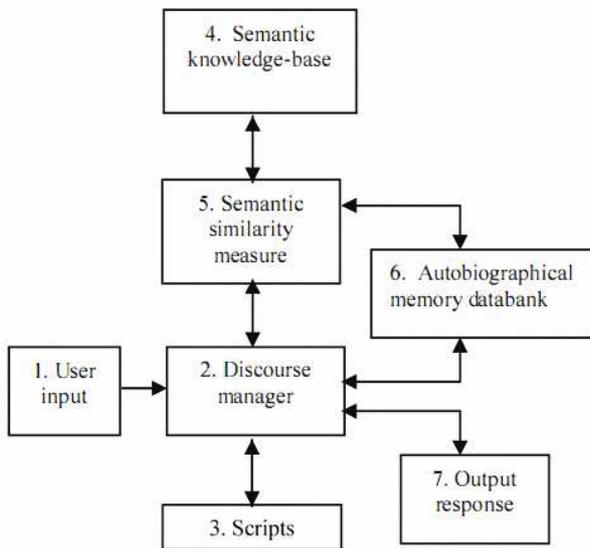


Figura 2. Arquitetura proposta em [6].

O Gerenciador de Diálogo (*Discourse Manager* na fig. 2) está claramente caracterizado, utilizando-se da memória episódica restrita à conversação corrente (*Autobiographical memory databank*) e utilizando-se de um processo de identificação de similaridade de conceitos para determinação do contexto da conversação (*semantic similarity measure*), com base na distância entre conceitos contidos em sua memória semântica (*semantic knowledge base*), que é formada pelos conceitos obtidos da WordNet [7]. O Gerenciador de Diálogo, neste caso, assemelha-se ao utilizado com AIML: identifica-se uma situação a partir do conceito mais similar encontrado e da estrutura da frase, gerando-se a saída a partir de um *template* de resposta.

Em [8] é proposta a arquitetura mostrada na Fig. 3, construída para auxiliar estudantes no aprendizado de redes de computadores. O Gerenciador de Diálogo (*Dialogue Manager* na figura 3) também está claramente identificado, bem como os módulos de memória. Neste trabalho, o Gerenciador de Diálogo é baseado em HTN (*Hierarchical Task Network*),

uma técnica comumente utilizada como *task planner* em robôs, que escolhe a situação de diálogo a aplicar. Em seguida, máquinas de estados finitos (FSM, *finite state machine*) são utilizadas para a condução do diálogo propriamente dito.

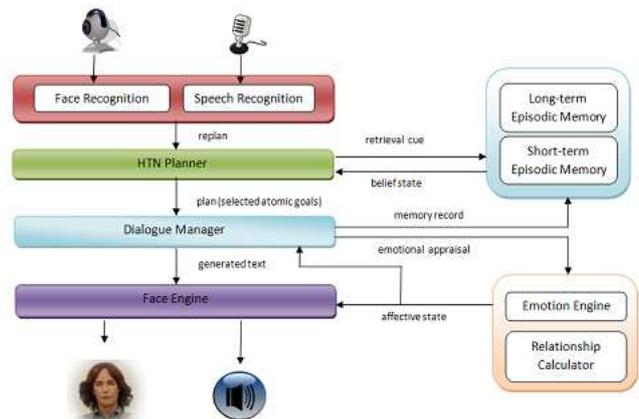


Figura 3. Arquitetura proposta em [8].

A utilização de HTN leva a uma estrutura relativamente fixa da conversação, mostrada na Fig. 4.

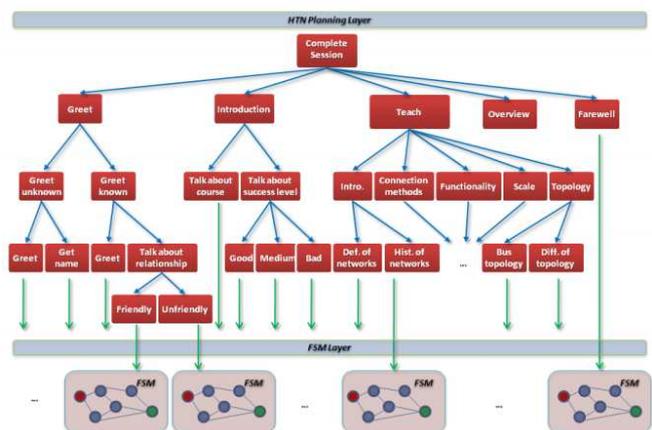


Figura 4. Estrutura da conversação em [8].

Uma estrutura também relativamente fixa, por conta do uso de FSM, é observada nos diálogos: por exemplo, uma pergunta é feita e, dependendo da resposta certa ou errada, passa-se à nova pergunta ou à uma explicação do assunto correlacionado.

Destaca-se que este trabalho leva em consideração, na condução do diálogo, o sentimento do robô em relação ao interlocutor (por exemplo, ressentimento por um grande número de respostas erradas) bem como o histórico de sua inter-relação (por exemplo, sessões anteriores em que o robô ressentiu-se do tratamento recebido), o que está representado na Fig. 3 (*emotion engine, relationship calculator*).

III. MÁQUINAS DE MARKOV ADAPTATIVAS

Neste trabalho, o conceito de adaptatividade está associado a Máquinas de Markov Adaptativas, utilizando-se a formulação em [9], baseada em [1], [2].

Uma máquina de Markov adaptativa M é definida por uma quintupla

$$M = (Q, S, T, q_0, F) \quad (1)$$

onde Q é um conjunto finito com n estados, S é o alfabeto de saída, q_0 é o estado inicial da rede, T é um conjunto de transições entre esses estados e F é um conjunto de funções adaptativas.

Toda função adaptativa pertencente a F pode ser definida como uma quádrupla

$$f = (\Psi, V, G, C) \quad (2)$$

onde:

- Ψ é um conjunto de parâmetros formais ($\psi_1, \psi_2, \psi_3, \dots, \psi_m$);
- V é um conjunto de identificadores de variáveis (v_1, v_2, \dots, v_n), cujos valores são desconhecidos no instante de chamada de f mas que uma vez preenchidos terão seus valores preservados durante toda a execução da função;
- G é um conjunto de identificadores de geradores ($g_1^*, g_2^*, \dots, g_n^*$), variáveis especiais que são preenchidas com novos valores, ainda não utilizados pelo autômato, a cada vez que a função é chamada;
- C é sequência de ações adaptativas elementares executadas em f .

Cada transição γ_{ij} é definida como uma quintupla

$$\gamma_{ij} = (q_i, q_j, \rho_{ij}, p_{ij}, a_{ij}). \quad (3)$$

onde ρ_{ij} é a probabilidade do estado q_j ser atingido estando em q_i , $p_{ij} \in S \cup \{\mathcal{E}\}$ é um símbolo inserido na cadeia de saída de M e a_{ij} é uma ação adaptativa associada a transição.

Observe que cada transição γ_{ij} é única em T , isto é, não há duas transições diferentes partindo de um mesmo estado q_i e chegando a q_j . Além disso, seja Γ_q como o conjunto de todas as transições que se originam em um estado q . Então

$$\forall q \in Q, \sum_{\gamma \in \Gamma_q} \rho_\gamma = 1 \quad (4)$$

Isto é, a soma das probabilidades de todas as transições iniciando em q é exatamente um, para todo q pertencente a Q .

A ação adaptativa a_{ij} tem a forma

$$a_{ij} = f(\omega_1, \omega_2, \dots, \omega_n) \cup \{\mathcal{E}\} \quad (5)$$

sendo f uma função adaptativa pertencente a F e $(\omega_1, \omega_2, \dots, \omega_n)$ uma lista de argumentos que correspondem posicionalmente à lista de parâmetros Ψ declaradas para f .

Define-se um SMA (Sistema de Markov Adaptativo) como um conjunto de máquinas adaptativas de Markov que podem se relacionar através de novas ações adaptativas, introduzidas a seguir. Assim, pode-se estabelecer uma relação de escopo sobre as ações adaptativas definidas para cada máquina, classificando-as em: ações que modifiquem apenas a topologia local da máquina e ações que podem interferir no comportamento de outras máquinas pertencentes ao sistema. Desta forma, cada máquina M^k pertencente a um sistema de Markov adaptativo é definida como uma quintupla:

$$M^k = (Q^k, \Sigma, T^k, q_0^k, F^k) \quad (10)$$

onde Σ é o alfabeto de saída, comum a todas as máquinas do sistema, e Q^k, T^k, q_0^k e F^k são como na definição de Q, T, q_0 e F , respectivamente.

IV. ARQUITETURA PROPOSTA

A arquitetura proposta está mostrada na Fig. 5. Os elementos arquiteturais representados têm as seguintes funções:

- ASR (*automatic speech recognition*): transformação da fala em texto. Uma solução existente, comercial ou não, será utilizada na implementação da arquitetura proposta.
- Detecção de silêncio: módulo auxiliar, voltado a auxiliar na identificação de frases completas bem como na atitude pro-ativa do robô em provocar uma conversa se não houver intervenção do interlocutor.
- Detecção de prosódia: determinar o tipo de frase (afirmativa, interrogativa, exclamativa) sendo pronunciada pelo interlocutor.
- Detecção de frase: transformação das saídas dos módulos anteriores em uma forma normalizada, contendo pontuação completa.
- POS (*part-of-speech*) *tagger*: detecção da classe de cada palavra, bem como sua função gramatical, sua flexão e forma primitiva.
- Resolvedor sintático: componente voltado a resolver as relações sintáticas (por exemplo, o pronome “que” em “meu amigo que é um aluno da Escola Politécnica ...”)
- Resolvedor de referência: componente voltado a resolver referências contextuais do diálogo, utilizando-se principalmente a memória episódica. Notar que, na arquitetura proposta, a memória episódica assume também as funções de memória de curto prazo.
- Construtor de Diálogo ou Gerenciador de Diálogo: responsável pela condução da conversação.
- Construtor de frases: responsável pela transformação da saída do Gerenciador de Diálogo em uma frase completa.
- Construtor de prosódia: responsável por adicionar elementos prosódicos à saída.
- TTS (*text-to-speech*): responsável pela geração da fala propriamente dita.
- Memória episódica: responsável pela memória episódica, como conceituada anteriormente
- Memória ontológica: trata-se da memória semântica, mas assim denominada porque, na proposta, este módulo será construído a partir de ontologias.

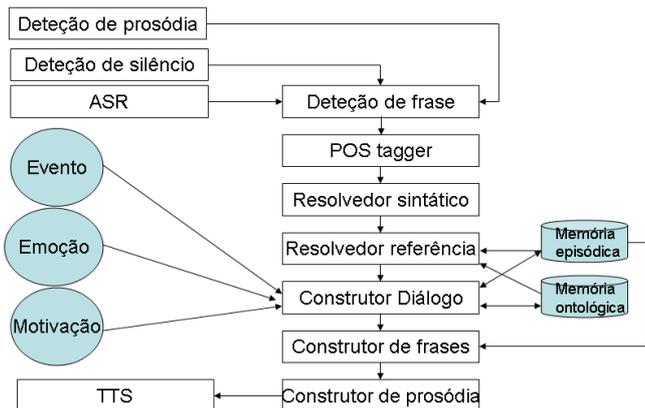


Figura 5. Arquitetura proposta.

A Fig. 5 ainda destaca os módulos:

- Emoção, responsável pela detecção da emoção demonstrada pelo interlocutor. Este módulo já conta com implementações descritas em [10], [11].
- Evento, responsável pela detecção de eventos, como a aproximação ou chegada de um interlocutor ou mesmo extra-conversaço, como um ruído muito forte;
- Motivação, responsável pela determinação da tarefa a ser realizada pelo robô. Uma implementação, similar à descrita em [8], está sendo realizada.

O foco deste trabalho é no módulo Gerenciador de Diálogo, cuja construção em curso baseia-se na adaptatividade, como apresentada anteriormente. De fato, propõe-se a utilização de um SMA (Sistema de Markov Adaptativo) em que uma máquina de nível superior, Condutor, será responsável por definir a condução do diálogo, seguindo idéias de estruturação como realizadas em [2]. Um possível estágio de evolução para esta máquina está mostrado na Fig. 6.

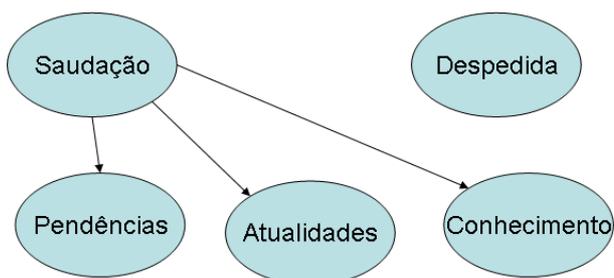


Figura 6. Gerenciador de Diálogo – nível Condutor.

Como ilustra a Fig. 6, a partir da Saudação inicial, pode-se dirigir o diálogo para Pendências (“Como você foi na prova de Cálculo?”), Atualidades (“Você soube da morte de Dennis Ritchie?”) ou Conhecimento (“Quantos irmãos você tem?”). Em função das regras de adaptação, a conversa eventualmente evoluirá até a Despedida.

Antes de caracterizar o nível Diálogo, que está subordinado ao nível Condutor, torna-se importante apresentar, de forma sucinta, a organização proposta para a memória ontológica. Sua estrutura está baseada nos blocos *vwCRK* (*Concept-*

relation-keyword) propostos em [12]. Um bloco *vwCRK* pode ser representado como mostra a Fig. 7.

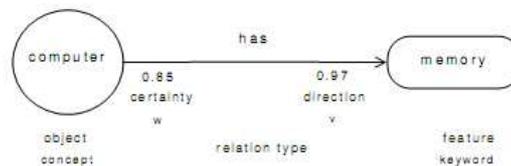


Figura 7. Bloco *vwCRK* [12].

Os parâmetros *w* e *v* representam, respectivamente, a certeza e a força da relação. Por exemplo, na frase “o cisne é branco”, tem-se *w* próximo de +1, enquanto em “o cisne é amarelo” tem-se *w* perto de -1. Entretanto, ambas as frases têm *v* próximo de 1, significando que tem-se certeza deste conhecimento. Entretanto, a memória ontológica não é simplesmente textual; na realidade, as memórias episódica e semântica são, na realidade, multimídia. A conceituação completa da memória ontológica pertencente à arquitetura não será aqui realizada; para os propósitos deste artigo, pode-se considerá-la como textual e semelhante à proposta em [7].

A memória semântica proposta está organizada nas seguintes ontologias:

- *Web of everything* (WoE), ontologia de referência, contendo todo o conhecimento do robô sociável sobre a estrutura do mundo;
- *Web of myself* (WoMe), conhecimento do robô sobre si mesmo
- *Web of others* (WoO), conhecimento do robô sobre cada conceito (pessoa, animal, lugar, livro, etc) com quem ele teve contato, seja virtual ou físico.

Assim, durante um diálogo, o robô utiliza o conhecimento que tem do mundo (WoE) para determinar se deve continuar a explorar aspectos de um mesmo conceito ou se deve alternar para outro conceito, anotando seu conhecimento na WoO específica do interlocutor e sua relação com o conceito em WoMe. Como exemplo, considere o conteúdo da WoE como mostrado na Fig. 8.

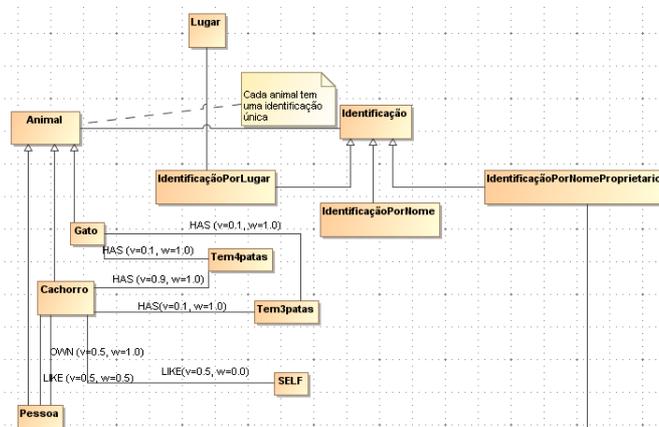


Figura 8. Exemplo de WoE

Supondo-se que em nível Diálogo esteja-se no estado Conhecimento. Uma possível realização de máquina para este estado, visando explorar a WoE, está mostrada na Fig. 8.

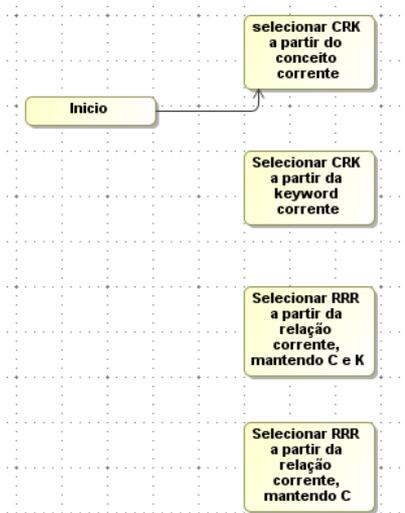


Figura 9. Um possível estágio de evolução da MMA.

Na Fig. 9, mostra-se o estágio inicial, que a única transição possível é a mostrada na figura. A partir desta transição, as funções de adaptação farão com que sejam explorados os caminhos possíveis na WoE. As funções de adaptação são condicionadas aos módulos constantes da Figura 5:

- O módulo Evento pode causar o fim da execução da submáquina corrente: por exemplo, a chegada de uma pessoa pode levar à interrupção da execução da submáquina Conhecimento, causando a adaptação de aumentar a probabilidade de acionar-se a máquina Despedida.
- O módulo Emoção pode detectar o desagrado do interlocutor, causando uma adaptação que eleve a probabilidade de alteração do conceito sendo explorado.
- O módulo Motivação pode causar uma ação de adaptação que leve ao fim da conversa, por cansaço.

Então, um possível diálogo que explore a WoE poderia ser aquele mostrado na Fig. 10, onde R representa o robô e J representa uma pessoa, chamada J.

R	1. Decisão MMA: Selecionar CRK (conceito-relação-keyword) a partir do conceito corrente (inicialmente, igual a Pessoa) 2. Selecionado: Pessoa-LIKE-Cachorro 3. Perguntar: Você gosta de cachorro?
J	Sim! (adicionar a WoO::J: J-LIKE(v=1.0,w=1.0)-Cachorro
R	1. Adicionar a WoO::J: J-LIKE(v=1.0,w=1.0)-Cachorro 2. Decisão MMA: Selecionar RRR (relação-relação-relação) a partir da relação corrente (LIKE) 3. Selecionado LIKE-OWN 4. Perguntar: Você possui um cachorro?

J	Sim!
R	1. Adicionar a WoE::J::J-OWN(v=1.0,w=1.0)-Cachorro 2. Erro: não pode criar identificador para Cachorro porque não conhece o nome do cachorro. 3. Perguntar: Qual o nome do cachorro?
J	Gaia.
R	1. Adicionar a WoE::J::J-OWN(v=1.0,w=1.0)-Cachorro 2. Decisão MMA:Selecionar CRK a partir do conceito corrente 3. Selecionado Pessoa-Like-Gato. 4. Perguntar: Você gosta de gato?

Figura 10. Um possível diálogo.

V. CONCLUSÃO

O artigo propôs uma arquitetura para a construção de diálogo humano-robô sociável, baseado em conceitos de adaptatividade. Assim como em composição musical, como demonstrado em [2], a adaptatividade apresenta-se como uma técnica viável para a construção de comportamentos que possam ser considerados como criativos (ou humanos), de forma aceitável. O modelo proposto está sendo implementado, para a comprovação desta tese.

REFERÊNCIAS

- [1] J. J. Neto, "Contribuições à metodologia de construção de compiladores", *Post-Doctoral Tese de Livre Docência*, EPUSP, São Paulo, 1993.
- [2] B. A. Basseto, "Um sistema de composição musical automatizada, baseado em gramáticas sensíveis ao contexto, implementado com formalismos adaptativos", *Dissertação de Mestrado*, EPUSP, São Paulo, 2000.
- [3] Alicebot Org. Disponível em www.alicebot.org. Acesso em 10 de outubro de 2011.
- [4] E. Tulving, "Episodic Memory: From Mind to Brain". *Annual Review of Psychology* 53, pp. 1-25, 2002.
- [5] Y. Miyashita, "Cognitive memory: cellular and network machineries and their top-down control". *Science Magazine*, vol.306, pp.435-440, 2004.
- [6] K. O'Shea, K. Crockett e Z. Bandar, "A proposed mechanism for memory simulation within a semantic-based conversational agent framework". *IEEE International Conference on Systems, Man and Cybernetics*, Istanbul, Turkey, 2010.
- [7] G. A. Miller, "WordNet: A Lexical Database for English", *Comm. ACM, Vol. 38, no. II*, 1995, pp. 39-41.
- [8] Z. Kasap e N. Magnenat-Thalmann, "Towards episodic memory-based long-term affective interaction with a human-like robot". *19th International Symposium on Robot and Human Interactive Communication*, Viareggio, Italy, 2010.
- [9] D. A. Alfenas, D. P. Shibata, D.P, J. J. Neto e M. R. Pereira-Barretto, "Sistemas de Markov Adaptativos: formulação e plataforma de desenvolvimento", *submetido ao WTA2012*.
- [10] D. R. Cueva, R. A. M. Gonçalves, F. G. Cozman e M. R. Pereira-Barretto, "Fusão de observações afetivas em cenários realistas". *Anais do ENIA (Encontro Nacional de Inteligência Artificial)*, Natal, RN, Brasil, 2011.
- [11] Gonçalves, R.A.M.; Cueva, D.R.; Pereira-Barretto, M.R.; Cozman, F.G. "Determinação da emoção demonstrada pelo interlocutor". *Anais do ENIA (Encontro Nacional de Inteligência Artificial)*, Natal, RN, Brasil, 2011.
- [12] J. Szymanski e W. Duch, "Knowledge Representation and Acquisition for Large-Scale Semantic Memory", *International Joint Conference on Neural Networks*, 2008.



Daniel Assis Alfenas é formado em Engenharia da Computação pela Escola Politécnica da Universidade de São Paulo (EPUSP) em 2004. Trabalhou, nos últimos dez anos, nas diversas áreas do desenvolvimento de sistemas, desde a definição da demanda até a implantação do sistema. Recentemente tem trabalhado como coordenador técnico no desenvolvimento de sistemas complexos que envolvem simulação, otimização e interfaces ricas. Atualmente, é mestrando no Laboratório de Técnicas Adaptativas da EPUSP e a área de pesquisa é a aplicação da tecnologia adaptativa no diálogo em linguagem natural entre usuários e sistemas.



Marcos Ribeiro Pereira-Barretto é graduado em Engenharia Elétrica (1983), mestre em Engenharia Elétrica (1988) e doutor em Engenharia Mecânica (1993) pela Escola Politécnica da Universidade de São Paulo. É professor da Escola Politécnica da USP desde 1986. Atua nas seguintes áreas de pesquisa: robôs sociáveis, computação afetiva e arquitetura de sistemas para aplicações críticas como Automação Industrial.