

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/315113128>

# Agentes Inteligentes Conversacionais: Conceitos Básicos e Desenvolvimento

Chapter · July 2016

CITATIONS

0

READS

449

2 authors:



[Flávia A. Barros](#)

Federal University of Pernambuco

59 PUBLICATIONS 621 CITATIONS

[SEE PROFILE](#)



[Patricia Azevedo Tedesco](#)

Federal University of Pernambuco

148 PUBLICATIONS 1,362 CITATIONS

[SEE PROFILE](#)

## Capítulo

# 4

## Agentes Inteligentes Conversacionais: Conceitos Básicos e Desenvolvimento

Flávia de Almeida Barros e Patrícia Azevedo Tedesco

Centro de Informática, Universidade Federal de Pernambuco  
Av. Jornalista Aníbal Fernandes, s/n - Cidade Universitária  
CEP 50.740-560 - Recife – PE - Brasil

{fab,pcart}@cin.ufpe.br

### *Resumo*

*Este curso tem por objetivo apresentar um panorama sobre Agentes Inteligentes Conversacionais, incluindo um histórico da área, conceitos básicos, desenvolvimento e aplicações. Tais agentes, também denominados de Chatterbots, são projetados para interpretar a pergunta/intervenção do usuário e oferecer respostas diretas, buscando manter a ilusão de que o interlocutor está conversando com outro ser humano. Atualmente, encontramos esses agentes nas mais diversas aplicações, tais como ensino a distância, jogos sérios, em SACs de lojas online, e mesmo em sistemas voltados a pessoas com deficiência física. Daremos maior ênfase aos Agentes Conversacionais Incorporados, que possuem representação gráfica, movimentos faciais e corporais, e possibilitam diálogos via voz, o que torna a interação mais natural e realista.*

### *Abstract*

*This course aims to present an overview of Intelligent Conversational Agents, including a historical background, basic concepts, development and applications of such agents. These agents, also named as Chatterbots, are designed to interpret the user's question/intervention and to offer direct answers, with the aim of keeping the illusion that the user is talking to another human being. Currently, we find these in a wide variety of applications, such as distance learning, serious games, in online shops, and as personal assistants. Here we focus on Embodied Conversational Agents, which bear a graphical representation, facial and body movements, and are capable of performing voice dialogues, thus enabling a more natural and realistic interaction.*

## 4.1. Introdução

A ideia de construir sistemas capazes de dialogar com o usuário em linguagem natural surgiu na década de 1950, a partir do trabalho de Alan Turing intitulado “*Computing Machinery and Intelligence*” [Turing, 1950], no qual ele discutia a questão: “As máquinas podem pensar?”. Nesse artigo, Turing propôs o “Jogo da Imitação”, atualmente conhecido como o Teste de Turing<sup>1</sup>. Um sistema seria considerado inteligente se ele conseguisse se fazer passar por humano em um diálogo com um interlocutor humano.

O trabalho de Turing abriu a possibilidade de se desenvolverem “máquinas de diálogo”, mais tarde conhecidas como *Chatterbots*<sup>2</sup> (ou *Chatbots*), para as mais diversas aplicações [Laven, 2016]. Tais sistemas, denominados também de *Agentes Conversacionais* (ACs), são projetados para interpretar a pergunta (ou intervenção) do usuário, e oferecer respostas diretas, sempre buscando manter a coerência do diálogo e a ilusão de que o interlocutor está conversando com outro ser humano.

Encontramos atualmente diversas competições mundiais que realizam versões do Teste de Turing. Destacamos aqui o *Loebner Prize*<sup>3</sup>, que realiza eventos anuais de premiação, e o *Turing Test 2014*, um evento organizado pela Universidade de Reading (Grã-Bretanha)<sup>4</sup> que premiou o chatterbot Eugene<sup>5</sup> como sendo o primeiro sistema a passar no Teste de Turing irrestrito. Vale mencionar que o trabalho de Turing também vem inspirando livros e filmes desde a década de 1960, com destaque para o filme autobiográfico “*The Imitation Game*”<sup>6</sup> (2014), e diversos filmes do gênero ficção científica, como *Her*<sup>7</sup> (2013), incluindo enredos onde o Teste de Turing é aplicado a “humanóides” – por exemplo, *Ex Machina*<sup>8</sup> (2015), e o insuperável *Blade Runner* (1982)<sup>9</sup>.

As primeiras implementações desses sistemas eram puramente baseadas em texto. Na última década, contudo, os chatterbots *stand-alone* textuais perderam espaço para os chamados *Agentes Conversacionais Incorporados* (ACIs), agentes conversacionais que possuem uma representação gráfica e síntese/interpretação da fala (por exemplo, REA [Cassell et al., 1999] e GRETA<sup>10</sup>). Essa mudança se deve à crescente tendência de se desenvolverem ambientes virtuais imersivos, baseados em realidade virtual. Nesses casos, o usuário espera que o agente conversacional faça parte do ambiente simulado, possibilitando diálogos via voz, a fim de tornar a interação mais natural e realista.

---

<sup>1</sup> [https://en.wikipedia.org/wiki/Turing\\_test](https://en.wikipedia.org/wiki/Turing_test)

<sup>2</sup> O termo “*chatterbot*” foi introduzido por Michael Mauldin em 1994.

Ver <http://en.wikipedia.org/wiki/Chatterbot>

<sup>3</sup> <http://www.loebner.net/Prizetf/loebner-prize.html>

<sup>4</sup> <http://www.reading.ac.uk/news-and-events/releases/PR583836.aspx>

<sup>5</sup> <http://www.reading.ac.uk/sse/news/sse-newsarticle-2014-06-08.aspx>

[https://en.wikipedia.org/wiki/Eugene\\_Goostman](https://en.wikipedia.org/wiki/Eugene_Goostman)

<sup>6</sup> <http://theimitationgamemovie.com/>

<sup>7</sup> <http://www.herthemovie.com/#/home>

<sup>8</sup> <http://exmachina-movie.com/synopsis>

<sup>9</sup> <http://www.warnerbros.com/blade-runner>

<sup>10</sup> <http://perso.telecom-paristech.fr/~pelachau/Greta/>

Esse cenário motivou o crescimento das pesquisas na área, a fim de prover ACIs mais credíveis. Essas pesquisas envolvem não apenas a área de Processamento de Linguagem Natural (em particular, Análise da Conversão), como também Processamento de voz (síntese e análise), Processamento de imagem, e outras áreas correlatas. No que concerne a modelagem e implementação desses agentes, nota-se uma tendência de uso de representações em XML. O objetivo é atender a requisitos da Engenharia de Software, como modularidade e reúso.

Atualmente, encontramos agentes conversacionais (incorporados ou não) nas mais diversas aplicações, tais como em sistemas de ensino a distância, jogos sérios, em SACs de lojas online, em sistemas voltados a pessoas com alguma deficiência física, entre outros [Chatbots.org, 2016].

Este curso tem por objetivo apresentar a área de *Agentes Inteligentes Conversacionais*, incluindo conceitos básicos, desenvolvimento e aplicações. Em particular, daremos maior ênfase aos Agentes Conversacionais Incorporados (ACIs), que são sua forma mais atual de apresentação.

Este documento prossegue com uma visão geral sobre os *Sistemas Conversacionais*, que incluem qualquer tipo de sistema que seja capaz de manter um diálogo em linguagem natural com o usuário (Seção 4.2.1). A Seção 4.2.2 define os *Agentes Conversacionais*, buscando caracterizá-los como *agentes inteligentes*, dentro da visão mais moderna da Inteligência Artificial [Russel e Norvig, 2010].

A Seção 4.3 apresenta brevemente conceitos básicos sobre duas áreas intimamente ligadas ao desenvolvimento de agentes conversacionais: *Teorias Linguística* para Análise da Conversação e Análise do Discurso (Seção 4.3.1), e *Computação Afetiva* - área de estudos que serve de base para se projetar agentes artificiais com traços de personalidade e emoção (Seção 4.3.2).

A Seção 4.4 trata dos *Agentes Conversacionais textuais* (ou *Chatterbots*), precursores dos Agentes Conversacionais Incorporados. A Seção 4.5 apresenta os ACIs, agentes conversacionais possuem uma representação gráfica e síntese/interpretação da fala. Veremos um breve histórico da área, seguido de técnicas utilizadas na construção desses agentes. Discutiremos algumas aplicações existentes na área, concluindo com uma visão das tendências futuras da área.

A Seção 4.6 conclui o documento trazendo algumas considerações finais, apresentando desafios e tendências futuras de pesquisa e desenvolvimento na área. Por fim, temos a lista de referências bibliográficas citadas no documento.

Por fim, gostaríamos de ressaltar que todos os sítios na Web (URLs) indicados aqui estavam ativos durante o período de escrita deste documento (entre abril e maio de 2016).

## 4.2. Sistemas e Agentes Conversacionais

Como já mencionado, os sistemas conversacionais são projetados para interpretar a pergunta (ou intervenção) do usuário, e oferecer respostas diretas, buscando manter a coerência do diálogo e a ilusão de que o interlocutor está conversando com outro ser humano. Devido a essas características, Chatterbots podem ser considerados como *agentes inteligentes*, dentro da visão mais moderna da Inteligência Artificial (IA) [Russel e Norvig, 2010]. Tais agentes têm como característica marcante a capacidade de dialogar com os usuários, sendo então chamados de *Agentes Conversacionais*.

Considerando sua grande utilização em sistemas Web (de livre acesso ou não), os agentes conversacionais podem ser considerados como uma forma de interface, que complementa ou substitui outras formas de acesso ao computador. Atualmente, encontramos tais agentes em sistemas de ensino/educação a distância (EaD), em jogos digitais sérios ou para simples entretenimento, em sistemas de FAQ (*Frequent Asked Questions*), em SACs (Sistemas de Atendimento ao Cliente) de lojas online, entre outros (ver [ALICE Bot, 2016], [Cassell et al., 1999], [Chatbots.org, 2016], [Laven, 2016]).

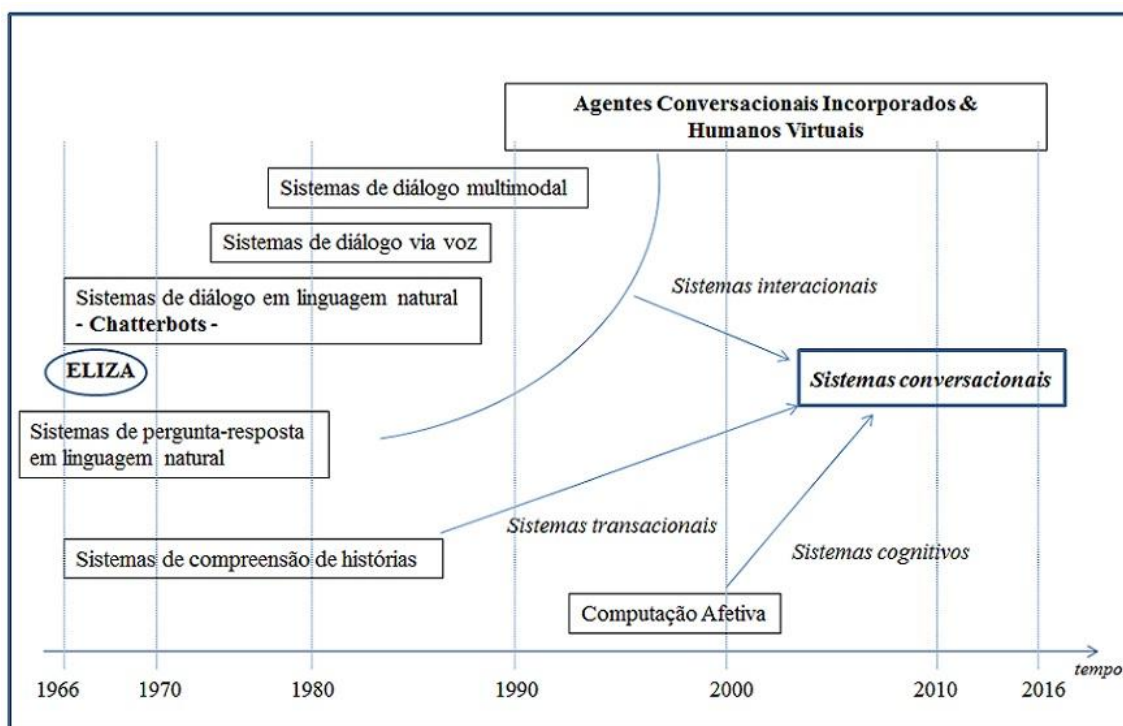
Antes de avançar na apresentação sobre os ACs, vamos situar esses agentes dentro da história do desenvolvimento dos Sistemas Conversacionais. A seguir, a Seção 4.2.2 busca caracterizar os *agentes conversacionais* como *agentes inteligentes*, trazendo definições e estabelecendo a terminologia adotada neste documento. Como veremos, os Agentes Conversacionais são um tipo de sistema conversacional que apresenta características de agente virtual autônomo.

### 4.2.1 Evolução dos Sistemas Conversacionais

Os primeiros sistemas que buscavam interagir com o usuário em linguagem natural, conhecidos como *Sistemas de Pergunta-Resposta*, datam do início da década de 1960. Exemplos de tais sistemas são BASEBALL [Green et al., 1961], LUNAR [Woods, 1973], e SHRDLU (o mais conhecido e bem sucedido dentre os sistemas iniciais) [Winograd, 1972]. Contudo, tais sistemas não eram capazes de travar um diálogo mais flexível com o usuário, limitando-se a responder perguntas previamente catalogadas.

O primeiro sistema com ampla divulgação capaz de travar diálogos (simples) em linguagem natural foi ELIZA [Weizenbaum, 1966] [Eliza Chat Bot, 2016]. Mais tarde, tais sistemas receberam a denominação de *Chatterbots* (robôs de conversação). Esses sistemas marcaram a década de 1970, ainda compartilhando espaço com os sistemas pergunta-resposta.

Em paralelo aos chatterbots, temos ainda os *Sistemas para compreensão e geração de histórias*, que eram baseados em *scripts* [Schank et al., 1975]. Esses não eram sistemas de diálogo, porém ajudaram a avançar as pesquisas no Processamento de Linguagem Natural (PLN) [Allen, 1995], contribuindo assim para a melhoria dos sistemas conversacionais.



**Figura 4.1: História do desenvolvimento dos Sistemas Conversacionais – figura adaptada de [Nishida et al., 2014].**

Os sistemas conversacionais via voz, uma evolução dos chatterbots textuais, ganharam espaço na década de 1980, acompanhando a evolução na área de processamento de voz (interpretação e síntese). Exemplos de tais sistemas são: *Hearsay-II* [Erman et al., 1980] e “*Put-That-There*” [Bolt, 1980].

No final da década de 1980, surgiu o sistema “*The Knowledge Navigator*”, o primeiro sistema conversacional via voz com uma imagem humana associada (conhecido como *Phil*)<sup>11</sup>. Phil já pode ser considerado como um agente inteligente virtual (ver Seção 4.2.2). A partir daí, a evolução na área foi bem mais rápida, com os sistemas de diálogo multimodal, chegando até os Agentes Conversacionais Incorporados (e.g.<sup>12</sup>, *REA* [Cassell et al., 1999]).

Vale ressaltar, contudo, que os chatterbots puramente textuais ainda são tema de pesquisa e desenvolvimento. Nesse caso, o foco das pesquisas é mais voltado à capacidade de manter diálogos credíveis com os interlocutores (Seção 4.4).

Por fim, destacamos os *Humanos Virtuais* (ou Atores Sintéticos), que também podem ser considerados como sistemas conversacionais, porém com foco não apenas na capacidade dialógica, mas principalmente na ilusão de vida (e.g., Projeto Oz [Mateas, 1999]). A fim de passarem mais credibilidade, os humanos virtuais também possuem personalidade e capacidade de demonstrar emoção, seguindo as pesquisas na área de

<sup>11</sup> [http://en.wikipedia.org/wiki/Knowledge\\_Navigator](http://en.wikipedia.org/wiki/Knowledge_Navigator); [https://www.youtube.com/watch?v=\\_a0t2Eb7YJk](https://www.youtube.com/watch?v=_a0t2Eb7YJk)

<sup>12</sup> e.g.: Abreviatura do termo em Latim “*exempli gratia*”, que, em Português, significa “por exemplo”.

Computação Afetiva [Picard, 2003] (Seção 4.3.2). Tais capacidades podem também ser embutidas nos ACIs, na tentativa de prover diálogos mais naturais e verossímeis.

#### 4.2.2 Agentes Conversacionais como Agentes Inteligentes

Agentes conversacionais têm sido tradicionalmente desenvolvidos usando-se técnicas de Inteligência Artificial [Russel e Norvig, 2010], principalmente as técnicas oriundas da subárea de Processamento de Linguagem Natural [Allen, 1995]. Na década passada, esses agentes/sistemas também passaram a observar requisitos da Engenharia de Software, como modularidade e reuso [Wallace, 2009].

Esta seção busca caracterizar os agentes conversacionais como agentes inteligentes dentro do paradigma mais atual da IA. É importante fazer uma distinção entre agentes conversacionais puramente textuais (aqui denominados de chatterbots – Seção 4.4), e os ACs incorporados – agentes animados que possuem uma representação gráfica, sendo capazes de realizar o diálogo por meio de fala e de outras modalidades não verbais, como olhares e gestos (Seção 4.5). Vale ainda ressaltar que alguns chatterbots textuais apresentam uma imagem associada, porém não são considerados ACs incorporados, por não apresentarem as outras características dos ACIs mencionadas aqui.

De início, é importante definir quais as características indispensáveis a um software para qualificá-lo como um agente. Segundo [Russel e Norvig, 2010], um *agente* é qualquer entidade humana ou artificial (de software ou de hardware) que, estando imersa (ou situada) em um ambiente físico, virtual ou simulado, é capaz de:

- *perceber* seu ambiente através de sensores (*e.g.*, teclado, câmeras, microfone, protocolos de comunicação em sistemas multiagentes), e
- *agir* sobre esse ambiente através de atuadores (*e.g.*, tela do computador, autofalante, impressora, funções internas que enviam mensagens a outros agentes, braços, pernas...).

*Agentes inteligentes* (ou *racionais*, como denominados por [Russel e Norvig, 2010]) são aqueles capazes de *raciocinar* com base em suas percepções e seu conhecimento do mundo, a fim de escolher as ações mais adequadas a realizar no ambiente para alcançar seus *objetivos próprios*. Essa definição parece se adequar tanto a agentes humanos quanto aos artificiais.

A partir de agora, discutiremos apenas os *agentes artificiais*, que são o foco deste curso. Assim, quando não houver menção ao tipo de agente (se humano ou artificial), deve-se assumir que estamos nos referindo a agentes artificiais. Vejamos então mais algumas definições sobre os agentes inteligentes.

Agentes inteligentes são ditos *autônomos*, pois não são diretamente controlados por humanos. Podemos citar ainda outras características (opcionais) dos agentes inteligentes, como *adaptabilidade*, e ainda *personalidade* e *emoção*, características estas que buscam aumentar o grau de realismo desses agentes.

Cabe agora examinar os tipos de agentes artificiais que podemos encontrar em ambientes virtuais ou simulados. Apesar de essa ser uma área de pesquisa já bem explorada, ainda não existe uma taxonomia clara e consensual sobre o assunto. Assim

sendo, definiremos os tipos de agentes de acordo com o nosso conhecimento da área. Indicamos algumas referências importantes nas notas de rodapé.

- **Agentes inteligentes**<sup>13</sup> – Já definidos anteriormente, podendo ser humanos ou artificiais.
- **Agentes virtuais** – Todos os agentes inteligentes que não são humanos.
- **Agentes conversacionais**<sup>14</sup> – Agentes virtuais com capacidade de dialogar com o usuário através de texto ou usando processamento de voz.
- **Agentes credíveis**<sup>15</sup> – Agentes virtuais com personalidade e emoção.
- **Agentes incorporados**<sup>16</sup> – Agentes virtuais que têm uma representação gráfica em 2 ou 3 dimensões dentro de um ambiente virtual/simulado.
- **Agentes avatares**<sup>17</sup> – Inicialmente, esse termo era usado apenas para se referir a agentes que representavam humanos em ambientes virtuais/simulados. Atualmente, o termo também é usado para se referir a agentes artificiais em mundos virtuais persistentes. O site [Chatbots.org, 2016] traz uma discussão muito pertinente sobre o termo e seus variados usos.
- **Agentes animados** (mais conhecidos como *Atores Sintéticos*<sup>18</sup>) – são agentes credíveis incorporados, que possuem características especiais como atitudes, emoções e personalidade. A representação gráfica desses agentes é, em geral, gerada pelo computador e renderizada em tempo real, apresentando movimentos corporais, gestos e olhares.

Por fim, vejamos alguns tipos de ACs que combinam várias características dos tipos agentes citados acima.

- **Agentes conversacionais credíveis** – ACs com personalidade e emoção.
- **Agentes conversacionais incorporados** – ACs com representação física, e geralmente com comunicação via voz.
- **ACIs credíveis** (ou **Humanos Virtuais**) – ACIs animados/atores sintéticos.

Vamos adotar essa nomenclatura ao longo deste documento, reconhecendo sempre que nem sempre há consenso com relação a essas definições.

---

<sup>13</sup> *Intelligent Agents*. [http://en.wikipedia.org/wiki/Intelligent\\_agent](http://en.wikipedia.org/wiki/Intelligent_agent)

<sup>14</sup> *Conversational agents*. Ver [https://www.chatbots.org/conversational\\_agent/](https://www.chatbots.org/conversational_agent/); <http://digitalmarketing-glossary.com/What-is-Conversational-agent-definition>

<sup>15</sup> *Believable agents*. Ver [Bates, 1994], referência clássica.

<sup>16</sup> *Embodied agents*. Ver [http://en.wikipedia.org/wiki/Embodied\\_agent](http://en.wikipedia.org/wiki/Embodied_agent)

<sup>17</sup> <https://www.chatbots.org/avatar/>

<sup>18</sup> *Synthetic actors*. Ver <http://www.rodny.net/kim/SyntheticActors.htm>



### 4.3. Assuntos Correlatos – *Background*

Esta seção introduz conceitos básicos indispensáveis ao bom entendimento do conteúdo deste curso. Inicialmente, a Seção 4.3.1 discute alguns conceitos linguísticos relacionados ao desenvolvimento de sistemas e agentes conversacionais. A Seção 4.3.2 trata da Computação Afetiva, uma área de estudos especialmente importante na construção de agentes artificiais (em particular, de ACIs) com traços de personalidade e emoção, buscando dar mais credibilidade a esses agentes.

#### 4.3.1 Um pouco de Linguística – Análise da Conversação e Análise do Discurso

Veremos aqui conceitos básicos de duas áreas de estudos linguísticos intimamente ligadas aos Sistemas/Agentes conversacionais: Análise da Conversação e a Análise do Discurso.

##### Análise da Conversação

A *Análise da Conversação* estuda a *interação social*, cobrindo comunicações verbais e não verbais em situações do dia a dia. Essa área de estudo tem grande interesse para o presente curso, pois seu arcabouço teórico é frequentemente usado como base para projetar e analisar agentes conversacionais. Ao longo do texto, veremos como generalizar os conceitos aqui apresentados, definidos para conversações entre interlocutores humanos, para cobrir também as interações com agentes conversacionais.

Os estudos iniciais na área da Análise da Conversação cobriam apenas conversações sociais, que podem ser definidas como a “troca informal de ideias” entre duas ou mais pessoas (em contraste com discursos, aulas, debates, etc.)<sup>19</sup>. Contudo, seus métodos foram mais tarde adotados para estudar também interações institucionais/formais (e.g., julgamentos em tribunais, consultas médicas), e conversações orientadas a tarefas (como ocorre nos ambientes educacionais).

Apresentamos aqui um resumo breve sobre a Análise da Conversação, a fim de facilitar o entendimento do presente curso. Segundo [Marcuschi, 1986], uma conversação se distingue de outros processos comunicativos (como discursos e aulas) por apresentar algumas características básicas. Dentre elas, destacam-se:

- 1) Interação entre pelos menos dois interlocutores;
- 2) Ocorrência de pelo menos uma contribuição de cada interlocutor;
- 3) Sincronismo – em contraste com processos comunicativos assíncronos, como diálogos via WhatsApp<sup>20</sup>, por exemplo; e
- 4) Existência de um tema comum aos falantes.

Assim definida, a conversação engloba interações por telefone, via Skype<sup>21</sup>, e em aplicações online variadas, através das quais os usuários podem conversar com outros usuários ou com agentes conversacionais virtuais. Exemplos dessas aplicações são salas de *chat*, sistemas de EaD, SACs, jogos digitais sérios ou para simples

---

<sup>19</sup> <http://www.oldandsold.com/articles10/voice-17.shtml>

<sup>20</sup> <https://www.whatsapp.com/>

<sup>21</sup> <http://www.skype.com/>

entretenimento, e ainda os mundos virtuais<sup>22</sup> na Web, geralmente desenvolvidos para jogos (games).

Apesar de parecer livre, a conversação é uma atividade estruturada, que apresenta dois níveis distintos de organização, o nível global e o nível local.

No que concerne a *organização global*, é possível identificar três fases distintas de uma conversação: abertura, desenvolvimento e fechamento. Na *abertura*, os interlocutores procuram identificar os demais participantes da conversa (ou se apresentar ao grupo, se necessário), cumprimentando-se. No *desenvolvimento*, os interlocutores conversam sobre os tópicos escolhidos, podendo ocorrer várias mudanças de tópico. No *fechamento*, em geral os usuários se despedem. Os agentes conversacionais devem ser projetados para obedecer a essa estrutura, a fim de manter conversações naturais e verossímeis com os usuários.

A *organização local* da conversação, por sua vez, tem por base elementos organizadores, dentre os quais destacam-se o *turno*, as *sequências*, a *preferência* e os *tópicos* da conversação, brevemente descritos a seguir.

*Turno* é definido como a produção de um falante enquanto está com a palavra, incluindo aí a possibilidade do silêncio, quando ele for significativo [Marcuschi, 1986]. Durante uma conversação, cada interlocutor deve ficar restrito a falar nos turnos a ele alocados. A *tomada de turno* é definida como o processo pelo qual os interlocutores alocam turnos de fala para si. Podemos identificar diferentes estratégias para tomada de turno, por exemplo, começar a falar assim que o outro falante fizer uma breve pausa ou completar uma oração. Claramente, a tomada de turno em conversações com mais de dois interlocutores é bem mais complexa, pois requer também determinar quem será o próximo a falar.

No caso dos agentes conversacionais textuais, em geral não há dificuldades para se sinalizar a troca de turno, pois isso fica claro quando o usuário conclui a digitação do seu texto. Porém, quando se trata de ACIs com processamento de voz, a situação é semelhante a uma conversação entre humanos. Como esperado, o problema se agrava quando lidamos com ACIs em ambientes virtuais multiusuário.

*Sequências*, mais conhecidas como *pares adjacentes*, estruturam as conversações de forma padronizada [Schegloff, 1972]. Cada par adjacente define uma sequência de dois turnos que coocorrem, organizando assim a conversação no nível local. Por exemplo, quando um falante cumprimenta o outro (*e.g.*, “Bom dia!”), o segundo falante quase obrigatoriamente responde ao cumprimento. Como exemplos de pares adjacentes, citamos: {pergunta, resposta}, {cumprimento, cumprimento/silêncio}, {convite, aceitação/recusa}, {ordem, execução}.

A noção de *preferência* se reflete na escolha do segundo turno de cada par adjacente, sendo “preferidas” as escolhas mais comumente observadas nas interações sociais. Tomando como exemplo o par {convite, aceitação/recusa}, quando um interlocutor convida o outro a realizar alguma atividade, ele espera que seu convite seja aceito, sendo esta a opção socialmente *preferida* na formação desse par adjacente. Contudo, o outro interlocutor poderá também recusar o convite.

---

<sup>22</sup> [http://en.wikipedia.org/wiki/Virtual\\_world](http://en.wikipedia.org/wiki/Virtual_world), [http://pt.wikipedia.org/wiki/Mundo\\_virtual](http://pt.wikipedia.org/wiki/Mundo_virtual)

Os agentes conversacionais projetados para trabalhar com pares adjacentes podem ter suas preferências determinadas a priori. Caso o agente tenha personalidade e seja capaz de demonstrar emoção, as preferências na formação dos pares adjacentes devem estar em consonâncias com seus estados emocionais e de humor (Seção 4.3.2).

É possível extrapolar a ideia dos pares adjacentes para *comunicação gestual* (não apenas verbal). Um sorriso ou um aceno podem indicar a intenção do interlocutor em cumprimentar o outro, mesmo sem pronunciar/escrever nenhuma palavra. Assim, podemos pensar em pares adjacentes “multimodais”<sup>23</sup>, não apenas falados ou escritos. Isso é mais um recurso importante para os ACIs.

Os *tópicos* (temas) da conversação devem ser escolhidos de comum acordo entre os interlocutores, havendo uma mudança natural de tópico durante o desenrolar da interação. Note que, em conversações sociais, a *mudança de tópico* é, em geral, negociada entre os falantes a cada turno; porém deve ocorrer de forma natural, para não quebrar a fluência da conversa. Assim, podemos fazer uma distinção entre *mudança* e *quebra de tópico*, sendo esta última identificada quando ocorre uma interrupção brusca ou inesperada na conversação.

No que concerne os agentes conversacionais, identificamos ACs projetados para conversar sobre tópicos predeterminados (e.g., agentes tutores), e outros que são projetados para conversações livres (Seções 4.4 e 4.5). Em qualquer dos casos, os ACs devem ser capazes de identificar automaticamente mudanças de tópico iniciadas pelo usuário, e responder adequadamente a isso sem comprometer a fluência do diálogo.

Além desses elementos organizadores, vale ainda citar os *reparos*, definidos como mecanismos através dos quais os falantes lidam com problemas/mal entendidos ocorridos durante a conversação. Nota-se que o falante que introduziu a fonte do problema é quem busca reparar/corrigir o erro (essa é a preferência). Assim, em geral, esse falante inicia seu próximo turno com um reparo (e.g., uma correção ou uma observação), ou tenta utilizar algum silêncio do outro falante para antecipar o reparo.

Esse é um recurso muito importante nos ACs, uma vez que podem ocorrer muitos mal entendidos em conversações com esses agentes. As razões são diversas: erros do sistema ao mapear som (voz) em texto; dificuldades do sistema em interpretar o texto (devido a imperfeições nas frases do usuário, ou à falta de conhecimento do sistema), entre outras.

## **Análise do Discurso**

A *Análise do Discurso* (AD) estuda “os princípios que governam a produção de sequências estruturadas de frases (*discurso* - escrito ou falado)” [Barros e Robin 2001]. Incluímos aqui interações faladas e/ou escritas, síncronas ou assíncronas, com um falante/escritor (e.g., monólogos, cartas, avisos) ou vários interlocutores.

Como visto anteriormente, a Análise da Conversação se ocupa de estudar a interação social síncrona entre pelo menos dois falantes. Assim, vemos que a AD se distingue da Análise da Conversação no seu foco, bem como nos métodos utilizados

---

<sup>23</sup> Multimodais – que integram mais de um modo de comunicação (escrita, fala, gestos, expressões faciais, movimentos corporais).

para estudar/processar a língua natural. A AD busca identificar como cada frase influencia na interpretação das frases subsequentes, sendo indispensável para a resolução de referências pronominais e de dêiticos (e.g., amanhã, aqui, depois, etc.). Esses processos não são cobertos pela Análise da Conversação. Veremos agora alguns conceitos básicos dessa área necessários para o bom entendimento deste curso: tópico e focos do discurso, segmentos/segmentação do discurso, e marcadores discursivos.

Em geral, cada discurso possui um tema geral/central, que motiva sua existência – e.g., ensinar a fazer um bolo, informar sobre algum acontecimento, ou apenas conversar sobre o cotidiano (caso das conversações sociais livres). Esse tema será aqui chamado de *tópico do discurso*. Além disso, os discursos apresentam um ou mais *focos*, temas secundários, porém ligados ao tópico principal, que são abordados durante o discurso (e.g., onde comprar os ingredientes, como misturá-los, como cozer o bolo, etc.). No caso das conversações sociais, a mudança de foco ocorre com maior frequência, dada a maior liberdade desse tipo de interação. Já em discurso muito curtos (como notícias curtas), pode haver apenas um foco, que coincide com o tópico geral.

Assim como as frases são estruturas com base nos sintagmas (grupos nominais, grupos verbais, etc.), o discurso é organizado em *segmentos*, que são os seus *elementos estruturantes*. Cada segmento agrupa as frases consecutivas que tratam de um mesmo foco. Um segmento longo pode ser subdividido em *subsegmentos*, formando assim uma estrutura hierárquica de segmentos “aninhados”.

Do ponto de vista prático, a segmentação do discurso auxilia na resolução de referências pronominais (e.g., ela, eles) e dêiticos (e.g., aqui, hoje). Ao tentar resolver uma referência pronominal (e.g., ela), damos preferência a candidatos (substantivos femininos no singular) que ocorrem dentro do mesmo segmento de discurso no qual o pronome ocorreu. Em seguida, investigam-se os referentes que ocorrem nos segmentos de mais alto nível, hierarquicamente superiores ao segmento onde o pronome ocorreu. O mesmo procedimento pode ser adotado para os dêiticos.

Contudo, no discurso (escrito ou falado), nem sempre encontramos um delimitador claro para indicar as fronteiras de cada segmento. Delimitamos “mentalmente” essas fronteiras com base na identificação de mudanças no foco do discurso – como dito, cada segmento se refere a um único foco de discussão. Assim, a resolução de referências pronominais e dêiticos geralmente ocorre de forma simples. Porém, a *segmentação automática* do discurso é bem mais complexa, sendo um dos maiores problemas dos sistemas para processamento do discurso.

Vale a pena observar que existem alguns fenômenos linguísticos que podem auxiliar nessa segmentação. Dentre eles, destaca-se a utilização de *marcadores discursivos* (e.g., a propósito, em seguida, ao contrário, mudando de assunto) como elementos que sinalizam manutenção ou mudança de foco. Outra dica a observar são as *repetições* no discurso. Observa-se que a manutenção do foco do discurso, em geral, vem marcada por repetições dos mesmos termos (geralmente, substantivos), ou pelo uso de pronomes (que substituem esses substantivos). Assim, o surgimento de novos termos no discurso sinaliza uma mudança de foco.

Apesar de ser uma tarefa difícil, a segmentação automática do discurso é indispensável aos sistemas conversacionais. Tal tarefa tem sido tratada de maneiras

distintas, como pode ser visto na literatura relacionada (e.g., [Grosz e Sidner, 1986], [Walker e Whittaker, 1990], [Galley et al., 2003]).

No caso dos ACs textuais (chatterbots), a resolução de referências pronominais e dêiticos pode ser mais simples, dado que os turnos são claramente marcados, e o chatterbot tem mais “tempo” para processar as informações de que dispõe a fim de encontrar os referentes corretos. Nesses casos, o agente pode até confirmar com o usuário se fez uma substituição correta. Já no caso dos ACIs com processamento de voz, essa tarefa se torna mais difícil do que no caso das conversações entre humanos. Note que tais ACIs trabalham em tempo real, e não dispõem de todo o conhecimento que um humano tem para interpretar corretamente essas referências.

Para simplificar o processamento da conversação, alguns ACs simplesmente não aceitam o uso de pronomes e dêiticos. Porém, isso não parece uma boa prática, dado que um dos objetivos desses agentes é prover uma conversação verossímil. Assim, alguns ACIs trazem um módulo específico para tratamento do diálogo, a fim de prover resolução de referências em geral (ver Seção 4.5).

#### **4.3.2 Computação Afetiva para ACs**

Com o objetivo de aumentar o grau de realismo dos agentes virtuais, pesquisadores vêm construindo modelos que buscam simular características humanas, como a personalidade e a emoção. Esses trabalhos utilizam conceitos da *Computação Afetiva*, que estuda a modelagem computacional de fenômenos afetivos e influências emocionais [Picard, 2003] [Calvo et al., 2015]. De forma geral, podemos dizer que o objetivo desses modelos é detectar, interpretar, processar e simular emoções humanas.

Segundo [Bates, 1994], que trabalha com atores sintéticos, esses agentes devem agir e reagir de modo semelhante ao que entendemos por “comportamento humano”. Essas características também são desejáveis no contexto dos agentes conversacionais, uma vez que ACIs também buscam se assemelhar aos humanos, a fim de criar empatia com os usuários, tornando a comunicação mais natural e efetiva.

Esta seção discute brevemente modelos para modelagem de personalidade e de emoção em agentes virtuais. Vale ressaltar que os modelos de personalidade e de emoção não são excludentes, podendo ser usados de maneira complementar na modelagem de agentes artificiais.

#### **Modelagem Computacional de Personalidade**

Dentro da Psicologia, entende-se que a *personalidade* é um componente mais estável no ser humano, sendo a característica mais duradoura de cada indivíduo. A personalidade em geral determina o comportamento, as atitudes e ações mais frequentes do indivíduo.

Estudos na área da Psicologia já propuseram diversas teorias para modelar personalidade, sendo a abordagem taxonômica a mais utilizada para esse fim<sup>24</sup>. Essa abordagem classifica os indivíduos por meio de suas características: um indivíduo é classificado como introvertido se ele apresenta características como “quieto”, “tímido” e “reservado”.

---

<sup>24</sup> <http://personality-project.org/revelle.html>

Dentro dessa abordagem, a teoria de maior destaque para a Computação Afetiva é a *Teoria dos Traços* [Allport, 1938], que busca determinar quais traços de personalidade (características) podem caracterizar o comportamento de um indivíduo. A partir da *Teoria dos Traços*, foram criados modelos reduzidos que podem ser utilizados na modelagem computacional da personalidade.

Dentre os modelos criados, os que mais se destacam são o *Big Five* [Saucier e Goldberg, 1998] e o *Five Factor* [McCrae e Costa Jr., 1998]. Embora apresentem pequenas diferenças, na literatura, os dois modelos são conhecidos pelo termo *Big Five*, sem distinção. Ambos trabalham com a distribuição e redução dos traços catalogados por Allport em cinco fatores/dimensões.

Essas cinco dimensões são conhecidas pelo acrônimo OCEAN. O acrônimo vem das iniciais dos termos em inglês usados para denominar os aspectos escolhidos pelos autores como sendo determinantes na personalidade:

- *Openness* – ser curioso, ser criativo, estar aberto a viver novas experiências;
- *Consciousness* – ter responsabilidade, ser confiável, ser consciente, pensar nas consequências de seus atos antes de agir;
- *Extraversion* – ser extrovertido, ser sociável;
- *Agreeableness* – ser delicado/gentil, ter espírito de cooperação; e
- *Neuroticism* – ser ansioso, nervoso, demonstrar instabilidade emocional.

Como dito, esse modelo assume que a personalidade pode ser caracterizada a partir desses cinco aspectos, indicando-se o grau/peso de cada um deles na personalidade daquele indivíduo. Quando desejamos modelar a personalidade de agentes virtuais, é necessário inicialmente definir o tipo de personalidade desejada para o agente (amigável, extrovertido, etc.). A partir daí, devemos definir o peso de cada um dos cinco aspectos na personalidade do agente (por exemplo, se queremos modelar um agente muito introvertido, devemos associar um valor muito baixo ao aspecto *Extroversão*).

O intervalo de variação desses valores (e.g., de 0 a 5), bem como o incremento de variação (de 1 em 1; de 0,5 em 0,5; etc.) serão escolhidos por quem está construindo o agente. Se desejarmos uma modelagem mais refinada, devemos trabalhar com intervalos de variação menores. É necessário também definir a interação entre esses vários aspectos, para que o agente apresente uma personalidade credível.

### **Modelagem Computacional de Emoção**

Voltando a atenção agora à modelagem de *emoções*, podemos ver que essa característica varia de acordo com a situação atual em que o indivíduo se encontra, podendo um indivíduo vivenciar várias emoções ao longo de um mesmo dia (ou até ao mesmo tempo – e.g., alegria e vergonha; medo e surpresa).

Estudos na área de emoção podem ser classificados em quatro perspectivas teóricas (abordagens) diferentes: Darwiniana, Jamesiana, Cognitiva e Sócio Construtivista. De acordo com [Kasap e Magnenat-Thalmann, 2007], atualmente modelos de emoção baseados na abordagem Cognitiva têm sido preferidos (em

detrimento dos modelos baseados na abordagem Darwiniana), uma vez que eles explicam melhor o processo geral de como emoções ocorrem e afetam nossa tomada de decisão. Tais modelos levam em conta a nossa “apreciação” (percepção) do ambiente, incluindo não apenas a situação atual, mas também eventos passados e nossas expectativas sobre o futuro.

No que concerne a modelagem/simulação de emoção em agentes virtuais, destaca-se o *Modelo OCC* (sigla derivada dos sobrenomes dos seus criadores - Ortony, Clore e Collins) [Ortony et al., 1988]. Esse modelo pode ser considerado como padrão para representação computacional de emoções, tendo sido usado na construção de diversos agentes virtuais [Bartneck, 2002].

Vale relembrar que as ações realizadas no ambiente (repostas do agente) são diretamente influenciadas pelos seus estados emocionais. Isso inclui a escolha das expressões faciais, dos olhares, gestos e do texto falado/escrito, bem como a coordenação entre todas essas facetas. Daí a grande importância de se modelar adequadamente essa característica nos agentes virtuais credíveis.

O modelo OCC propõe 22 categorias emocionais, tais como: raiva, satisfação, amor, vergonha, admiração, esperança, etc. Contudo, esse modelo é bastante flexível, não impondo o uso de todas essas categorias na modelagem da emoção dos agentes. Essa escolha fica a critério do desenvolver do agente. Além disso, é possível também definir novas categorias de emoções a partir da combinação de emoções mais básicas (e.g., desapontamento = tristeza + surpresa), uma vez que, como já dito, é bastante comum vivenciarmos mais de uma emoção ao mesmo tempo.

Como no caso do OCEAN, aqui também devemos estipular intervalos e incrementos de variação para descrever os valores que as categorias de emoção podem assumir na modelagem do agente. O ideal seria trabalhar com espectros contínuos, mas isso não é possível.

O modelo OCC não deixa explícito como a emoção é influenciada pela personalidade, dando liberdade ao projetista para criar modelos diferentes. Contudo, independente do modelo de personalidade adotado, é necessário considerar a influência dessa característica na escolha das emoções que o agente vai demonstrar a cada turno de percepção-ação. Relembramos que essa escolha deve considerar também a percepção atual do ambiente, eventos passados e os objetivos e preferências do agente.

Por fim, destacamos que alguns autores propõem ainda a modelagem do *humor*, como um complemento ao estado emocional do agente. Segundo [Wilson, 2000], humor é um estado emocional prolongado, causado pelo efeito cumulativo de emoções momentâneas, sendo uma propriedade dinâmica, que varia no tempo. Contudo, a depender da personalidade, podemos sorrir, mesmo estando de mau humor.

#### 4.4. Agentes Conversacionais Textuais – *Chatterbots*

Esta seção inicia com a apresentação de algumas competições internacionais que realizam o Teste de Turing para avaliar o desempenho de chatterbots. A seguir, temos um histórico sobre chatterbots, com foco nas técnicas utilizadas para manutenção do diálogo com o usuário. Por fim, discutiremos os problemas ainda apresentados por esses bots, a algumas direções futuras na área. Como já dito, alguns chatterbots textuais apresentam uma imagem associada, porém não são considerados agentes incorporados, por não apresentarem as outras funcionalidades que caracterizam os ACIs.

##### 4.4.1 Prêmios e competições

Uma das competições mais conhecidas na área é o Prêmio Loebner (*Loebner Prize*)<sup>25</sup>, sendo vista como “a primeira instanciação formal de um Teste de Turing”. Trata-se de uma competição anual que premia o chatterbot que conseguir “enganar” os juízes por mais tempo. A competição implementa um “Teste de Turing restrito”, que se tornou mais viável do que a proposta inicial de passar no teste de modo irrestrito. No decorrer do texto, faremos menções a chatterbots vencedores do prêmio Loebner.

Recentemente, em 2014, um Teste de Turing foi realizado como parte de um evento especial, organizado pela *Royal Society* (Grã-Bretanha) para celebrar os 60 anos da morte de Turing. Nesse evento, o chatterbot *Eugene Goostman* foi premiado e reconhecido como sendo o primeiro sistema a passar no Teste de Turing. Eugene se fez passar por um adolescente com pouco conhecimento de inglês, e assim conseguiu enganar 33% dos juízes participantes do evento (mais do que os 30% considerados suficientes para que o software seja considerado inteligente)<sup>26</sup>. Destacamos ainda o *Chatterbox Challenge*<sup>27</sup>, outro evento de grande visibilidade na área.

##### 4.4.2 Histórico – Gerações de Chatterbots

A partir de uma análise sobre a literatura relacionada, é possível dividir a história dos chatterbots em três gerações, que se diferenciam principalmente pelas técnicas utilizadas na manutenção do diálogo com o usuário [Neves, 2005].

A *primeira geração* utilizava técnicas simples de casamento de padrão [Weizenbaum, 1966]; a *segunda geração* passou a utilizar técnicas mais complexas do ponto de vista de Inteligência Artificial, como, por exemplo, redes de ativação [Mauldin, 1994]; e a *terceira geração* se apoia em técnicas mais complexas de casamento de padrão, baseadas em XML [ALICE Bot, 2016].

##### Primeira Geração

O primeiro chatterbot com ampla divulgação foi ELIZA<sup>28</sup> [Weizenbaum, 1966], marcando o início da primeira geração de agentes conversacionais textuais. Esse sistema simula um psicanalista da linha Rogeriana [Eliza Chat Bot, 2016].

---

<sup>25</sup> [http://en.wikipedia.org/wiki/Loebner\\_Prize](http://en.wikipedia.org/wiki/Loebner_Prize)

<sup>26</sup> Ver [http://pt.wikipedia.org/wiki/Eugene\\_Goostman](http://pt.wikipedia.org/wiki/Eugene_Goostman) e <http://www.zdnet.com/article/computer-chatbot-eugene-goostman-passes-the-turing-test/>

<sup>27</sup> <http://www.chatterboxchallenge.com>

<sup>28</sup> <http://en.wikipedia.org/wiki/ELIZA>



Os *bots* da primeira geração buscam manter diálogos com o usuário utilizando técnicas muito simples de casamento de padrões com base em palavras-chave, podendo também tratar frases inteiras. De modo simples, esses sistemas utilizam pares <Pergunta-Resposta> (ou mais precisamente, pares <Estímulo, Resposta>) para manter o diálogo. A cada nova intervenção do usuário, o sistema busca em sua base a entrada mais semelhante à frase do usuário, devolvendo a resposta associada à entrada selecionada. Esses sistemas apresentam apenas um módulo de processamento.

O Exemplo 4.1 ilustra a técnica de casamento de padrões utilizada por ELIZA. Apresentamos trechos de diálogos com ELIZA, incluindo o padrão de pergunta-resposta selecionado para cada caso. Os termos sublinhados indicam as palavras-chave usadas no processo de casamento de padrão. Empregamos aqui os seguintes símbolos: <I> texto de entrada do usuário; <E> padrão de entrada de um par; <R> padrão de resposta associado a essa entrada na base de pares <E,R>; <C> resposta do chatterbot.

#### Exemplo 4.1: Trechos de diálogo com ELIZA.

<I1> Hi!

Padrão selecionado:

{<E> Hi/Hello; <R> Hello, how are you?}

<C1> Hello, how are you?

<I2> Today I am feeling a bit sad.

{<E> (slot1) I am feeling (slot2);\_<R> Why are you feeling (slot2)?}

<C2> Why are you feeling a bit sad?

O sucesso de ELIZA abriu caminho para essa nova aplicação, e as técnicas utilizadas na sua implementação influenciam a construção desses sistemas até os dias de hoje [Eliza Chat Bot, 2016]. Outro chatterbot dessa geração que alcançou fama foi PARRY [Colby et al., 1972], que simulava um doente mental com esquizofrenia e paranoia.

Apesar do sucesso desses bots, as limitações das técnicas utilizadas não favorecem a construção de um diálogo verossímil, coeso e coerente por muito tempo. Para se obter um grau aceitável de realismo nos diálogos, é necessário construir uma base de padrões muito grande, a fim de tentar cobrir as possíveis variações de frases dos usuários. Além disso, os sistemas iniciais não armazenavam conhecimento sobre um domínio particular ou informações sobre o usuário (seu perfil), o que resultava em diálogos muito superficiais e iguais para qualquer usuário. Por fim, os sistemas iniciais não tratavam adequadamente fenômenos discursivos, como referências pronominais e repetições, uma vez que eles mantêm apenas o registro de sua última réplica (com o intuito de evitar a repetição das suas próprias respostas). Algumas dessas questões foram tratadas pelos chatterbot da geração seguinte.

## Segunda Geração

A segunda geração de chatterbots (início da década de 1990) passou a incorporar técnicas mais avançadas, oriunda da IA e do PLN, a fim de aumentar o realismo dos diálogos. Os sistemas passaram a apresentar uma arquitetura mais modular, seguindo assim os avanços da Engenharia de Software.

Tais sistemas contam com um módulo de processamento morfossintático, a fim de cobrir variações morfológicas e sintáticas das frases dos usuários. Esse módulo implementa um mecanismo de normalização de palavras e frases, com base em uma tabela de substituições. Apesar de simples, essa solução conseguiu melhorar bastante a fluência dos diálogos com os chatterbots [Neves, 2005]. Além disso, alguns sistemas da dessa geração passaram a incorporar bases com conhecimento sobre tópicos específicos, seguindo também uma tendência da IA de construir Sistemas Especialistas. Por fim, tais sistemas também passaram a considerar informações referentes aos interlocutores, através da manutenção de um modelo simplificado do usuário baseado em frames.

O bot mais popular dessa geração é JULIA [Mauldin, 1994], desenvolvido para um ambiente de entretenimento virtual multiusuário de 1989 (TinyMUD<sup>29</sup>). JULIA participava do ambiente como se fosse um usuário humano interagindo com outros humanos através de *salas de chat*. Para dar mais credibilidade ao diálogo, JULIA guardava um perfil do usuário com informações básicas, como nome, país e cidade de origem. Essas informações eram adquiridas pelo *bot* durante o diálogo com o usuário.

A base de diálogos de JULIA consistia em um grafo (Rede de Ativação), no qual cada nó possui um título, palavras-chave, um valor de ativação e um molde de réplica (semelhante aos padrões de ELIZA, com *slots* a serem preenchidos) – ver Exemplo 4.2. Os nós estão ligados por arestas com pesos positivos ou negativos. O valor de ativação de um nó pode ser aumentado ou diminuído durante o processamento da entrada. Ao final, o nó com maior valor de ativação é selecionado.

A fim de aumentar a credibilidade, JULIA utilizava as informações do perfil do usuário para compor suas réplicas. Assim, no exemplo abaixo, o nome do usuário faria parte da réplica do bot, e.g.: "Do you have any pets Ryan?".

**Exemplo 4.2: Nó <havepet> da Rede de Ativação de JULIA.**

Título	<havepet>
Valor de ativação	1.0
Palavras-chave	*pet*
Réplica associada	Do you have any pets <name>?
Nós com ligações positivas	<havepet-1> <havepet-2>
Nós com ligações negativas	<havepet-6> <havepet-7>

Assim como ELIZA, JULIA influenciou toda uma geração de chatterbots, tendo servido como ponto de partida para muitos outros bots de sucesso, tais como JFRED<sup>30</sup> (um *framework* para o desenvolvimento de chatterbots). JFRED foi desenvolvido por Robby Garner<sup>31</sup>, vencedor do prêmio Loebner de 1998 e 1999 com o bot Albert One<sup>32</sup>.

Apesar do avanço em relação à primeira geração, os chatterbots dessa geração ainda apresentavam problemas variados. Assim, alguns sistemas da primeira geração continuaram superando os sistemas dessa geração nos concursos do Prêmio Loebner. O problema principal desses bots é a necessidade de se refazer a Rede de Ativação a cada

<sup>29</sup> TinyMUD (Tiny Multi-User Dungeons) - Um ambiente virtual de jogo multiusuário desenvolvido por James Aspnes em 1989 - <http://www.cs.yale.edu/homes/aspnes>

<sup>30</sup> <http://www.turinghub.org/JFRED.php>

<sup>31</sup> [http://en.wikipedia.org/wiki/Robby\\_Garner](http://en.wikipedia.org/wiki/Robby_Garner)

<sup>32</sup> [http://en.wikipedia.org/wiki/Albert\\_One](http://en.wikipedia.org/wiki/Albert_One)

novo domínio de aplicação (uma vez que essas redes são dependentes do domínio). Outras questões são a necessidade de controlar o andamento global da conversação (início, desenvolvimento e fechamento), e de oferecer tratamento adequado a intervenções (perguntas) desconhecidas [Neves et al., 2006].

### Terceira Geração

A terceira geração de chatterbots (iniciada no final da década de 1990) apresenta uma arquitetura modular e mais completa do ponto de vista do PLN, tendo como marco principal o uso de linguagens de marcação do tipo XML (*eXtensible Markup Languages*)<sup>33</sup>.

A primeira e mais importante linguagem dessa geração é a AIML (*Artificial Intelligence Markup Language*) [Wallace, 2009]. Tal linguagem possui uma estrutura simples do ponto de vista computacional (uma vez que se inspirou nas técnicas usadas por ELIZA), porém oferece uma capacidade de processamento de diálogos bem superior às técnicas usadas nos bots das duas primeiras gerações.

AIML pode ser vista como um novo “*framework*” para o desenvolvimento de chatterbots, oferecendo tratamento de informação em quatro níveis linguísticos: morfossintático, semântico, discursivo e pragmático. Para isto, o framework utiliza diversas variáveis de controle, bem como bases de dados/conhecimento em AIML.

#### Nível morfossintático:

- A *Base de substituição* é usada tanto para corrigir erros de digitação (e.g., `<substitute find= "becuase" replace= "because"/>`), quanto para padronizar termos a partir de variações de digitação que não são necessariamente um erro (e.g., `<substitute find= "ohh" replace= "oh"/>`).
- A *Base de categorias* realiza o tratamento sintático da entrada do usuário. Tal base consiste em unidades identificadas pelo elemento `<category>`, compostas por um padrão de entrada associado a um ou mais moldes de réplica, semelhante aos sistemas de primeira geração (Exemplo 4.3). Os padrões de entrada podem usar caracteres especiais (e.g., \* \_) para casar com frases variadas digitadas pelos usuários, substituindo esses caracteres por qualquer expressão (Exemplo 4.4). Assim, se um usuário digita “*What is an iPad?*”, e o sistema não tiver informação sobre isso, o bot pode responder “*I don't know what is an iPad*”.

#### Exemplo 4.3: Modelo geral de categoria AIML.

```
<category>
<pattern>padrão de entrada</pattern>
<template>
  <random>
    <li>molde de réplica 01
    <li>molde de réplica n
  </random>
</template>
</category>
```

<sup>33</sup> <https://pt.wikipedia.org/wiki/XML>

#### Exemplo 4.4: Categoria AIML com “\*”.

```
<category>
<pattern>what is *</pattern>
<template>I don't know what is <star>.</template>
</category>
```

##### Nível semântico:

- A *Base de descrição do Chatterbot* é formada por uma série de elementos `<property>`, que descrevem as características de cada bot em particular (e.g., `<property name="name" value="ALICE"/>`; `<property name="gender" value="female"/>`; `<property name="birthday" value="1995"/>`, etc.).
- A *Base de Tópicos* é formada por categorias classificadas por tópico. Essas são categorias mais específicas, que carregam informação semântica. Essa base pode ter, por exemplo, um conjunto de categorias específicas associadas ao tópico "LOVE", que só são utilizadas quando o tópico vigente do diálogo for "LOVE".

##### Nível discursivo:

- A *Base de substituição de pronomes* é semelhante à Base de substituição morfológica, porém os elementos `<substitute>` estão vinculados a um elemento `<person>`, indicando que essas substituições são pronominais (e.g., `<person>` `<substitute find=" he was " replace=" I was "/>`; ... `<substitute find=" to you " replace=" to me "/>`; ...; `</person>`).
- As *Variáveis de controle de dêiticos* (e.g., aqui, hoje, mais tarde, etc.) são definidas em tempo de execução. Essas variáveis são definidas pelo elemento `<set>` e recuperadas pelo elemento `<get>`, inseridos nos moldes de réplica das categorias AIML.

##### Nível pragmático:

- As *Variáveis com informação sobre o usuário* (seu perfil) são definidas da mesma forma que as variáveis de controle de dêiticos, e incluem informações sobre, por exemplo, o nome e gênero do interlocutor.

AIML foi usada na implementação de ALICE [ALICE Bot, 2016], que conseguiu destaque internacional ao vencer três edições do *Prêmio Loebner*, bem como uma edição do *Chatterbox Challenge*. Devido ao seu grande sucesso, ALICE foi a responsável pela difusão de AIML, que se tornou o estado da arte em tecnologia para criação de chatterbots. AIML já deu origem a centenas de chatterbots em línguas e domínio diversos<sup>34</sup>.

Por fim, vale citar aqui um trabalho que estendeu AIML padrão, incorporando informações de intencionalidade (no nível pragmático), a fim de evitar a quebra de fluência do diálogo e a perda de credibilidade do chatterbot [Neves, 2005]. O autor propôs a incorporação de novas *tags* de intenção ao AIML padrão, tendo por base a Análise da Conversação [Marcuschi, 1986].

<sup>34</sup> AIML bots - <http://www.alicebot.org/aimlbots.html>

#### 4.4.2 Aplicações Atuais, Limitações e Tendências Futuras

Esta seção conclui a discussão sobre agentes conversacionais chatterbots apresentando brevemente suas limitações e tendências futuras.

##### Aplicações Atuais

Apesar das suas limitações do ponto de vista do PLN, os chatterbots são utilizados por milhares de usuários na Web, não só para manter diálogos genéricos, mas também para intermediar o acesso a outros sistemas [Laven, 2016], podendo ainda desempenhar o papel de “acompanhantes digitais” do usuário em sistemas imersivos.

As aplicações de chatterbots são bastante variadas, podendo-se encontrar tais agentes como aplicações isoladas (*stand-alone*) na Internet para simples conversação (e.g., Eliza e ALICE), já bastante discutidos neste documento, ou embutidos em sistemas interativos variados (algumas vezes servindo como interface de acesso ou como acompanhante digital do usuário). Neste segundo caso, citamos: Chatterbots assistentes virtuais<sup>35</sup> (incluindo aqui bots do tipo FAQ - *Frequent Asked Questions* e SAC - Sistemas de Atendimento ao Cliente), bots em jogos digitais (incluindo jogos sérios), Chatterbots em sistemas de ensino/educação<sup>36</sup>, chatterbots interface de acesso a outros sistemas computacionais, entre outras.

As referências a seguir, algumas já indicadas anteriormente, oferecem uma grande quantidade e variedade de exemplos de chatterbots para os mais variados fins: [Laven, 2016], [ALICE Bot, 2016], [Chatbots.org, 2016].

Por fim, ressaltamos que existem várias empresas dedicadas à criação de chatterbots<sup>37</sup>. Algumas oferecem plataformas gratuitas para que o próprio usuário desenvolva seu bot (e.g., Pandora Bots<sup>38</sup>).

##### Limitações e Estratégias simples para manutenção dos diálogos

Apesar do seu grande sucesso, chatterbots apresentam limitações na manutenção de diálogos fluentes e verossímeis, como já dito. Essas limitações se devem principalmente à grande dificuldade em se construir um sistema completo de processamento de linguagem natural (que realize tanto interpretação como geração de LN), principalmente nas etapas de análise semântica e pragmática, que envolvem interpretação de sentido/significado e intenção/intensionalidade. Assim, esses bots incorrem em erros, tendo dificuldade para manter diálogos coerentes.

Algumas estratégias simples e bem-humoradas para driblar essas situações são apresentadas no texto online “10 Truques que Chatbots usam para fazer você acreditar que eles são humanos”<sup>39</sup>. Exemplos desses truques são: escreva uma frase completa a cada turno, mantenha o turno de fala (usando respostas como “Umm...”, que dão a impressão de que o bot está pensando), limpe a tela a cada resposta (para o interlocutor

---

<sup>35</sup> [https://www.chatbots.org/virtual\\_assistant/](https://www.chatbots.org/virtual_assistant/)

<sup>36</sup> [https://www.chatbots.org/industry/learn\\_lookup/](https://www.chatbots.org/industry/learn_lookup/)

<sup>37</sup> Empresas desenvolvedoras de chatterbots: <https://www.chatbots.org/companies/>

<sup>38</sup> <http://www.pandorabots.com/>

<sup>39</sup> 10 Tricks That Chatbots Use to Make You Believe They're Human. <http://io9.com/5901579/10-tricks-that-chatbots-use-to-make-you-believe-theyre-human>.

não ter a memória do diálogo, e então não perceber as quebras de coesão), mude de assunto com frequência (como acontece em diálogos reais), simule uma briga (isso não exige memória!), entre outros.

Além das dificuldades inerentes ao PLN, restam ainda as dificuldades relacionadas à criação de Bases de Conhecimento (BC) com informação pertinente ao domínio da aplicação, um problema geral da IA baseada em conhecimento. Por exemplo, se queremos construir um chatterbot para um sistema de EaD sobre História do Brasil, esse bot precisa de uma base de conhecimento que seja o mais completa possível, e que seja sempre mantida atualizada.

Quando o bot não encontra na BC a informação necessária para responder ao usuário, ele terá que contornar a situação, a fim de não interromper o diálogo. Na maioria das aplicações vistas, em caso de falta de informação na BC, os bots simplesmente oferecem links na Web relacionados à informação solicitada.

### **Tendências Futuras**

Chatterbots certamente continuarão a existir tanto como aplicações *stand-alone* como fazendo parte dos sistemas interativos do futuro (funcionando como interfaces), provendo acesso em linguagem natural aos sistemas de computação em geral. Contudo, em sua maioria, tais sistemas irão requerer comunicação via voz (em lugar de texto digitado). Para isto, é necessário utilizar sintetizadores e interpretadores de voz, capazes de intermediar a comunicação entre chatterbot e usuário. Um exemplo desses sistemas é o Siri do iPhone<sup>40</sup>.

Por fim, destacamos uma evidência de outra ordem: chatterbots *stand-alone* puramente textuais estão perdendo espaço nos dias atuais, com a crescente tendência de se desenvolverem ambientes virtuais imersivos, baseados em realidade virtual. Nesses casos, o usuário espera que o agente conversacional faça parte do ambiente simulado, possibilitando diálogos via voz, a fim de tornar a interação mais natural e realista. Claramente, quando o diálogo é realizado somente pelo meio textual, podem-se perder aspectos importantes da interação, fazendo com que o significado da mensagem seja mal interpretado.

Assim, vemos como tendência futura o crescimento dos agentes conversacionais incorporados com comunicação via voz, que oferecem, além das capacidades dialógicas, também uma imagem associada, ampliando assim sua esfera de uso. Esse será o tema da próxima seção deste documento.

---

<sup>40</sup> <https://www.apple.com/ios/siri/> & [http://pt.wikipedia.org/wiki/Siri\\_\(software\)](http://pt.wikipedia.org/wiki/Siri_(software))

## 4.5. Agentes Conversacionais Incorporados

Como já discutido, apesar do seu grande sucesso, chatterbots têm uso limitado em sistemas imersivos. Assim, voltamos agora atenção para os *Agentes Conversacionais Incorporados*. ACIs são agentes que possuem uma representação gráfica em 2D ou 3D (em geral, com aparência humana), sendo capazes de realizar diálogos por meio de texto, de fala e de outras modalidades não verbais, como olhares e gestos. Assim como os ACs textuais, encontramos ACIs na Web ou como aplicações *stand-alone*. Em geral, ACIs são usados em aplicações online, para prover serviços personalizados, como assistentes pessoais, ou ainda para auxiliar usuários a navegar em sites.

Alguns autores definem os ACIs como “Agentes Interface” (*Interface Agents*). Por exemplo, [Cassell e Vilhjálmsón, 1999] descrevem os ACIs como “interfaces multimodais” que implementam, tanto quanto possível, todos os elementos que os humanos usam para se comunicar. Incluem-se aqui fala, gestos e movimentos corporais, expressões faciais, linguagem corporal, e respostas com entonação baseada em emoção. Segundo [Cassell e Vilhjálmsón, 1999], ACIs podem ser definidos como interfaces de computador que conseguem lidar bem com as dificuldades de uma conversação, elaborando suas “falas” em função das exigências do diálogo, considerando emoção, personalidade e algumas características das conversações sociais.

Há também autores que nomeiam alguns ACIs como “Humanos Virtuais” (do inglês, *Virtual Humans*), principalmente quando eles incluem capacidades mais avançadas de raciocínio e demonstram emoção (e.g., [Egges et al., 2004]). Assim, optamos por não excluir nenhum trabalho apenas com base na terminologia utilizada para nomear os agentes, e sim com base na descrição do agente.

A seguir, a Seção 4.5.1 traz um breve histórico da área. Mudando o foco para questões mais técnicas, a Seção 4.5.2 apresenta uma arquitetura genérica para construção de ACIs, discutindo detalhes sobre seus componentes. A Seção 4.5.3 discute algumas abordagens adotadas na implementação de desses agentes. A Seção 4.5.4 traz aplicações atuais, e a Seção 4.5.5 apresenta limitações e tendências futuras na área.

### 4.5.1 Breve Histórico

No final da década de 1990, em paralelo ao ainda intenso trabalho no campo da construção de diálogos coerentes para chatterbots, observamos os primeiros esforços na construção de ACs incorporados. Os primeiros pesquisadores a se destacarem nessa área foram Justine Cassell<sup>41</sup> (um dos seus maiores expoentes), que iniciou seus trabalhos com *agentes conversacionais autônomos animados* [Cassell et al., 1994], e Joseph Bates, que trabalhava na modelagem de emoção para agentes credíveis [Bates, 1994].

O termo “Agente Conversacional Incorporado” (do inglês, *Embodied Conversational Agent*<sup>42</sup>) foi usado pela primeira vez em 1998, no *First Workshop on Embodied Conversational Characters*<sup>43</sup>. O foco inicial das pesquisas em ACIs foi na *modelagem da comunicação entre humanos*, uma vez que essa é a atividade central dos

---

<sup>41</sup> <http://www.justinecassell.com/>

<sup>42</sup> [https://www.chatbots.org/embodied\\_conversational\\_agent/](https://www.chatbots.org/embodied_conversational_agent/)

<sup>43</sup> <http://www.sigchi.org/chi98/events.html>

ACIs. Os pesquisadores buscavam definir quais aspectos da comunicação entre humanos poderiam ser simulados através dos ACIs, a fim de torná-los mais credíveis, e quais seriam as áreas de pesquisa relacionadas a essa simulação.

Segundo [Cassell et al., 2000a], tais agentes devem ter capacidade de: reconhecer entradas verbais e não verbais; responder com saídas verbais e não verbais; lidar com elementos organizadores da conversação, tais como tomada de turno e reparo (Seção 4.3.1); reconhecer a fase da conversação; e contribuir com o discurso gerando novas proposições (não apenas respostas), sendo assim um agente proativo.

Contudo, apenas estudos na área de comunicação (interpretação/geração de texto e processamento de voz) não eram suficientes para prover todas as características “humanas” aos ACIs. Assim, observamos o surgimento de grupos multidisciplinares, capazes de integrar diferentes tecnologias a fim de criar ACIs mais complexos e multifacetados. Destacam-se aqui o *MIT Media Lab* (em particular, o *Gesture and Narrative Language Group*<sup>44</sup>); e o *Information Science Institute (ISI)*<sup>45</sup> da *University of Southern California (USC)*.

A partir daí, a pesquisa passou a priorizar a construção de agentes animados, com representação visual em 3D, capazes de observar seu contexto, e ainda de exibir emoções que mudariam de acordo com o contexto do diálogo. Um dos primeiros ACIs desenvolvidos no ISI foi STEVE (*Soar Training Expert for Virtual Environments*)<sup>46</sup>, um agente tutor que ensinava, interativamente, como operar um painel de controle de navios [Rickel e Johnson, 1999]. STEVE se comunicava com o usuário através de voz sintetizada e possuía imagem em 3D. O ACI inicial foi expandido e usado em um projeto de construção de ambientes mais imersivos para treinamento militar.

Por volta de 2000, a USC criou o *Institute for Creative Technologies (ICT)*<sup>47</sup>, que passou a abrigar muitos dos pesquisadores do ISI. No ICT, esses pesquisadores desenvolveram muitos outros agentes pedagógicos (tutores) que ajudavam usuários a desenvolver habilidades em análise de tarefas, negociação, tomada de decisão e lideranças complexas. Tais agentes eram capazes de compreender entradas escritas ou faladas, apresentando uma representação visual que incluía linguagem corporal, expressões faciais, olhares variados, entre outras habilidades [Gratch et al., 2002].

Esses agentes já apresentavam uma modelagem de emoções sofisticada, sendo capazes de demonstrar raiva, e até de se recusar a continuar interagindo com o usuário. Para tanto, eles necessitavam manter uma memória com conhecimento sobre seu ambiente e sobre os outros participantes do ambiente (virtuais ou humanos). Atualmente, o ICT conta com vários protótipos de Humanos Virtuais<sup>48</sup>, a maioria deles voltada a aplicações militares de ajuda, havendo também alguns agentes tutores.

Em paralelo aos trabalhos do ISI e do ICT, o *Gesture and Narrative Language Group (GNL)* do *MIT Media Lab* também contribuiu muito para o desenvolvimento da área. Esse grupo, extinto em 2003, era liderado por Justine Cassell - atualmente

---

<sup>44</sup> GNL - <http://www.media.mit.edu/gnl/>

<sup>45</sup> ISI - <http://www.isi.edu/>

<sup>46</sup> STEVE - <http://www.isi.edu/isd/VET/steve-demo.html>

<sup>47</sup> ICT - <http://ict.usc.edu/>

<sup>48</sup> Protótipos do ICT - <http://ict.usc.edu/prototypes/>



trabalhando na *Carnegie Mellon University*<sup>49</sup>. Apesar de extinto, o site do GNL ainda guarda uma lista dos projetos desenvolvidos na época<sup>50</sup>.

Os pesquisadores do GNL dividiram o trabalho realizado pelo grupo em três gerações de humanoides (ou humanos virtuais)<sup>51</sup>. A primeira geração tem como representante o sistema *Animated Conversation*<sup>52</sup>, no qual dois caracteres autônomos animados conversam entre si. Apesar de não haver nenhum humano participando do diálogo, esse foi o primeiro sistema capaz de gerar automaticamente comunicação na forma verbal (entonação e fala) e não verbal (gestos com as mãos e expressões faciais).

A segunda geração é representada por *Gandalf*<sup>53</sup>, um *cartoon* animado (representado apenas pelo seu rosto) que era capaz de responder perguntas sobre o sistema solar usando processamento de voz. Gandalf também era capaz de perceber movimentos a partir de um sistema eletromagnético de rastreamento usado pelo usuário. Assim, ele era capaz de responder também a comportamentos não verbais. Apesar de Gandalf operar em tempo real, suas saídas (*outputs*) eram selecionadas a partir de uma biblioteca de respostas prontas (não sendo, portanto, muito flexível).

A terceira geração de humanoides passou a apresentar um corpo totalmente articulado (com maior flexibilidade de movimentos), sendo também capaz de perceber o usuário através de câmeras. Dentre eles, destaca-se REA, certamente o ACI mais conhecido do grupo [Cassel et al., 1999], e que serviu de base para o desenvolvimento de muitos outros ACIs.

REA (*Real Estate Agent*)<sup>54</sup> desempenha o papel de uma corretora de imóveis que mostra propriedades virtuais aos usuários e tenta vender uma casa. Esse domínio foi escolhido por possibilitar conversações tanto orientadas a tarefas (e.g., vender imóveis), como conversações sociais. Ao contrário de Gandalf, REA sintetiza suas respostas em tempo real (incluindo aqui fala e gestos com a mão), levando em conta o contexto comunicativo. Para tanto, o sistema utiliza uma gramática e um léxico, contando ainda com um modelo simples de discurso para determinar atos de fala [Searle, 1976], bem como para resolver e gerar referências pronominais.

Atualmente, encontramos no *MIT Media Lab* dois grupos de pesquisa que mantêm alguns projetos relacionados ao desenvolvimento de humanos virtuais/ACIs: o grupo *Personal Robots*<sup>55</sup> e o grupo *Affective Computing*<sup>56</sup>.

Observando os trabalhos desenvolvidos na última década, nota-se a preocupação em prover uma sincronização mais realista entre som, gestos e olhares, bem como com a coerência do comportamento do agente em relação ao contexto do diálogo com o usuário. Além disso, tais agentes devem incorporar características como personalidade e emoção. Alguns desses trabalhos mais recentes serão mencionados nas Seções 4.5.3 e 4.5.4.

---

<sup>49</sup> <http://www.justinecassell.com/>

<sup>50</sup> <http://www.media.mit.edu/gnl/projects.html>

<sup>51</sup> <http://www.media.mit.edu/gnl/projects/humanoid/>

<sup>52</sup> *Animated Conversation* - <http://www.media.mit.edu/gnl/projects/animated/>

<sup>53</sup> *Gandalf* - <http://alumni.media.mit.edu/~kris/gandalf.html>

<sup>54</sup> *REA* - <http://www.media.mit.edu/gnl/projects/humanoid/>

<sup>55</sup> <http://www.media.mit.edu/research/groups/personal-robots>

<sup>56</sup> <http://www.media.mit.edu/research/groups/affective-computing>

Por fim, destacamos um trabalho recente, que surgiu da demanda por mais realismo nas aplicações em tempo real com imersão: os *ACIs Avatares* (ACIAs) [Morie et al., 2013]. Tais agentes, especialmente projetados para ambientes (mundos) virtuais persistentes, são incorporados em avatares 3D - o mesmo tipo de representação que os participantes humanos usam-, estando portanto sujeitos às mesmas regras e restrições (de interação, de design, etc.) utilizadas na construção dos avatares de usuários humanos desses ambientes virtuais. Note que esses ACIs avatares são diferentes dos apresentados por [Cassell e Vilhjálmsdóttir, 1999], que são avatares representando usuários em ambientes virtuais multiusuários.

Os ACIAs devem ser capazes de se adaptar a ambientes em mudança constante, bem como de reconhecer os usuários do mundo virtual, incorporados em uma grande variedade de avatares. Além disso, eles devem ser persistentes em um mundo que, a princípio, nunca é “desligado”. Além disso, caso o mundo virtual sofra uma interrupção e seja reiniciado (um *reboot*), o ACIA deve ser capaz de se reiniciar automaticamente.

#### 4.5.2 Modelando/Projetando ACIs

Esta seção apresenta uma arquitetura genérica para ACIs, e discute alguns aspectos centrais no *design* desses agentes. Segundo [Kasap e Magnenat-Thalmann, 2007], esses aspectos são: a forma de *interação* com o agente, a sua *autonomia* para controlar dinamicamente o andamento do diálogo, e a *personificação* do agente, através da inclusão de características humanas, como personalidade e emoção. O objetivo central de cobrir todos esses aspectos é simular a experiência da conversação face a face entre humanos, buscando estabelecer um melhor engajamento do usuário no diálogo.<sup>57</sup>

##### 4.5.2.1 Arquitetura genérica

Com o crescimento das aplicações em tempo real, como *games* e interfaces, *comportamento autônomo* tornou-se uma característica vital para agentes virtuais - em particular, para os ACs (incorporados ou não), que devem ser capazes de dialogar de forma natural e coerente com seus interlocutores (usuários do sistema).

Contudo, prover comportamento autônomo a ACIs é uma tarefa complexa, por envolver não apenas a escolha do conteúdo (textual ou falado) a ser apresentado ao usuário, como também o tom de voz, e por fim a imagem (movimentos, olhares, expressões faciais), que devem estar sincronizados e devem ser coerentes com o contexto do diálogo em curso. Deve-se ainda considerar o estado emocional do agente em cada turno do diálogo.

Voltando às definições vistas na Seção 4.2.2, relembramos que agentes inteligentes funcionam em ciclos “percepção-raciocínio-ação”, definindo seu comportamento com base nas percepções oriundas do ambiente em que estão imersos. Mais detalhadamente, os agentes autônomos devem ser capazes de:

- *Perceber* seu ambiente através de sensores (e.g., teclado do computador, câmeras, microfone, protocolos de comunicação em sistemas multiagentes);

---

<sup>57</sup> Note que esses autores trabalham com “humanos virtuais inteligentes”, que são em geral mais complexos do que os ACIs.

- *Raciocinar*, utilizando as percepções do ambiente, conhecimento do domínio e regras, para escolher a melhor ação possível a realizar no ambiente naquele momento; e
- *Agir* sobre esse ambiente através de atuadores (*e.g.*, tela do computador, autofalante, funções internas que enviam mensagens a outros agentes, etc.).

Assim, uma arquitetura genérica para modelar agentes inteligentes deve contar com três módulos básicos, um para cada “tarefa macro” (perceber, raciocinar e agir). Contudo, observamos uma modularização mais refinada nas arquiteturas propostas para ACIs, que geralmente dividem esses três módulos básicos em componentes menores e mais específicos. Isso se deve à complexidade de modelar e construir ACIs.

De acordo com [Kasap e Magnenat-Thalmann, 2007], a maioria dos trabalhos separa a modelagem da “mente” do agente da modelagem das suas ações, uma vez que a mente humana trabalha com objetivos de mais alto nível (como “ir ao supermercado” para então comprar os produtos desejados), em lugar de planejar detalhes de movimento como “vou dar um passo com meu pé esquerdo”. Isso justifica as arquiteturas modulares usadas para a construção desses agentes.

A arquitetura de alto nível apresentada a seguir foi inspirada em [Kenny et al., 2007], que trabalha com humanos virtuais, contando com 7 componentes:

1. *Reconhecimento de fala* - produz texto a partir da entrada de voz do usuário;
2. *Reconhecimento de imagens* - produz uma representação interna da percepção visual do agente (gestos, expressões do usuário);
3. *Interpretação de linguagem natural* - constrói uma representação semântica interna a partir do texto da entrada;
4. *Raciocínio*<sup>58</sup> - raciocina com base nas percepções atuais usando regras/planos (base de conhecimento), e gera ações (na forma de representação interna do texto e da imagem de saída).
  - Em geral, os agentes têm um *modelo do mundo* (memória), para guardar os estados do ambiente, bem como as percepções e ações anteriores;
  - Agentes mais sofisticados contêm um *modelo do discurso* ou *modelo de diálogos orientados a tarefas* (para ACIs orientados a tarefa), *modelo do usuário*, e *modelo de emoção/personalidade*.
5. *Geração de comportamento não verbal* - seleciona gestos, posturas e olhares para o agente com base no texto de resposta, no histórico do diálogo e na emoção do agente;
6. *Síntese de voz* – mapeia o texto de saída do agente para som (voz sintetizada);
7. *Animação* – gera a ação/reação de saída do ACI em resposta às percepções atuais, sincronizando a voz com os gestos e os olhares selecionados previamente.

Observe que, como essa arquitetura geral foi proposta para Humanos Virtuais, nem todos os ACIs implementam todos os componentes citados aqui. O tipo de

---

<sup>58</sup> [Kenny et al., 2007] nomeiam esse módulo como “Agente Inteligente”.

interação e o tipo de comportamento escolhidos para o ACI vão influenciar diretamente na escolha da arquitetura ideal para cada ACI.

Antes de avançar na apresentação dos componentes dessa arquitetura, consideramos importante discutir sobre formalismos usados para representação interna das percepções e ações do agente (fala/texto, gestos, movimentos, e emoção do usuário). Note que essas representações internas são utilizadas na passagem de informação entre os componentes da arquitetura geral.

#### 4.5.2.2 XML como padrão para comunicação multimodal

Seguindo a tendência iniciada com os chatterbots (e.g., ALICE com AIML), os ACIs atuais também adotam linguagens do tipo XML como representação interna do conhecimento relacionado ao diálogo multimodal - percepções e ações do agente, bem como para representar a base de padrões de entrada e de saída desses agentes.

Como visto anteriormente, o texto de entrada, produzido pelo componente (1), é mapeado em uma representação interna pelo componente (3). Os padrões de entrada da base são então utilizados pelos componentes (2) e (3), a fim de gerar uma representação interna da entrada completa do usuário (texto e dados multimodais). Já os padrões de saída são usados na composição da resposta do agente para o turno atual do diálogo.

Podemos citar aqui diversas linguagens tipo XML usadas na construção de ACIs, com destaque para o XML usado pela ferramenta BEAT [Cassel et al., 2001], e para as linguagens *Function Markup Language* (FML) e *Behavior Markup Language* (BML), utilizadas pelo framework SAIBA (ver detalhes na Seção 4.5.3.3). Existem também linguagens específicas para representar comunicação não verbal através de gestos, por exemplo, a SiGML (*Signing Gesture Markup Language*)<sup>59</sup>, uma linguagem de marcação para gestos do agente.

A seguir, temos um exemplo de saída em XML gerada pela ferramenta BEAT [Cassel ET AL., 2001].

##### Exemplo 4.5: Exemplo de saída para o texto “*It is some kind of a virtual actor.*”

```
<AnimEvent: GAZE w=1 t=0.0spec=AWAY_FROM_HEARER>
<AnimEvent: GAZE w=3 t=0.517 spec=TOWARDS_HEARER>
<AnimEvent: R_GESTURE_START w=3 t=0.517 spec=BEAT>
<AnimEvent: EYEBROWS_START w=3 t=0.517 spec=null>
<AnimEvent: L_GESTURE_START w=7 t=1.338 spec=ICONIC_VIRTUAL >
<AnimEvent: R_GESTURE_START w=7 t=1.338 spec=ICONIC_VIRTUAL >
<AnimEvent: EYEBROWS_START w=7 t=1.338 spec=null>
<AnimEvent: L_GESTURE_END w=9 t=2.249 spec=null>
<AnimEvent: R_GESTURE_END w=9 t=2.249 spec=null>
<AnimEvent: EYEBROWS_END w=9 t=2.249 spec=null>
```

Esse exemplo traz a definição do comportamento do agente em relação a cada modo de interação: *gaze* = olhar, *gesture* = gesto, *eyebrows* = posição das sobrancelhas. Os códigos que aparecem nas definições são ponteiros para o banco de dados que

<sup>59</sup> SiGML - [http://www.visicast.cmp.uea.ac.uk/eSIGN/Images/Flyer\\_SiGMLSigning.pdf](http://www.visicast.cmp.uea.ac.uk/eSIGN/Images/Flyer_SiGMLSigning.pdf)  
<http://jeroenarendsen.nl/2007/01/esign-editor-turns-hamnosys-into-sigml-for-signing-avatars/>

armazena a representação de cada elemento visual. A síntese da fala é realizada em outra etapa de processamento.

Gostaríamos ainda de apresentar brevemente uma iniciativa do consórcio W3C na área de interação multimodal. Trata-se do trabalho do *W3C Multimodal Interaction Working Group*<sup>60</sup>, cujo objetivo central é propor e desenvolver padrões que possibilitem estender a Web, para permitir *interação multimodal* (e.g., voz, texto digitado, caneta, visão, gestos, interfaces gráficas, interfaces hápticas, etc.) *de modo amplo* (todos, em qualquer lugar, a qualquer hora, usando qualquer dispositivo), com serviços que se adaptem ao dispositivo, ao usuário e às condições do ambiente.

Destacamos aqui a *Arquitetura Multimodal* proposta pelo grupo, uma arquitetura genérica para interfaces multimodais que pode ser estendida à medida que novos modos de acesso sejam criados/disponibilizados. Essa arquitetura trabalha com dados representados usando linguagens do tipo XML:

- *EMMA - Extended Multimodal Annotation Language*, utilizada para representar informação multimodal automaticamente extraída da entrada do usuário;
- *InkML - Ink Markup Language*, projetada para lidar com escrita cursiva, gestos, desenhos, notação musical, entre outras linguagens notacionais; e
- *EmotionML - Emotion Markup Language*, utilizada para representar emoção e outros estados afetivos.

Apesar de essa iniciativa não ser voltada ao desenvolvimento de ACIs, e sim a interfaces multimodais em geral, ele é relevante para nós, uma vez que trata do estabelecimento de padrões para comunicação multimodal. Além do seu valor técnico, esse trabalho aponta para uma clara tendência de se trabalhar cada vez mais com linguagens do tipo XML.

Passaremos agora a discutir os componentes da arquitetura genérica para ACIs.

#### 4.5.2.3 Componentes de interação com o ambiente

Observando a arquitetura genérica descrita na Seção 4.5.2.1, vemos que os componentes (1) e (2) são responsáveis pela percepção do agente, enquanto os componentes (6) e (7) são responsáveis pela realização de ações no ambiente. De acordo com [Pauchet e Sabouret, 2012], a *interação* entre usuário e ACI pode ocorrer de três maneiras:

- *Percepção com múltiplos inputs*: textual (via caixas de texto), fala (via microfone), sensores de movimento (via câmera), etc.
- *Ação com múltiplos outputs*: textual, áudio, expressões faciais, movimentos corporais, iluminação da cena, etc.
- *Percepção e ação multimodal*: incluindo conversação via fala, e podendo incluir percepção do estado do ambiente e do usuário, tratamento de emoção, etc.

A escolha de quais componentes serão implementados no ACI vai depender diretamente do tipo de interação desejada.

---

<sup>60</sup> W3C MMI Group - <http://www.w3.org/2011/03/mmi-charter>

Não será possível discutirmos aqui as técnicas utilizadas pelos componentes de interação com o ambiente, nem os principais problemas enfrentados pelas áreas de pesquisa relacionadas com processamento (interpretação e síntese) de voz e imagem, uma vez que as autoras do curso não trabalham nessas áreas de pesquisa. Para os interessados nesses temas, indicamos como ponto de partida o trabalho de [Kasap e Magnenat-Thalmann, 2007], que traz uma ampla revisão do estado da arte sobre Humanos Virtuais. Apesar de não ser uma publicação muito recente, essa foi a única revisão realmente abrangente que encontramos sobre o assunto. As outras publicações tendem a relatar pesquisas particulares, com uma seção sobre trabalhos relacionados muito reduzida e focada no tema central da pesquisa sendo relatada.

Kasap e Magnenat-Thalmann (2007) discutem, na seção 2.1 do seu artigo, a modelagem da *percepção visual* do agente como um todo, indicando referências na área. A seção 3 desse mesmo artigo foca especificamente em ACIs, apresentando técnicas e trabalhos sobre modelagem de *expressões faciais* do ACI, e trabalhos sobre comunicação não verbal do ACI através de *gestos*. Por fim, a seção trata ainda da sincronização de apresentações multimodais (movimentos, expressões faciais, olhares, fala), destacando a SiGML, já mencionada aqui (Seção 4.5.2.2). Veremos mais tarde que alguns ACIs guardam apenas o padrão textual de saída, gerando os dados multimodais correspondentes (voz e imagem) em tempo real, enquanto outros ACIs mantêm registros completos com texto e os dados multimodais correspondentes.

#### **4.5.2.4 Componentes de Raciocínio**

Os componentes (4) e (5) da arquitetura genérica apresentada aqui são responsáveis por determinar as ações que o ACI deve realizar a cada turno de percepção-raciocínio-ação, levando em conta as percepções atuais do agente, os seus objetivos (indispensável para diálogos orientados a tarefa), e o conhecimento disponível através dos modelos mantidos pelo agente (modelo do mundo, modelo do discurso, modelo do usuário e modelo de emoção/personalidade).

Veremos mais tarde que algumas implementações de ACIs aglutinam esses componentes em um só, enquanto que outras implementações desmembram esses componentes em vários módulos específicos, sendo alguns dedicados a gerenciar os modelos mantidos pelo agente durante a interação com o usuário (Seção 5.3). Por serem de natureza distinta, cada um desses módulos pode utilizar abordagens diferentes de representação do conhecimento (e.g. lógica, objetos) e raciocínio.

A partir da literatura relacionada, podemos identificar diferentes abordagens para a construção dos componentes de tomada de decisão de ACIs/humanos virtuais (e.g., raciocínio simbólico baseado em regras, aprendizagem de máquina, abordagem estatística) [Russel e Norvig, 2010].

Ao examinar as implementações de ACIs, nota-se uma predominância da abordagem baseada em IA simbólica, na qual o agente “raciocina” aplicando regras predefinidas ao conhecimento disponível [Kasap e Magnenat-Thalmann, 2007]. Contudo, já existem algumas implementações de ACIs baseadas em Aprendizagem de Máquina (AM), tema que será brevemente discutido no final desta seção.

## Raciocinando com base em Regras e Conhecimento

Observando os modelos e as implementações de ACIs existentes, podemos distinguir a utilização predominante de três abordagens de raciocínio simbólico: a abordagem deliberativa, a reativa, e a abordagem baseada em objetivos. Vale ressaltar que essas abordagens não são excludentes, como veremos na Seção 4.5.3, que traz um modelo de ACI que combina as duas primeiras abordagens aqui citadas.

Inicialmente, ressaltamos que as arquiteturas mencionadas nesta seção se referem apenas ao componente de raciocínio do ACI<sup>61</sup>, diferentemente da arquitetura geral descrita na Seção 4.5.2.1. Contudo, os pesquisadores da área denominam esses modelos como “arquiteturas”. Assim, optamos por manter a nomenclatura utilizada na literatura relacionada.

Como representante principal da *abordagem deliberativa*, temos a arquitetura SOAR<sup>62</sup> (um acrônimo para *State, Operator And Result*), baseada em regras de produção. Essa arquitetura foi lançada em 1983, e alcançou grande sucesso, encontrando-se atualmente na sua versão 9. Um dos primeiros agentes construídos com base na Soar foi STEVE, já citado na Seção 4.5.1. O site oficial do projeto Soar<sup>63</sup> apresenta a evolução dessa arquitetura, que passou a incorporar também aprendizagem baseada em reforço e modelagem de emoção.

Apesar do sucesso das versões iniciais da Soar, a abordagem deliberativa não se mostrou adequada para sistemas em tempo real. Surgiram então arquiteturas com foco nos *sistemas reativos dinâmicos*, tendo como representante principal a *arquitetura de subordinação* de Brooks [Brooks, 1991]. Essa arquitetura, voltada aos sistemas multiagentes, mantém uma hierarquia de subordinação que define a precedência de cada agente em relação aos outros. Os comportamentos dos agentes são representados dentro dessa hierarquia com estruturas simples (semelhantes a estruturas *IF-THEN*). Além disso, existem soluções baseadas em Máquinas de Estado Finito, bem como abordagens baseadas em Probabilidade e em Lógica Difusa. Uma leitura importante sobre esse assunto é [Wooldridge, 2002].

Atualmente, as arquiteturas mais usadas em ambientes virtuais são as *baseadas em objetivo* (do inglês, *goal-driven architectures*), tendo como principal representante a arquitetura BDI<sup>64</sup> (acrônimo para *Belief-Desire-Intention* – Crenças, Desejos e Intenções). Esse modelo, que surgiu na década de 1970, ainda é amplamente usado na implementação de tipos variados de agentes, incluindo ACIs e atores sintéticos. Devido à sua importância, esse modelo será brevemente apresentado a seguir.

## Modelo/Arquitetura BDI

No final da década de 1980, surgiu o modelo BDI (*Belief-Desire-Intention* – Crenças, Desejos e Intenções) [Rao e Georgeff, 1995], que foi e ainda é largamente usado para modelar agentes virtuais creíveis, devido à sua acurácia. Outra referência importante

---

<sup>61</sup> Excetuando-se aqui a arquitetura SOAR, que é completa.

<sup>62</sup> [http://en.wikipedia.org/wiki/Soar\\_\(cognitive\\_architecture\)](http://en.wikipedia.org/wiki/Soar_(cognitive_architecture))

<sup>63</sup> <http://soar.eecs.umich.edu/>

<sup>64</sup> [http://en.wikipedia.org/wiki/Belief-desire-intention\\_software\\_model](http://en.wikipedia.org/wiki/Belief-desire-intention_software_model) & [http://www.inf.ufrgs.br/prosoft/bdi4jade/?page\\_id=46](http://www.inf.ufrgs.br/prosoft/bdi4jade/?page_id=46)

sobre o assunto é [Wooldridge, 2002]. O modelo/arquitetura BDI é baseado na Teoria do Raciocínio Prático de Bratman [Bratman, 1987]. Essa teoria busca explicar o raciocínio prático humano, que é o processo de decidir o que fazer a cada momento no tempo (turno percepção-ação), de modo a alcançar os objetivos desejados.

Os agentes BDI trabalham com crenças, desejos e intenções, sendo modelados para demonstrar “atitudes mentais”. As *crenças* do agente consistem no conhecimento que ele tem sobre o mundo, incluindo o conhecimento sobre si mesmo (o que ele é capaz de fazer) e sobre os outros agentes do ambiente. As crenças podem incluir regras de inferência, que permitem ao agente deduzir novas crenças a partir do conhecimento atual. O termo “crença” é usado para deixar claro que as crenças do agente podem não corresponder à verdade, e que podem mudar ao longo do tempo.

Os *desejos* representam o estado motivacional do agente, consistindo em situações ou objetivos a serem alcançados. O agente deve buscar ativamente alcançar seus objetivos. Note que o agente pode ter desejos conflitantes (por exemplo, economizar dinheiro e comprar o melhor carro do mercado), porém o conjunto de *desejos ativos* em cada instante não pode ser conflitante. Os desejos que o agente se compromete a atingir, i.e, seus desejos ativos, também são conhecidos como Objetivos.

As *intenções* representam o estado deliberativo do agente (o que ele escolheu fazer). As intenções podem ser vistas como os desejos que o agente se comprometeu a realizar (em geral, executando um *plano* para chegar aos objetivos). Planos podem ser definidos como sequências de ações que o agente realiza para alcançar suas intenções.

A atividade (ou reatividade) dos agentes é, em geral, disparada por *eventos*, que podem ser externos (percebidos via sensores), ou internos (gerados com o intuito de disparar atualizações ou planos de atividades). Eventos podem atualizar as crenças do agente, disparar planos ou modificar alguns de seus objetivos.

Apesar de ser amplamente utilizada em aplicações diversas, a arquitetura BDI tem algumas limitações já conhecidas, destacando-se aqui a falta de uma representação explícita dos objetivos, e de mecanismos para planejamento “para frente”. Além disso, o modelo BDI não oferece mecanismos explícitos para aprendizagem, o que dificulta a implementação de agentes com capacidade de aprender e de se adaptar. Existem, contudo, diversos trabalhos que estendem o modelo BDI, buscando sanar suas limitações (e.g., [Ancona e Mascardi, 2004], [Pokahr et al., 2005], entre outros).

### **Modelo do Discurso/Diálogo**

Como visto, ACIs são projetados para suportar diálogos multimodais, combinando texto, voz, gestos, e por vezes também movimentos corporais. Assim, o gerenciamento de diálogos com ACIs é naturalmente mais complexo do que no caso dos ACs textuais.

A fim de prover um controle mais sofisticado do fluxo do diálogo, alguns ACIs implementam um módulo exclusivo para Gerenciamento de Diálogo<sup>65</sup>, com regras que definem estratégias para guiar a escolha dos padrões de saída que o agente deve

---

<sup>65</sup> Contudo, notem que existem ACIs cujo diálogo é gerenciado por módulos de mais alto nível, que coordenam o comportamento do ACI como um todo.



apresentar ao usuário. Esse módulo mantém uma *Base de Padrões* (e.g., pares adjacentes) usados durante o diálogo, e um *Modelo do Discurso*, que armazena o histórico do diálogo atual. Esse histórico é essencial na resolução de referências pronominais e de dêiticos. No caso de ACIs com personalidade e/ou humor, o histórico também registra os estados emocionais do agente.

Essas informações são usadas como contexto na escolha dos pares adjacentes que formam o diálogo, incluindo também a escolha dos modos de comunicação a serem utilizados pelo agente (e.g., os cumprimentos verbais podem ser acompanhados ou substituídos por acenos, uma vez que estamos tratando com agentes multimodais). Esse módulo também deve gerenciar a tomada de turno, decidindo quando o ACI mantém o turno e quando passa o turno para o usuário. O ideal é que exista um equilíbrio entre os turnos do usuário e os turnos do ACI, de modo a simular um diálogo natural.

É possível identificarmos dois tipos de diálogo com agentes conversacionais (e até mesmo com humanos): (1) os diálogos livres/sociais (e.g., [Bickmore e Cassell, 2004]); e (2) os diálogos orientados a tarefa (do inglês, *task-oriented dialogues*), nos quais o agente faz o papel de um especialista que guia um humano novato na realização de uma tarefa, ou na assimilação de algum conteúdo a ser aprendido (e.g., STEVE, já mencionado anteriormente). As estratégias de controle do andamento diálogo devem levar em conta o tipo de diálogo que o ACI realiza.

Relembramos aqui que os diálogos são organizados em três fases: abertura, desenvolvimento e fechamento. Para ACIs voltados a diálogos livres, a fase de desenvolvimento não apresenta uma estrutura (linear ou hierárquica) rígida entre os pares adjacentes que podem surgir durante a conversação. Além disso, o tópico e os focos da conversação devem ser selecionados de comum acordo com o usuário, durante o desenrolar do diálogo.

Já os diálogos com ACIs orientados a tarefa apresentam uma estrutura mais rígida, uma vez que têm um objetivo a cumprir. Esses ACIs deverão manter também um *Modelo da Tarefa*, com seus objetivos (implícitos ou explícitos). Nesse caso, o desenrolar do diálogo deve ser estruturado de acordo com as subfases da tarefa a ser realizar (ou nas subfases de apresentação do conteúdo a ser estudado pelo aluno, no caso de ACIs para Educação). Na fase de fechamento, o agente deve verificar se o seu objetivo foi alcançado (com a realização da tarefa por parte do usuário). Tais agentes são mais fortemente baseados em conceitos da Análise do Discurso (Seção 4.3.1).

### **Modelo do Usuário**

Modelos do Usuário (MU) são uma representação do conhecimento sobre o usuário que o sistema armazena [Kobsa e Pohl 1995]. Esse modelo consiste em uma coleção de dados sobre um usuário específico, sendo o ponto de partida para quaisquer adaptações do comportamento do sistema a esse usuário. Assim, o MU é mais uma base de conhecimento que pode ser mantida pelos ACIs.

Podemos classificar os MU, em três dimensões, considerando: (1) a frequência de atualização das suas informações; (2) a forma de coleta de informações sobre o usuário; e (3) o tipo de raciocínio utilizado para inferir conhecimento sobre o usuário.

No que concerne a dimensão (1), os MUs podem ser *estáticos* ou *dinâmicos*. No primeiro caso, as informações são geralmente coletadas no momento do cadastramento do usuário, quando ele responde a um questionário provendo dados pessoais, como nome, idade, interesses/preferências e habilidades. Essas informações não serão modificadas ao longo da interação do usuário com o sistema. Já os modelos dinâmicos são atualizados à medida que o usuário interage com o sistema. A informação estática do MU é comumente chamada de *perfil do usuário*.

A coleta de informações (dimensão 2) pode ser realizada de forma *explícita* ou *implícita*. No primeiro caso, o sistema faz perguntas periódicas ao usuário, a fim de diagnosticar suas preferências, seus objetivos ou seu conhecimento sobre algum tema (no caso de sistemas de aprendizagem). Já no caso da coleta implícita, o sistema observa a interação do usuário, e coleta as informações relevantes, como, por exemplo, tarefas realizadas, navegação, entre outras. As informações a serem coletadas são especificadas pelo projetista do sistema.

Por fim, vejamos a dimensão (3). O tipo de raciocínio utilizado para inferir conhecimento sobre o usuário é, na maioria dos casos, baseado em regras. Contudo, já encontramos atualmente modelos mais sofisticados, baseados em Aprendizagem de Máquina. Em ambos os casos, o objetivo dessa inferência/aprendizagem é prover adaptação à medida que o usuário interage com o sistema.

No entanto, pode ser difícil coletar informações sobre o usuário. Nesses casos, quando há informações disponíveis sobre um grupo de usuários, podemos utilizar modelos baseados em estereótipos [Rich, 1989], que classificam o usuário de acordo com as características do grupo a que ele pertence. Essas informações são então usadas na adaptação do sistema. No caso dos modelos dinâmicos que coletam informação implicitamente, é possível iniciar a interação com base em um modelo baseado em estereótipo, e atualizar essas informações à medida que a interação avança, a fim de prover adaptações mais personalizadas.

### **Modelando Emoção e Personalidade - ACIs com Afetividade**

Modelos de Computação Afetiva para ACIs têm por objetivo detectar, interpretar, processar e simular emoções humanas. Aplicações de ACIs com afetividade seguem o chamado “*loop-afetivo*” [Hook, 2008]. Esse *loop* conta com três passos, que estão alinhados com o ciclo geral de percepção-raciocínio-ação dos agentes inteligentes, já discutido aqui:

- 1) Reconhecer a emoção do usuário;
- 2) Selecionar a ação mais adequada, considerando também o estado do usuário; e
- 3) Sintetizar a resposta afetiva mais apropriada.

Os modelos OCEAN e OCC já foram largamente usados na construção de ACIs [Bartneck, 2002], bem como na elaboração de modelos e arquiteturas genéricos para ACIs com afetividade (personalidade, humor e/ou emoção) – ver Seção 4.3.2. Podemos citar aqui o MCAChat, um Modelo de Computação Afetiva para a construção de Chatterbots com emoção usando o modelo OCC construído em coautoria com uma das autoras deste curso [Jacob Jr. et al., 2013].

Algumas aplicações utilizam modelos de personificação em três camadas, considerando o humor como uma característica “meio-termo” entre as emoções temporárias e a personalidade permanente do agente [Kasap e Magnenat-Thalmann, 2007]. Como exemplo, temos ALMA (A Layered Model of Affect) [Gebhard, 2005], que busca prover um *profile* de personalidade, bem como humor e emoção em tempo real usando os modelos OCEAN e OCC.

ACIs com afetividade podem ser encontrados em várias aplicações, tais como ACIs para “contação” de histórias (do inglês, *storytelling*), que interagem com o usuário, buscando envolvê-lo na história [Gelin et al., 2010]; ACIs tutores – e.g., INES [Heylen et al., 2009], ACIs assistente pessoais [Mancini et al., 2008], entre outras. Vale a pena também observar o trabalho desenvolvido pelo *Affective Computing group* do MIT Media Lab<sup>66</sup>.

### **Adaptação e Aprendizagem**

Como já mencionado aqui, a maioria dos modelos e frameworks para a construção de agentes virtuais trabalha com regras predefinidas de percepção-raciocínio-ação. Contudo, ambientes dinâmicos requerem dos agentes a capacidade de se *adaptar* a situações novas (não previstas). Adaptação pode ser obtida através de mecanismos de *aprendizagem em tempo real* (online) [Hapfelmeier et al., 2012], que diferem da aprendizagem em modo *batch*, realizada com base em exemplos previamente coletados e, em geral, etiquetados.

A aprendizagem online possibilita a indução de padrões de comportamento durante a interação entre agente e usuário. Por exemplo, o agente pode construir “regras sociais” durante uma dada interação, aprendendo a distinguir entre comportamentos “aceitáveis” e “inaceitáveis” para aquela aplicação (e.g., xingamentos, mau humor, etc.). Essas regras também podem ser usadas pelo agente para classificar as percepções oriundas do ambiente (e.g., o agente pode se negar a continuar o diálogo com um usuário que demonstrar um comportamento inaceitável, desrespeitoso).

Contudo, aprendizagem online baseada na interação com humanos é uma área de pesquisa ainda relativamente recente, e bastante árdua. [Kasap e Magnenat-Thalmann, 2007] citam algumas iniciativas nessa área:

- O trabalho do extinto *MIT Synthetic Characters Group*, que tinha por objetivo treinar um cachorro virtual usando aprendizagem por reforço [Blumberg et al., 2002].
- A arquitetura ALEC para aprendizagem emocional, que provê aprendizagem cognitiva e de emoções, além de tomada de decisão, usando principalmente redes neurais [Gadanh, 2003].
- O trabalho de [Wan e Tang, 2002], que apresenta um agente virtual (ator sintético) capaz de aprender a realizar tarefas específicas (como pular obstáculos) usando algoritmos evolucionários.

---

<sup>66</sup> <http://www.media.mit.edu/research/groups/affective-computing>

- [Gatzoulis et al. 2004] apresentam um agente virtual robô servente baseado em aprendizagem difusa por reforço.
- O trabalho de [Dinerstein e Egbert, 2005] combina aprendizagem de emoções e imitação, apresentando um agente virtual (ator sintético) que se adapta rapidamente através da interação com humanos.

Claramente, ainda há muito trabalho a ser feito nessa área, a fim de prover os ACIs com capacidade de aprendizagem online e adaptação. Esse caminho fica então indicado como uma tendência futura nas pesquisas.

### 4.5.3 Implementando ACIs

Discutiremos aqui algumas abordagens adotadas por pesquisadores da área na implementação de ACIs. Inicialmente, salientamos que, apesar de a arquitetura apresentada na Seção 4.5.2.1 ser bastante completa e geral, por simplificação, a maioria dos ACIs não implementa todos os componentes ali detalhados. [Kopp et al., 2014] argumentam que um grande número de ACIs tem sido construído com base em modelos que não apresentam processamento incremental da entrada/saída. Tais modelos integram o processamento da entrada, raciocínio (planejamento e deliberação) e geração da saída, a fim de simplificar a implementação dos ACIs.

Contudo, como veremos abaixo, já existem trabalhos que buscam sanar essa deficiência, oferecendo frameworks com processamento incremental, que possibilitam a criação de ACIs mais flexíveis e verossímeis. A seguir, veremos alguns dos trabalhos mais proeminentes na área: o trabalho de Cassell e seu grupo, seguido da arquitetura SAIBA e de algumas de suas extensões.

#### 4.5.3.1 O trabalho de Cassell e seu grupo

A arquitetura inicial proposta por [Cassell et al. 2000b] caracteriza-se por apresentar processamento sequencial e não incremental. Essa arquitetura é mais restrita do que a vista na Seção 4.5.2.1, consistindo em um módulo *gerenciador de entradas* (que fusiona áudio, voz e entradas gestuais), que aciona o *módulo de raciocínio*, responsável por gerar o comportamento do agente com base na interpretação da entrada e no *planejamento da saída* (mais detalhes abaixo). Além disso, a arquitetura conta com um *módulo reativo*, que mapeia diretamente algumas entradas em saídas, a fim de implementar reações rápidas.

A cada turno, o agente deve optar por seguir o caminho *deliberativo* (através do módulo de raciocínio) ou o *reativo* (disparando o módulo reativo). Contudo, em ambos os casos, o processamento utiliza planos autocontidos ou especificações de conjuntos completos de comportamentos de entrada e saída. Assim, o sistema produz uma lista completa de comportamentos, que podem ser executados com agendamento prévio (baseado em tempo absoluto), ou então podem ser disparados por eventos específicos. Essa lista especifica antecipadamente (de forma não incremental) como alguns comportamentos do agente devem se desenvolver ao longo do tempo.

Uma importante extensão desse trabalho foi a ferramenta BEAT (*The Behavior Expression Animation Toolkit*) [Cassell et al., 2001], para criação de agentes animados com aspecto humano a partir de texto. Segundo os autores, da mesma forma como

sistemas de texto-voz sintetizam voz a partir de texto, BEAT sintetiza/gera comportamento conversacional não verbal (i.e., gestos, olhares) em agentes incorporados a partir de texto.

A ferramenta mapeia o texto digitado pelo construtor do agente em comportamentos através de um conjunto de regras que consideram a informação linguística e contextual contida no texto para controlar os movimentos (mãos, braços, face) e a entonação do agente animado. Os comportamentos gerados são armazenados para uso futuro, quando o agente estiver de fato dialogando com algum usuário dentro de alguma aplicação.

BEAT utiliza uma linguagem tipo XML desenvolvida pelos autores para representar “funções comunicativas”<sup>67</sup>. O sistema associa ao texto de entrada (digitado pelo construtor do ACI) *tags* que informam sobre os sinais não verbais pertinentes ao texto/contexto da entrada. A saída do agente também contém *tags* especificando o tipo de animação a ser realizada, e por quanto tempo.

Vale ressaltar que nem todos os ACIs trabalham com bancos de dados de imagens pré-prontas. Existem sistemas mais sofisticados que são capazes de gerar esses elementos gráficos em tempo real, de acordo como andamento do diálogo (a parte textual) e das percepções do agente sobre os movimentos do usuário [Kasap e Magnenat-Thalmann, 2007].

O sistema foi construído para operar em tempo real, sendo modular e extensível. BEAT permite que o usuário insira regras para a criação de novos comportamentos (movimentos e olhares), possibilitando assim a criação de agentes com diferentes personalidades e humor. Apesar de ainda não utilizar processamento incremental na geração da saída (uma vez que os comportamentos são previamente gerados), essa ferramenta oferece um grande avanço na construção de ACIs.

O trabalho de Cassell e seu grupo alcançou grande sucesso, por ter qualidade e também por ter sido um dos pioneiros na área. De fato, esses trabalhos serviram como base/inspiração para a maioria dos ACIs existentes, como pode ser facilmente visto na literatura relacionada (e.g., Max [Kopp et al., 2005], [Kenny et al., 2007], dentre dezenas de outros trabalhos).

#### **4.5.3.2 Buscando Processamento Incremental**

Apesar de se observar o uso predominante da arquitetura de Cassell et al. por tantos anos, já existem atualmente modelos que oferecem processamento incremental, buscando superar esse estilo rígido de processamento.

De acordo com [Kopp et al., 2014], a primeira proposta de processamento incremental para gerar comportamento multimodal (fala e gestos sincronizados) foi o sistema ACE (*Articulated Communicator Engine*) [Kopp e Wachsmuth, 2004]. Nesse sistema, o processamento se dá conectando automaticamente *unidades de comportamento* que mantêm sua sincronização interna. Essa abordagem foi estendida pela arquitetura ASAP [van Welbergen et al., 2012] [Kopp et al., 2014], que provê

---

<sup>67</sup> Ver [http://en.wikipedia.org/wiki/Jakobson's\\_functions\\_of\\_language](http://en.wikipedia.org/wiki/Jakobson's_functions_of_language)

unidades de comportamento mais granulares, que podem ser ligadas para antecipar eventos de entrada. Voltaremos a falar nesse trabalho mais abaixo.

O trabalho de [Nijholt et al., 2008], por sua vez, propõe uma abordagem que vincula diretamente a ocorrência (*timing*) do comportamento do agente às previsões (expectativas) de ocorrência de eventos realizados pelo usuário, alcançando assim um grau mais refinado de coordenação temporal com os movimentos do usuário.

Processamento incremental também passou a ser considerado como indispensável para a modelagem de diálogos mais naturais. Por exemplo, o trabalho de [Schlangen e Skantze, 2011] descreve os mecanismos internos em ACs em termos de módulos abstratos que se comunicam através de unidades incrementais (*incremental units - IUs*), que são passadas entre os módulos. Muitas implementações desse modelo foram desenvolvidas e muitos aspectos de ACs têm sido modelados com sucesso dentro desse framework de processamento incremental. Contudo, as implementações da arquitetura UI têm sido uni modais até o momento (tipicamente, fala e texto).

#### 4.5.3.3 O Framework SAIBA

Atualmente, podemos observar uma preocupação em padronizar a geração da saída dos ACIs, de modo a permitir arquiteturas mais modularidade e maior flexibilidade do comportamento gerado. A arquitetura/framework SAIBA [Kopp et al., 2006] é o exemplo mais proeminente dessa tendência. Como podemos ver na página web do projeto<sup>68</sup>, o principal objetivo do SAIBA (*Situation, Agent, Intention, Behavior, Animation*) é oferecer um framework unificado para geração de comportamento multimodal para ACIs, de modo a facilitar a interação e o compartilhamento de recursos entre pesquisadores da área.

Esse framework especifica a geração multimodal em três etapas de processamento de mais alto nível, cada etapa podendo ser desmembrada em módulos menores:

- 1) Planejamento do conteúdo (intenção comunicativa do agente);
- 2) Planejamento do comportamento do agente (i.e., planejamento da realização multimodal do conteúdo); e
- 3) Realização dos comportamentos planejados.

Essas etapas de processamento se comunicam via representações em duas linguagens tipo XML: *Function Markup Language* (FML) e *Behavior Markup Language* (BML). Buscando neutralidade, essas linguagens são: livres de aplicação e de domínio; independentes do processador de som e do gerador de imagens; e distinguem com clareza informações de natureza distinta (informações relacionadas a funções *versus* informações relacionadas a processos de especificação do comportamento).

A informação passada da etapa (1) para a etapa (2) descreve *intenções comunicativas*<sup>69</sup> e de expressão do agente, sem ainda fazer referências ao seu comportamento físico (i.e., como ele irá realizar essas intenções). Essa informação é

---

<sup>68</sup> SAIBA - <http://www.mindmakers.org/projects/saiba/wiki>

<sup>69</sup> Pragmática – intenções comunicativas e atos de fala – [http://pt.wikipedia.org/wiki/Atos\\_de\\_fala](http://pt.wikipedia.org/wiki/Atos_de_fala)  
<http://plato.stanford.edu/entries/pragmatics/> & <http://plato.stanford.edu/entries/speech-acts/>

especificada usando a linguagem FML, que oferece uma descrição semântica dos aspectos relevantes no planejamento do comportamento verbal e não verbal do agente.

A informação oriunda da etapa (2) descreve o modo como os comportamentos multimodais devem ser realizados pela etapa final do processo de geração (etapa 3). Essa informação é representada usando a linguagem BML, que provê uma descrição geral e independente dos realizadores de som e imagem que podem ser usados pelo ACIs. Contudo, essa representação deve prover um bom nível de detalhe na descrição do comportamento do agente, desde a mera ocorrência e o *timing* relativo das ações envolvidas, até a definição detalhada de uma forma particular de comportamento.

Esse modelo, bem como as linguagens FML e BML, têm se tornado um padrão para geração de comportamento em ACIs. Contudo, segundo [Kopp et al., 2014], SAIBA foi planejado para ser um modelo de alto nível. Assim, o framework não detalha como deve ser realizada a coordenação entre entradas e saídas do agente, nem como deve ser implementado o processamento incremental.

#### **4.5.3.4 Extensões de SAIBA**

Podemos encontrar na literatura relacionada muitos trabalhos que se inspiram no framework SAIBA. Porém, não haveria tempo para discutirmos todas essas iniciativas aqui. Assim, apresentaremos brevemente duas de suas extensões mais importantes: os trabalhos de [Scherer et al., 2012] e de [Kopp et al., 2014].

Trabalhos recentes têm estendido o framework SAIBA para incluir também a representação do comportamento não verbal do usuário, com o objetivo de considerar suas intenções comunicativas na interação com o ACI. Essa tarefa requer a percepção e a integração de uma grande variedade de pistas (dicas) de comportamento, como por exemplo, risadas, palavras faladas, mudanças de entonação e gestos simultâneos.

O trabalho de [Scherer et al., 2012], na área de reconhecimento automático de comportamento, propõe a linguagem PML (*Perception Markup Language*) como um padrão para representar percepções de comportamento não verbal do usuário. Segundo os autores, PML possibilita que os agentes virtuais reajam ao comportamento do usuário de modo adequado, melhorando a fluidez das interações entre usuário e agente. PML também trabalha com probabilidades para representar incerteza. Os diferentes níveis de processamento de PML foram projetados para controlar/coordenar atualizações do estado do diálogo, planejamento de intenções e geração adequada de comportamento.

De forma geral, os pesquisadores na área de ACI estão buscando desenvolver modelos que favorecem processamento incremental do diálogo, bem como uma integração mais flexível entre percepção e ação. Dentre essas iniciativas, destaca-se ASAP (*Artificial Social Agent Platform*), um framework para construção de agentes capazes de prover respostas rápidas e adaptação a novas situações, a fim de oferecer diálogos multimodais fluentes em tempo real.

Segundo [Kopp et al., 2014], ASAP estende SAIBA para uma arquitetura completa de entrada-saída com processamento incremental. ASAP inclui uma extensão dos modelos de representação (em especial, BML), para cobrir especificação e geração incrementais; e a utilização de PML para representar a percepção do agente.

ASAP provê uma coleção de módulos de software para construção de agentes conversacionais virtuais e robôs. Estão disponíveis também ferramentas para viabilizar a construção desses agentes, como IPAACA, um *middleware* que implementa uma arquitetura para processamento incremental. Mais informações podem ser obtidas na página web do projeto<sup>70</sup>.

#### 4.5.4 Aplicações Atuais

Assim como no caso dos ACs textuais (chatbots), podemos encontrar ACIs como aplicações na Internet<sup>71</sup>, ou como agentes embutidos em sistemas variados. Contudo, as possibilidades de uso dos ACIs são bem mais amplas, devido a eles apresentarem um estilo de interação mais rico do que os chatbots textuais – ver [Nijholt, 2004].

As aplicações citadas nesta seção são tanto de ACIs mais simples quanto de humanos virtuais, uma vez que alguns autores não diferenciam esses dois tipos de agentes. Dentre as diversas aplicações de ACIs, destacam-se ACIs tutores, ACIs assistentes pessoais, ACIs para SAC, ACIs para entretenimento, terapia online, entre outros. Abaixo discutiremos suas características principais e daremos alguns exemplos de aplicações encontradas na literatura.

- *Os ACIs tutores* são companheiros virtuais em ambientes de aprendizagem [Kerry et al., 2009] [Veletsianos e Russell, 2014] ou em ambientes de treinamento [Kenny et al., 2007a]. Esses agentes são bastante utilizados em ambientes de EaD, com o objetivo de diminuir a sensação de isolamento dos alunos. Moyo [2014] mostra que ACIs conseguem promover o aprendizado dos estudantes, por causar empatia e motivá-los a permanecer no ambiente. Além disto, Tais agentes são capazes de demonstrar tarefas complexas, buscando trazer o foco de atenção do aluno para detalhes importantes da apresentação através de movimentação e gestos (como movimentar-se dentro do ambiente virtual, e apontar para informações importantes). Alguns desses agentes são capazes de demonstrar emoção. Segundo [Johnson et al., 2000], esses agentes melhoram o aprendizado dos alunos/aprendizes.
- *ACIs acompanhantes/assistentes pessoais*<sup>72</sup> têm usos muito diversos, como por exemplo, servir como acompanhantes de idosos que vivem sozinhos [Ring et al., 2013], [Garrido et al., 2013]. Nesses casos, tais agentes atuam como interfaces para os ambientes inteligentes [de Carolis et al., 2012], auxiliando os habitantes a utilizarem os recursos disponíveis. Ou, no caso de idosos, ajudando-os com suas necessidades de medicação.
- *ACIs para sistemas de terapia online* (do inglês, *e-therapy*), que, como acontece com sistemas de EaD, também possuem grandes taxas de evasão [Grolleman et al., 2006]. Um exemplo desse tipo de ACI é o do SimCoach, projeto do Institute for Creative Technologies, que provê três ACIs diferente para auxílio<sup>73</sup>.

---

<sup>70</sup> <http://asap-project.ewi.utwente.nl/>

<sup>71</sup> Contudo, não é comum encontrar ACIs para simples conversação na Web, como ELIZA e ALICE.

<sup>72</sup> Alguns autores denominam os ACIs tutores como assistentes pessoais, uma vez que essas duas aplicações apresentam muitas semelhanças.

<sup>73</sup> <http://ict.usc.edu/prototypes/simcoach/>



- *ACIs para e-fitness*, que atuam como *personal trainers* [Bickmore, 2003].
- *ACIs comentaristas e apresentadores* são uma aplicação que vem emergindo nos últimos anos, e podem ser utilizados para comentar eventos em tempo real, jogos, passeios turísticos etc. Como exemplo dessa categoria de ACI temos Eric [Strauss e Kipp, 2008].
- *ACIs para Serviço de Atendimento ao Consumidor (SAC)* são um tipo de aplicação já consolidada, e que vem crescendo ao longo dos anos [Zdenek, 2007] [Kuligowska, 2015]. Como exemplos temos Anna, assistente da IKEA<sup>74</sup> e a Lu, assistente do Magazine Luiza<sup>75</sup>.
- *ACIs em Jogos Digitais* para simples entretenimento. Esses assistentes buscam aumentar o grau de realismo do jogo [Lankes e Bernhaupt, 2011] [Downs e Sundar 2011].
- *ACIs em Jogos Sérios* têm o objetivo de aumentar o interesse e o grau de imersão do jogo [Laufer e Tatai, 2004] [Rosmalen et al., 2012]. Esse tipo de ACI pode se confundir com ACIs tutores, uma vez que em ambos os casos o objetivo do ACI é auxiliar no aprendizado.
- *ACIs para “Contação” de Histórias* (do inglês, *storytelling*) não são ainda muito difundidos, porém são muito interessantes [Ryokai et al., 2003], [Gelin et al., 2010].

Assim como no caso dos chatterbots, com ELIZA e ALICE, também existem ACIs que ganharam destaque, como STEVE e REA. Um ACI mais atual que também se destacou foi GRETA, que é capaz de se comunicar em 6 idiomas distintos<sup>76</sup>. GRETA foi desenvolvida dentro do projeto europeu SEMAINE<sup>77</sup>, que deu origem a alguns ACIs muito interessantes, como POPPY e SPIKE, que variam seu comportamento e seu estado emocional durante o desenrolar da conversação. Mais exemplos de ACIs podem ser encontrados no site chatbots.org, em constante atualização.

Voltando a atenção novamente para Computação Afetiva, destacamos os ACIs tutores com afetividade, que buscam reconhecer as emoções dos estudantes a fim de promover um aprendizado mais efetivo. Um exemplo desse tipo de aplicação é o agente PrimeClimb [Davoodi et al., 2013], que usa o modelo OCC para tentar reconhecer o motivo da emoção do usuário (se ela é devida a algo que o estudante realizou –e.g., orgulho ou vergonha, – ou devida ao comportamento do estudante –e.g., admiração ou reprovação).

Quando o agente é capaz de detectar corretamente a emoção do usuário, ele pode adequar sua resposta, auxiliando aquele estudante de forma mais efetiva. Mais um exemplo de ACI que utiliza o modelo OCC pode ser visto no trabalho de [Zong et al., 2000]. Nesse trabalho, o ACI expressa emoção por meio de diferentes ações (gestos faciais e corporais), e modificando parâmetros do discurso, como volume, velocidade e ênfase em certas palavras.

<sup>74</sup> [http://www.ikea.com/ms/en\\_US/customer\\_service/contact\\_us/contact.html](http://www.ikea.com/ms/en_US/customer_service/contact_us/contact.html)

<sup>75</sup> <http://www.magazineluiza.com.br/portaldalu/>

<sup>76</sup> <http://perso.telecom-paristech.fr/~pelachau/Greta/>

<sup>77</sup> <http://www.semaine-project.eu/>

#### 4.5.5 Limitações e Tendências Futuras

Devido à grande quantidade e à variedade de implementações de ACIs, fica difícil apontar as limitações desses agentes de forma geral (como foi feito no caso dos chatterbots textuais). Aqui, cada aspecto do ACI tem suas limitações próprias (e.g., limitações na percepção visual do agente, na interpretação/síntese de voz, na construção dos diálogos, etc.).

Assim, cada aspecto interacional do ACI deve ser inicialmente investigado separadamente: sensores e softwares (que podem limitar a percepção do agente); atuadores e processadores de voz e imagem (que podem degradar a qualidade da saída gerada pelo agente - suas ações), e módulos de gerenciamento do diálogo (que nem sempre conseguem manter uma conversação plausível e natural com o interlocutor). Contudo, vale notar que as limitações existentes podem afetar uma aplicação e não afetar outra, tudo depende do foco da aplicação.

Nesse contexto, citamos a dissertação de mestrado de [Jonkman, 2012], que discute as limitações do comportamento não verbal de um agente para contação de histórias. Já [Bartneck e Lyons, 2009] discutem as limitações na modelagem de expressões faciais em geral. Encontramos ainda muitos artigos que apontam as limitações dos reconhecedores de voz como responsáveis pelo nem sempre satisfatório desempenho dos ACIs (e.g., [López-Cózar et al., 2011]).

Existem também as dificuldades em integrar todas as modalidades da comunicação, coordenando satisfatoriamente gestos, olhares e fala a fim de prover uma conversação natural [Kopp et al., 2014]. Em relação às técnicas utilizadas, nota-se que ainda há muito trabalho a ser feito em médio e longo prazo, especialmente no que se refere à aprendizagem online e à adaptação dos ACIs.

Apesar das limitações atuais, podemos identificar um uso crescente de ACIs em aplicações diversas, como visto na Seção 4.5.4. Destaca-se aqui seu uso em interfaces do usuário, principalmente nas aplicações que requerem comunicação multimodal em tempo real [André e Pelachaud, 2010]. Essa tendência é justificada por esses agentes oferecerem aos usuários um estilo de interação natural (via fala e gestos). Além disso, ACIs podem também demonstrar emoção, tornando a interação mais verossímil e menos “anônima” para o usuário. Agentes bem projetados tornam a interação com os computadores mais agradável, fomentando bons sentimentos nos seus usuários, dando um *feedback* positivo de que o usuário é capaz de interagir com um computador. E nada pode ser mais simples do que interagir com uma máquina via fala e gestos, como é a proposta dos ACIs.

Outro aspecto muito relevante é o *comportamento social* dos ACIs. Acreditamos que os ACIs do futuro deverão possuir características colaborativas, sendo capazes de interagir com outros atores, levando em consideração sinais como empatia, emoções, papéis dos usuários e relações sociais. Também é uma tendência futura a participação dos ACIs em grupos de usuários, a fim de melhorar os relacionamentos de longo prazo. Por fim, destacamos a personificação de tais agentes como característica central para o seu sucesso.

## 4.6. Conclusão

Este documento apresentou um panorama abrangente sobre os Agentes Inteligentes Conversacionais, com destaque para os ACs incorporados. Discutimos *background*, histórico, modelos, implementações e aplicações de tais agentes. Apresentamos ainda as limitações desses agentes, bem como as tendências futuras na área.

Encontramos atualmente diversos grupos de pesquisa com projetos relacionados ao desenvolvimento de agentes conversacionais e/ou de humanos virtuais, contribuindo de forma consistente para o crescimento da área. Muito provavelmente, os principais trabalhos futuros nessa área terão sua semente nesses grupos.

O site [Chatbots.org, 2016] afirma que foram identificadas 14.052 universidades e instituições que desenvolvem algum trabalho relacionado à área de conversação com Inteligência Artificial. Uma lista de instituições acadêmicas pode ser vista nesse site<sup>78</sup>. Abaixo, destacamos algumas instituições indicadas no site, e mais outras que julgamos também relevantes para essa área:

- ArticLab - Carnegie Mellon University<sup>79</sup> (Estados Unidos)
- Media Lab - Massachusetts Institute of Technology<sup>80</sup> (Estados Unidos)
- Institute for Creative Technology - University of Southern California<sup>81</sup> (Estados Unidos)
- TELECOM ParisTech<sup>82</sup> (França)
- Human-Centered Multimedia - Augsburg University<sup>83</sup> (Alemanha)
- Human Media Interaction - University of Twente<sup>84</sup> (Holanda)
- DFKI (Deutsche Forschungszentrum für Künstliche Intelligenz) – Centro de pesquisa alemão para inteligência artificial<sup>85</sup> (Alemanha)

Além das instituições acadêmicas, encontramos também empresas de softwares desenvolvendo produtos com ACs. O site [Chatbots.org, 2016] apresenta uma lista de empresas que atuam nessa área<sup>86</sup>. Dentre elas, destacamos *Artificial Solutions*<sup>87</sup>, que alcançou sucesso produzindo ACs assistentes virtuais para empresas variadas. Mais exemplos de ACs comerciais, com indicação das empresas desenvolvedoras, podem ser vistos no artigo de [Kuligowska, 2015].

Assim, acreditamos que o futuro aponta para um uso sempre crescente de agentes conversacionais em sistemas com diversas finalidades (design, educação, etc) e sobre plataformas variadas.

---

<sup>78</sup> <https://www.chatbots.org/universities/>

<sup>79</sup> <http://www.articulab.justinecassell.com/index.html>

<sup>80</sup> <http://media.mit.edu/research/groups-projects>

<sup>81</sup> <http://ict.usc.edu/research/>

<sup>82</sup> <http://perso.telecom-paristech.fr/~pelachau/site/projects.html#>

<sup>83</sup> <http://www.informatik.uni-augsburg.de/en/chairs/hcm/>

<sup>84</sup> <http://hmi.ewi.utwente.nl/>

<sup>85</sup> <http://www.dfki.de/web/research>

<sup>86</sup> <https://www.chatbots.org/companies/>

<sup>87</sup> <http://www.artificial-solutions.com/natural-language-interaction-solutions/digital-employee/>

#### 4.7. Referências Bibliográficas<sup>88</sup>

- ALICE Bot - Artificial Linguistic Internet Computer Entity. *ALICE A.I. Foundation* (2016). Online em <http://www.alicebot.org/>.
- Allen, J.F. (1995). *Natural Language Understanding*. The Benjamin Cummings Pub. Company, New York.
- Allport, G. W. (1938). *Personality: a psychological interpretation*. Henry Holt and Company, New York.
- Ancona, D., Mascardi, V. Hübner, J.F. e Bordini, R. (2004). "Coo-AgentSpeak: Cooperation in AgentSpeak through Plan Exchange". In: *Proc. of AAMAS 2004 (Int. Conf. on Autonomous Agents and Multiagent Systems)*, ACM press, pp.698-705.
- André, E. e Pelachaud, C. (2010). "Interacting with embodied conversational agents". In: *Speech technology*. Springer US, pp. 123-149.
- Barros, F.A. e Robin, J. (2001). "Processamento de Linguagem Natural". *Revista Eletrônica de Iniciação Científica*, v.1, pp. 1 – 61.
- Bartneck, C. (2002). "Integrating the OCC model of emotions in embodied characters". In: *Workshop on Virtual Conversational Characters: Applications, Methods and Research Challenges*, Melbourne.
- Bartneck, C. e Lyons, M. J. (2009). "Facial Expression Analysis, Modelling and Synthesis: Overcoming the Limitations of Artificial Intelligence with the Art of the Soluble". In: *Handbook of Research on Synthetic Emotions and Sociable Robotics: New Applications in Affective Computing and Artificial Intelligence*. IGI Global., pp. 33-53.
- Bates, J. (1994). "The role of emotion in believable agents". *Communications of the ACM*, v. 37, n.7, pp. 122-125.
- Bickmore, T. (2003). "Relational Agents: Effecting Change through Human-Computer Relationships". Tese de Doutorado. Media Arts & Sciences, Massachusetts Institute of Technology. <http://affect.media.mit.edu/pdfs/03.bickmore-phd.pdf>
- Bickmore, T. e Cassell, J. (2004) "Social dialogue with embodied conversational agents." *Advances in Natural Multimodal Dialogue Systems*. Springer Netherlands, pp. 23-54.
- Blumberg, B., Downie, M., Ivanov, Y., Berlin, M., Johnson, M. e Tomlinson, B. (2002). "Integrated learning for interactive synthetic characters". In: *Proceedings of the ACM SIGGRAPH*.
- Bolt, R. A. (1980) "Put-that-there": Voice and gesture at the graphics interface, v.14, n.3, *ACM*.
- Bratman, M. (1987). *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.

---

<sup>88</sup> Todos os links da Web indicados neste documento estavam ativos na data de finalização deste documento.

- Brooks, R.A. (2001). "How to build complete creatures rather than isolated cognitive simulators". In: *Architectures for Intelligence*, Lawrence Erlbaum Associates, K. VanLehn, ed., pp. 225–239.
- Calvo, R. A., D'Mello, S., Gratch, J. e Kappas, A. (Eds.) (2015). *The Oxford Handbook of Affective Computing*. Oxford University Press.
- Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S. e Stone, M. (1994). "Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents". In: *Computer Graphics, ACM SIGGRAPH*, pp. 413–420.
- Cassell, J., Bickmore, T., Billingham, M., Campbell, L., Chang, K., Vilhjálmsón, H. e Yan, H. (1999). "Embodiment in Conversational Interfaces: Rea". In: *Proceedings of the CHI'99 Conference*, pp. 520-527.
- Cassell, J. e Vilhjálmsón, H. (1999). "Fully Embodied Conversational Avatars: Making Communicative Behaviors Autonomous". *Autonomous Agents and Multi-Agent Systems*, v.2, n.1 pp. 45-64.
- Cassell, J., Sullivan, J., Prevost, S. e Churchill, E. (eds) (2000). *Embodied Conversational Agents*, MIT Press.
- Cassell, J., Bickmore, T., Vilhjálmsón, H., e Yan, H. (2000a). "More than Just a Pretty Face: Affordances of Embodiment". In: *Proceedings of the Conference on Intelligent User Interfaces*, pp. 52-59.
- Cassell, J., Bickmore, T., Campbell, L., Vilhjálmsón, H., e Yan, H. (2000b). "Human Conversation as a System Framework: Designing Embodied Conversational Agents". In: [Cassel et al. 2000].
- Cassel, J., Vilhjálmsón, H., e Bickmore, T. (2001). "BEAT: the Behavior Expression Animation Toolkit". In: *Proceedings of SIGGRAPH '01*, pp. 477–486.
- Chatbots.org (2016). *Virtual Agents / Chatbots Directory*. <https://www.chatbots.org/>.
- Chatterbot. (2016). "Wikipedia, The Free Encyclopedia". <http://en.wikipedia.org/wiki/Chatterbot>.
- Colby, K.M., Hilf, F.D., Weber, S., e Kraemer, H.C. (1972). "Turing-like Indistinguishability Tests for the Validation of a Computer Simulation of Paranoid Processes". *Artificial Intelligence*, v.3, pp. 199-222.
- Davoodi, A., Kardan, S. e Conati, C. (2013). "Understanding Users' Interaction Behavior with an Intelligent Educational Game: Prime Climb." In: *AIED 2013 Workshops Proceedings Volume 2 Scaffolding in Open-Ended Learning Environments (OELEs)*. pp. 9-16.
- de Carolis, B. Ferilli, S. e Novielli, N. (2012) "Recognizing the User Social Attitude in Smart Environments", *Ambient Intelligence*, pp. 240-255.
- Dinerstein, J. e Egbert, P. K. (2005) "Fast multi-level adaptation for interactive autonomous characters". *ACM Transactions on Graphics*, v. 24 n.2, pp.262– 288.

- Downs, E. P., e Sundar, S. S. (2011). "We won" vs. "they lost": Exploring egoenhancement and self-preservation tendencies in the context of video game play. *Entertainment Computing*, v. 2, pp. 23-28.
- Egges, A, Kshirsagar, S, Magnenat-Thalmann, N (2004). "Generic personality and emotion simulation for conversational agents". *Computer Animation and Virtual Worlds*, v.15, n.1 pp. 1-13.
- Eliza Chat bot (2016). "Is she a Rogerian psychotherapist, a semi intelligent chat bot, or just a toy?" <http://nlp-addiction.com/eliza/>
- Erman, Lee D., et al. (1980) "The Hearsay-II speech-understanding system: Integrating knowledge to resolve uncertainty." *ACM Computing Surveys (CSUR)*, v.12, n.2, pp. 213-253.
- Gadanhó, S. (2003). "Learning behavior-selection by emotions and cognition in a multi-goal robot task". *Journal of Machine Learning*, v. 4, pp. 385–412.
- Galley, Michel, et al. (2003). "Discourse segmentation of multi-party conversation". In: *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*. Association for Computational Linguistics,
- Garrido, P., Sanchez, A., Martinez, F. J., Baldassarri, S., Cerezo, E. e Seron, F. J. (2013). "Using 3d virtual agents to improve the autonomy and quality of life of elderly people". In: *Ambient Intelligence-Software and Applications*. Springer International Publishing. pp. 129-136.
- Gatzoulis, C., Tang, W., Wan, T.R. (2004). "Fuzzy Reinforcement Learning for an Evolving Virtual Servant Robot". In: *IEEE International Workshop on Robot and Human Interactive Communication held at Kurashiki, Japan*. IEEE, pp. 685- 690.
- Gebhard, P. (2005). "Alma –A Layered Model of Affect". In: *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multi Agent Systems*. pp. 29–36.
- Gelin, R., d'Alessandro, C., Le, Q., Deroo, O., Doukhan, D., Martin, J. C., Pelachaud, C., Rilliard, A. e Rosset, S. (2010). "Towards a Storytelling Humanoid Robot". In: *AAAI Fall Symposium: Dialog with Robots*.
- Gratch, J., Rickel, J., André, E., Cassell, J., Petajan, E., e Badler, N. (2002). "Creating Interactive Virtual Humans: Some Assembly Required". *IEEE Intelligent Systems*, July/August, pp. 54-61.
- Green Jr, B., et al. (1961) "Baseball: an automatic question-answerer". In: *Western Joint IRE-AIEE-ACM computer conference*. ACM Press.
- Grolleman, J., van Dijk, B., Nijholt, A. e van Emst, A. (2006). "Break the habit! designing an e-therapy intervention using a virtual coach in aid of smoking cessation". In: *Persuasive Technology*. Springer Berlin Heidelberg. pp. 133-141.
- Grosz, Barbara J., e Candace L. Sidner (1986). "Attention, intentions, and the structure of discourse." *Computational linguistics*, v.12, n.3, pp. 175-204.
- Hapfelmeier, A., T. Hothorn, K. Ulm, e C. Strobl (2012). "A new variable importance measure for random forests with missing data". *Statistics and Computing*, pp. 1–14.

- Heylen, D., Theune, M., Op den Akker, R. e Nijholt, A. (2009). "Social agents: The first generations". In: 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops (ACII 2009). IEEE. pp. 1-7.
- Höök, K. (2008). "Affective loop experiences – what are they?" Lecture Notes in Computer Science, v.5033, pp. 1–12.
- Jacob Jr., A., da Mata, E., Santana, A., Francês, C., Costa, J. e Barros, F. (2013). "Adapting Chatterbots Interaction for Use in Children's Education". In: Emerging Research and Trends in Interactivity and the Human-Computer Interface. pp. 413-428.
- Johnson, W.L., Rickel, J.W., e Lester, J.C. (2000) "Animated pedagogical agents: Face-to face interaction in interactive learning environments". International Journal of Artificial Intelligence in Education, v. 11, pp. 47–78.
- Jonkman, F. (2012). "Nonverbal behaviour of an embodied storyteller". Dissertação de Mestrado. University of Twente, NL. [http://essay.utwente.nl/61766/1/Master\\_thesis\\_Ferry\\_Jonkman.pdf](http://essay.utwente.nl/61766/1/Master_thesis_Ferry_Jonkman.pdf)
- Kasap, Z. e Magnenat-Thalmann, N. (2007). "Intelligent virtual humans with autonomy and personality: State-of-the-art". Intelligent Decision Technologies. vol 11, n.2. pp 3-15.
- Kenny, P., Hartholt, A., Gratch, J., Traum, D., Marsella, S. e Swartout, B. (2007). "The More the Merrier: Multi-Party Negotiations with Virtual Humans". In: Proc. of the National Conference on Artificial Intelligence (AAAI 2007), pp. 1970-1971.
- Kenny, P., Hartholt, A., Gratch, J., Swartout, W., Traum, D., Marsella, S. e Piepol, D. (2007). "Building interactive virtual humans for training environments". In: Proceedings of I/ITSEC. Vol. 174.
- Kerry, A., Ellis, R. & Bull, S. (2009). "Conversational agents in E-Learning". In: Applications and Innovations in Intelligent Systems XVI. Springer, pp. 169-182
- Kobsa, A. e Pohl, W. (1995) "The user modeling shell system BGP-MS" User Modeling and User-Adapted Interaction. v.4, n.2. pp 59–106
- Kopp, S., Gesellensetter, L., Krämer, N., e Wachsmuth, I. (2005), "A conversational agent as museum guide – design and evaluation of a real-world application". In: Panayiotopoulos et al. (eds.): Intelligent Virtual Agents Berlin: Springer-Verlag. pp. 329-343.
- Kopp, S., e Wachsmuth, I. (2004). "Synthesizing multimodal utterances for conversational agents". Journal of Computer Animation and Virtual Worlds, v.15, pp. 39-52.
- Kopp, S., van Welbergen, H., Yaghoubzadeh, R. e Buschmeier, H. (2014). "An architecture for fluid real-time conversational agents: integrating incremental output generation and input processing". Journal on Multimodal User Interfaces, v.8, n.1, pp. 97-108.

- Kuligowska, C. (2015). "Commercial Chatbot: Performance Evaluation, Usability Metrics and Quality Standards of Embodied Conversational Agents". In: Professionals Center for Business Research, Vol 2.
- Lankes, M. e Bernhaupt, R. (2011). "Using embodied conversational agents in video games to investigate emotional facial expressions". In: Entertainment Computing. v.2, n.1, pp. 29-37.
- Laufer, L. e Tatai, G. (2004). "Learn, Chat and Play—An ECA supported stock markets elearning curricula". In: Proceedings of the IASTED International Conference on Web-Based Education (WBE 2004). ACTA Press, Innsbruck. pp. 16-18.
- Laven, S. (2016). Chatterbot Central - *The Simon Laven Page*. <http://www.simonlaven.com/>
- López-Cózar, R., Callejas, Z., Espejo, G. e Griol, D. (2011). "Enhancement of Conversational Agents by Means of Multimodal Interaction". In: Conversational Agents and Natural Language Interaction: Techniques and Effective Practices. IGI Global. pp. 223-242.
- Mancini, M., Niewiadomski, R., Bevacqua, E. e Pelachaud, C. (2008). "Greta: a SAIBA compliant ECA system". In: Workshop on Animated Conversational Agents (WACA'08). Paris, França.
- Marcuschi, L.A. (1986) *Análise da Conversação*. Editora Ática, São Paulo.
- Mateas, M. (1999) *An Oz-centric review of interactive drama and believable agents*. Springer Berlin Heidelberg.
- Mauldin, M. L. (1994). "Chatterbots, TinyMUDs, and the Turing Test: Entering the Loebner Prize Competition". In: Proceedings of Twelfth National Conference on Artificial Intelligence. pp. 16-21.
- McCrae, R. R. e Costa Jr., P. T. (1998). "A five-factor theory of personality". In: Handbook of personality: Theory and Research. Vol. 2, pp. 139-153.
- Morie, J.F., Chance, E., Haynes, K. e Rajpurohit, D. (2013). "Embodied conversational agent avatars in virtual worlds: Making today's immersive environments more responsive to participants". In: Believable bots. Springer, pp.99-118.
- Moyo, S. G. (2014). "Effective Tutoring with Empathic Embodied Conversational Agents". Tese de Doutorado. The Open University.
- Neves, A. M. M.(2005). "IAIML - Um Mecanismo para o Tratamento de Intenção em Chatterbots". Tese de Doutorado em Ciência da Computação. Universidade Federal de Pernambuco.
- Neves, A. M. M., Barros, F. A. e Hodges, C. (2006). " iAIML: a Mechanism to Treat Intentionality in AIML Chatterbots". In: Proceedings of the 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2006). pp. 225-231.
- Nijholt, A. (2004). "Where computers disappear, virtual humans appear". Computers & Graphics. v.28, n.4, pp. 467-476.
- Nijholt, A., Reidsma, D., van Welbergen, H., op den Akker, H. , e Ruttkay, Z.M. (2008). "Mutually coordinated anticipatory multimodal interaction". In Verbal and



- nonverbal features of human–human and human–machine interaction. Springer, Berlin. pp 70–89.
- Nishida, T., Nakazawa, A., Ohmoto, Y. e Mohammad, Y. (2014). "History of Conversational System Development". In: *Conversational Informatics: A Data-Intensive Approach with Emphasis on Nonverbal Communication*. Springer. p. 43-62.
- Ortony, A., Clore, G. L. & Collins, A (1988). *The Cognitive Structure of Emotions*. Cambridge University Press, 1st. ed.
- Pauchet A. e Sabouret N. (2012). "Embodied Conversational Agents and Affective Computing". In: *European Agent Systems Summer School -EASSS*, Valencia, Espanha.
- Picard, R. (2003). "Affective computing: challenges". *International Journal of Human-Computer Studies*. v.59, n.1–2, pp. 55-64.
- Pokahr, A., Braubach, L. e Lamersdorf, W. (2005) "A Flexible BDI Architecture Supporting Extensibility". In: *Proc. IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT)*, pp. 379–385,
- Rao, A. e Georgeff M. (1995). "BDI Agents, from Theory to Practice" In: *Proceedings of the First International Conference on Multiagent Systems*. pp 312-319
- Rich, E. (1989) "Stereotypes and User Modelling", In: Kobsa A. And Wahlster W. (eds.) *User Models in Dialog System*, Springer, Berlin.
- Rickel, J. e Johnson W.L. (1999). "Animated agents for procedural training in virtual reality: Perception, cognition, and motor control." *Applied Artificial Intelligence*. v.13, pp. 343-382.
- Ring, L., Barry, B., Totzke, K. e Bickmore, T. (2013). "Addressing Loneliness and Isolation in Older Adults: Proactive Affective Agents Provide Better Support". In: *Conference on Affective Computing and Intelligent Interaction (ACII)*, 2013 Humaine Association pp. 61-66.
- Russell, S. e Norvig, P. (2010). *Artificial Intelligence: A Modern Approach*. Prentice Hall. 3rd Edition.
- Ryokai, K., Vaucelle, C. e Cassell, J. (2003). "Virtual Peers as Partners in Storytelling and Literacy Learning." *Journal of Computer Assisted Learning*, v.19, n.2. pp. 195-208.
- Saucier, G. e Goldberg, L. (1998). "What is beyond the big five". *Journal of Personality*. v.66, pp. 495–524.
- Schank, R. C., e Abelson, R. (1975). *Scripts, plans, and knowledge*. New Haven, CT: Yale University.
- Schegloff, E. A. (1972) "Sequencing in conversational openings." *Directions in sociolinguistics*, pp. 346-380.
- Scherer, S., Glodek, M., Schwenker, F., Campbell, N. e Palm, G. (2012), "Spotting Laughter in naturalistic multiparty conversations: a comparison of automatic online and offline approaches using audiovisual data", *ACM Transactions on Interactive*

- Intelligent Systems: Special Issue on Affective Interaction in Natural Environments, v.2, n.1, pp. 4:1-4:31
- Schlangen D. e Skantze. . (2011) "A General, Abstract Model of Incremental Dialogue Processing". *Dialogue & Discourse*, v. , n.1. pp:83–111.
- Searle, J. R. (1976). *Expression and Meaning*. Cambridge University Press.
- Strauss, M. e Kipp, M. (2008) "ERIC: a generic rule-based framework for an affective embodied commentary agent", In: *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1*, pp. 47-54
- Kenny, P., et al. (2007). "Building Interactive Virtual Humans for Training Environments" In: *Proc. of the Simulation and Education Conference*.
- Tang, W. e Wan, T. R. (2002) "Integrated Learning for Interactive Synthetic Characters". In: *WSCG– Short Papers*, pp. 137–144
- Turing, A. (1950). "Computing machinery and intelligence". *Mind*. v. 49, pp. 433-460. <http://www.csee.umbc.edu/courses/471/papers/turing.pdf>
- van Welbergen H, Reidsma D, Kopp S (2012) , "An Incremental Multimodal Realizer for Behavior Co-Articulation and Coordination" In: *LNCS*. Nakano Y, Neff M, Paiva A, Walker M (Eds) v.7502, Springer. pp. 175-188.
- Veletsianos, G., e Russell, G. S. (2014) "Pedagogical agentes". In: *Handbook of research on educational communications and technology*. Springer, pp. 759-769.
- Walker, M. e Steve W. (1990) "Mixed initiative in dialogue: An investigation into discourse segmentation." In: *Proceedings of the 28th annual meeting on Association for Computational Linguistics*. Association for Computational Linguistics.
- Wallace, R. (2009). "The Anatomy of A.L.I.C.E.". *Parsing the Turing Test*. Springer. pp. 181-210.
- Weizenbaum, J. (1966). "ELIZA: A computer program for the study of Natural Language Communication between man and machine". *Communications of ACM*. v.9, n.1, pp. 35-36. <http://dl.acm.org/citation.cfm?doid=365153.365168>
- Wilson, I. (2000). "The artificial emotion engine, driving emotional behaviour". In *AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment*. pp. 20-22.
- Winograd, Terry (1972). "Understanding natural language." *Cognitive psychology*, v.3, n.1, pp.1-191.
- Woods, W. A. (1973) "Progress in natural language understanding: an application to lunar geology." In: *Proceedings of the June 4-8, 1973, national computer conference and exposition*. ACM..
- Wooldridge, M. (2002). *An Introduction to MultiAgent Systems*. John Wiley & Sons, Chichester, England.
- Zdenek, S. (2007). "Just Roll Your Mouse over Me: Designing Virtual Women for Customer Service on the Web". *Technical Communication Quarterly*. v.16, n.4, pp. 397-430.

Zong, Y., Dohi, H. e Ishizuka, M. (2000). "Multimodal Presentation Markup Language MPML with emotion expression functions attached". In: Proceedings of Third International Symposium on Multimedia Software Engineering. pp. 359-365