**1. Introdução**

Os seres humanos se comunicam através de sinais verbais e não verbais: postura corporal, gestos (apontando para algo, que descreve as dimensões do objeto ...), expressões faciais, olhar (fazer contato visual, olhando para baixo ou para cima, para um objeto específico), e usando entonação e prosódia, em combinação com palavras e frases. A maneira pela qual as pessoas se comunicam e, portanto, os sinais que eles empregam, é influenciado por sua personalidade, objetivos e estado afetivo e pelo contexto em que ocorre a conversa [12]. Uma pesquisa muito ativa no campo de agentes inteligentes é dedicada à construção de Agentes de conversação Corporificada (ECAs). Uma ECA é um agente incorporado em um corpo virtual que interage com outro agente de um modo semelhante à humana, e particularmente de um modo convincente. A credibilidade está principalmente relacionada com a capacidade de expressar emoções [3] e por expor uma determinada personalidade [20]. No entanto, de acordo com a literatura recente [36, 13], um agente é mais crível se ele pode se comportar de maneiras típicas de determinadas culturas, e, finalmente, se ele tem um estilo comunicativo pessoal [5, 34]. Desenvolver tal "computador conversável" que é capaz de apresentar estas dimensões adicionadas de comunicação exige passar de geração de geração de linguagem natural (NLG) a geração comportamental multimodal. Uma abordagem possível é considerar o corpo e mente como estrita e necessariamente interdependentes. Outra é vê-los principalmente como independentes um do outro. A primeira abordagem implica que o planejamento para transmitir significados é concebido tendo em conta os possíveis sinais. A segunda abordagem vê uma ECA como uma entidade constituída por uma "mente" e um "corpo". Ao nível da mente, apenas o significado de um ação comunicativa é representado, deixando para o corpo a tarefa de decidir qual sinal empregar. Neste caso, a fim de evitar redundância do sinal ou conflitos, é necessário escrever regras dependentes do corpo em um nível intermediário de planejamento de sentença. Usar essas regras tem pelo menos duas vantagens: primeiro você pode adaptar os diferentes organismos para a mesma mente (digamos, o conhecimento e capacidade de conversação de um doutor podem ser transmitidos por uma menina bonita ou por um velho homem de cabelos brancos); segundo, o mesmo organismo pode assumir diferentes comportamentos, determinada por diferenças individuais, devido à cultura, estilo de comunicação, personalidade [25], em que expressa o mesmo significado. Para construir a arquitetura do nosso ECA, no contexto do projeto da UE MagiCster, 5 adotamos esta segunda abordagem. A "mente" e "corpo" são interligados por uma linguagem baseada em XML, de modo a superar os problemas da integração e para permitir a sua independência e modularidade. Durante a conversa, a mente do agente decide o que deve ser comunicado considerando diferentes fatores que desencadeiam o objetivo de comunicar e influenciar os contextos de comunicação (estado cognitivo, emoções, o contexto, a sensibilidade do usuário, e assim por diante). Em cada instante de um dado interação comunicativa, todos esses aspectos combinam uns com os outros para determinar o que o agente vai dizer [11]. O Corpo "lê" o que a mente decide se comunicar e interpreta e renderiza ao nível da superfície, de acordo com os canais de comunicação disponíveis. Além disso, este passo pode ser influenciados por diversos fatores (personalidade, estilo, identidade social, cultura) que determinam qual a combinação de sinais verbais e não verbais é o mais apto para expressar uma meta comunicativa particular.

Para conseguir uma expressividade rica, a saída da mente do agente não pode ser apenas uma combinação de descrições simbólicas de atos comunicativos. Deveria incluir, também, uma especificação dos "significados" que o corpo terá que anexar a cada um deles. Esses significados incluem as funções comunicativas que são normalmente utilizados em diálogos humanos-humano: tópico-comentário, afetivo, meta-cognitivas, performativas, dêiticas, adjetivas e funções de relação de crença [31].

Para atingir esta granularidade em expressar comportamentos críveis, definimos um conjunto de idiomas para especificar o formato de movimentos de diálogo na abstração em diferentes níveis. Em particular, para especificar o formato da movimentação em que o diálogo deve agir como interface entre a mente do agente e seu corpo, nós projetamos uma interface mente-corpo. Esta interface toma como entrada uma especificação de um plano de discurso em uma linguagem XML (Discourse Plan Markup Language, DPML), em que apenas metas comunicativas e as relações entre estes objetivos são especificados e gera como saída uma formalização do comportamento do agente em uma nova linguagem XML chamado APML (Affective Apresentation Markup Language) capaz de expressar o conteúdo de diálogo ao nível significado. Desta forma, a tarefa de interpretar a forma de tomar cada sentido, ou uma combinação de significados, ao nível da superfície, pode ser deixada para invólucros específicos do corpo. No nível dos sinais, nós desenvolvemos uma linguagem para descrever expressões faciais. Os sinais podem ser descritos de forma recursiva (um sinal pode ser definido pela combinação de sinais já definido) ou especificando todos os parâmetros (ações faciais). Estas descrições são compreensíveis e assim, interpretadas pelo jogador facial que produz a animação do agente.

Neste capítulo, depois de descrever as principais características da arquitetura subjacente, nos apresentamos o idioma APML e como ele tem sido utilizado, no contexto do projeto MagiCster. Em particular, vamos mostrar como ele foi usado ​​em interface com um rosto realista 3D chamado Greta [24] e uma voz sintética [4]. Para ilustrar a abordagem, iremos utilizar um exemplo no domínio médico. Conclusões serão discutidas na última seção.

**2 expressando comportamentos Criveis**

Como é que a comunicação se originou? A comunicação é um meio para influenciar outros [8, 7]. Um sistema envolve em um ato de comunicação, pois tem a objetivo de influenciar a outro sistema, isto é, para fazer com que outro sistema tenha algum objetivo que não tem, ou abster-se de um objetivo que ele tem. Há muitas diferentes formas de influenciar os outros: força, sedução, agressão, sendo um exemplo a imitar, induzindo emoções. A maneira peculiar em que um sistema influencia outro sistema através da comunicação é através do fornecimento de crenças a ele. Toda vez que um sistema S (remetente) comunica algo para um sistema de A (Destinatário), S proporciona crenças sobre metas de A, e através disto pergunta a A para "adotar" a meta de S, isto é, a persegui-lo como se fosse uma meta de A. Qualquer tempo que nós nos comunicamos, nós fornecemos aos outros crenças sobre os nossos objetivos a fim de fazê-los perseguir nosso próprio objetivos. Por exemplo, quando S diz a A "Leve uma aspirina", S proporciona A com uma crença sobre o objetivo da S (a meta de fazer A tomar alguma ação, e especificamente a ação de tomar uma aspirina), a fim de ter A perseguindo a meta de S, isto é, de A tomar a aspirina. Todos os atos comunicativos de fala, como uma frase ou um discurso, mas também atos não-verbais, como um gesto, um olhar ou uma expressão facial-proporcionar crenças sobre o remetente da objetivo, isto é, sobre a ação que o remetente quer que o destinatário tome: logo qualquer ato de comunicação é uma forma de pedir a outras pessoas que tomem uma ação, mas podemos pedir-lhes para fazer diferentes tipos de ações. Com um pedido, como "Pegue uma aspirina", S pede a A para fazer alguma ação; com uma série de perguntas, como "Você tomou uma aspirina? “ S pede a A para fazer uma determinada ação, a ação de prover a S uma informação; com uma informação, como “A aspirina alivia a dor” S ainda esta pedindo a A para fazer alguma ação: a ação cognitiva de acreditar em que S está dizendo”. Uma vez que qualquer ato de comunicação fornece a o destinatário uma crença complexo sobre o objetivo do remetente, e o objetivo do remetente é que A faça algo em específico , fornecendo informações, acreditar no ato de comunicação deve também mencionar todas as crenças que especificam o objetivo do emissor. Primeiro, o objetivo de S deve ser especificado numa performativa [2, 30], isto é, um objetivo específico que reivindica alguma relação social particular entre o emissor e o destinatário ("Tome uma aspirina" pode ser uma ordem, se A é meu filho, um conselho, se ele é um amigo meu); segundo, o ato de comunicação também deve especificar qual é a ação ou informação específica solicitada, ou quais são as crenças que S quer que A acredite. Por conseguinte, a menor unidade de comunicação é um ato comunicativo, que é composta de duas partes, dois pacotes de crenças: a performativa e um conteúdo proposicional. Qualquer hora em que nós nos comunicamos, devemos ter concebido pelo menos estes dois pacotes de crenças, e na medida em que estas crenças são crenças que temos o objetivo de comunicar a um destinatário, podemos dizer que eles formam um significado.

Um significado pode ser visto, então, como um conjunto de crenças que um sistema tem como objetivo transmitir para outro sistema, ou seja, crenças S tem a meta de que A também acredite. E, claro, os significados S podem ter o objetivo de transmitir para A significados muito complexos, assim, eles podem necessitar de ser embalados em vários atos comunicativos diferentes, que compõem um ato comunicativo complexo. Isto é o que acontece quando uma frase não é o suficiente, e nós precisamos recorrer a um discurso todo, um romance, um manual, uma peça de teatro, um filme, para especificar todos os significados que queremos dizer. Mas desde que as crenças são informações simples, não padrões físicos de matéria ou energia, como S pode causar essas crenças (significados) passar de sua mente para a mente de A? Para significados assim, o imateriais devem estar ligados a estímulos perceptíveis, que chamamos de sinais comunicativos. Cada significado ou conjunto de significado deve ser ligado a um sinal particular ou conjunto de sinais, e ambos S e A devem compartilhar um sistema comum de comunicação, isto é, um sistema que indica como os significados e sinais correspondem. Portanto, toda vez que um sistema S tem de comunicar um conjunto de significados para outro sistema A, ele tem de saber, em que o sistema de comunicação assume estar compartilhado com A, os sinais específicos que correspondem aos significados a serem transmitidos.

Vamos primeiro ver as crenças que podem formar o conteúdo de um ato comunicativo. Três classes de significados podem ser distinguidas [29]:

**Informações sobre o mundo:** conforme nós nos comunicamos, nós fornecemos informações sobre eventos concretos ou abstratos, seus atores e objetos, o tempo e relações espaciais entre eles. Tais informações são fornecidas principalmente através de palavras, mas também por gestos ou olhar.

**Informações sobre a identidade:** traços fisionômicos do nosso rosto, olhos, lábios, as características acústicas da nossa voz e nossa postura muitas vezes fornecem informações sobre o nosso sexo, idade, raízes socioculturais, e personalidade. E, claro, nossas palavras podem informar o destinatário nossa auto-apresentação, isto é, a maneira que nós queremos nos apresentar.

**Informação sobre a Mente do Orador:** Enquanto mencionamos acontecimentos do mundo externo, nós também comunicamos porque falar desses eventos, o que pensam e sentem sobre eles, como pretendemos falar deles. Nós fornecemos informações sobre as crenças que estamos mencionando, nossas próprias metas sobre como falar sobre eles e as emoções que sentimos ao falar [29].

Aqui vamos nos concentrar em alguns deles, que são aplicados em nossa ECA, Greta. Informações sobre o mundo incluem:

1. Dêiticos: Para mencionar os referentes de nosso discurso, podemos apontar para eles por gestos ou olhares dêiticos;

2. Adjetival: Para fazer referência a algumas propriedades de objetos podemos usar gestos icônicos ou simbólicos e até mesmo olhar (como quando se espremer os olhos para dizer "pequeno" ou "difícil").

Dentro de informações sobre a Mente do Orador, e nomeadamente as informações sobre as crenças do falante, informamos sobre:

1. Grau de certeza: palavras como 'talvez', 'certamente'; modos verbais condicionais ou do subjuntivo; mas também franzindo a testa, o que significa "Eu sou sério em afirmando isto '; abrindo as mãos, o que significa 'Isto é auto-evidente ";

2. informações metacognitivas: ou seja, a fonte de crenças mencionados, mesmo vindo da memória, inferência ou de comunicação (nós olhamos para cima quando tentando fazer inferências, movimente os dedos ao tentar lembrar, ...)

Considerando metas nós informamos sobre:

1. Performatividade da sentença (por verbos performativos, entonação, expressão facial);

2. Distinção por comentário-tópico ou tema;

3. Relações retóricas: exemplo de classe (dizendo primeiro, segundo, terceiro ...; contagem nos dedos) mudança de tópico (expressa através da troca de postura);

4. Tomada de turnos e backchannel: levantar a mão para pedir sua vez; acenar para contar ao Interlocutor que estamos seguindo o que ela diz.

Finalmente, informamos ao nosso destinatário sobre as emoções que sentimos ao falar (por palavras, gestos afetivos, entonação, expressão facial, olhar e postura).

As emoções podem ser implicadas na comunicação em pelo menos duas formas: i) pode ser a razão que provoca comunicação: nós ativamos o objetivo de se comunicar apenas porque queremos expressar a nossa emoção; ii) podem intervir durante a nossa comunicação, como uma reação ao que o nosso interlocutor está dizendo, ou para algum pensamento que de repente vem à nossa mente, quer relacionadas com o diálogo em curso ou não.

Em ambos os casos, o desencadeamento de emoção não necessariamente implica que o Agente as sente. Existem muitas razões pelas quais podemos deixar de expressar nossa emoção, e a decisão final (consciente ou não consciente) de exibi-lo podem depender de um certo número de fatores [32, 10]. Algumas delas dizem respeito à própria natureza da emoção sentida (natureza emocional), outros a interação de vários fatores contextuais (cenário).

**3 Arquitetura MagiCster.**

Para ilustrar quais os requisitos que a CEA tem de cumprir, vamos começar ilustrando um exemplo, que será utilizado, através do todo o papel, conselhos de diálogo no domínio médico (Tabela 1) em que os movimentos do agente são denotados com Gi e o usuário com Uj.

G0: Bom dia Sr. Smith.

U1: Bom dia doutor Greta. Você viu meus testes?

G1: Sim, e eu sinto muito em dizer-lhe que você foi diagnosticado como sofrendo de angina de peito, o que parece ser leve.

U2: O que é angina?

G2: A angina é um espasmo de peito, resultante do excesso de esforço quando o coração é doente.

U3: É possível curá-lo?

G3: Sim, uma terapia medicamentosa existe. Para resolver o seu problema, você devem tomar duas drogas. O primeiro é a aspirina e o segundo um é Atenolol.

Neste diálogo, o Agente (chamada Greta) assume o papel de um médico e o interlocutor é um paciente pedindo informações sobre a sua doença. Como explicado na seção anterior, a fim de mostrar um comportamento crível, o agente tem que agir de forma coerente com o seu papel, o estado mental, objetivos, personalidade e contexto social; isso é especialmente importante em áreas de conversação delicadas, como dar conselhos médicos. Além disso, o agente tem de decidir se uma emoção é sentida, de acordo com o contexto de interação, se tem de ser transmitida ao nível do sinal.

Por exemplo, durante a conversação com o paciente (ver Exemplo 1 no Quadro), Greta irá coordenar o seu discurso com várias expressões:

• No movimento G1, ela manifesta a sua empatia com o usuário. Ela faz isso não só verbalmente ("Lamento dizer-lhe"), mas também não verbal, através da apresentação da expressão de "desculpe por". Para minimizar a gravidade da doença, Greta vai enfatizar tanto verbalmente e não verbalmente o fato de que ainda é uma forma "suave".

• No movimento G2, Greta indica o peito ao dizer 'um espasmo de peito ", enquanto, por sua vez, G3, ela olha para o usuário, enquanto dizendo "seu problema". As duas expressões são realizadas através de um olhar numa direção especial que desempenha uma função dêitica para indicar um determinado ponto no espaço.

Vamos ver agora como a arquitetura do MagiCster suporta a geração de diálogos deste tipo.

A arquitetura do sistema MagiCster é composta de dois componentes principais (uma mente e um corpo), interligados por um Plano Enricher. A mente do agente inclui um planejador de conteúdo, um gerente de diálogos e um módulo de modelagem afetiva do Agente. Este módulo é responsável pela atualização do estado mental do agente, ou seja, seus objetivos e crenças. O Corpo é uma face 3D ou avatar, com um sintetizador de voz [4] para falas animadas. Nós iremos descrever brevemente cada módulo, para concentrar a nossa descrição sobre na interface Mente-Corpo.

**O módulo de modelagem afetiva do Agente** decide se um determinado estado afetivo deverá ser ativado e com que intensidade e se a emoção sentida deve ser exibida em um determinado contexto [12].

O planejador de conteúdo é responsável pela geração do plano de discurso adequado para o contexto [9]. Neste nível, a ênfase é sobre os objetivos que o agente tem de alcançar neste pedaço da conversa. Sem informação sobre como expressá-los em termos de comportamento do agente é representado no plano do discurso. De acordo com Moore e Paris [22], é uma árvore identificada com seu nome; seus componentes principais são nós que são identificados, bem como, por um nome; nós incluem atributos obrigatórios que descrevem o objetivo comunicativo, o foco do discurso e os elementos retóricos (papel na Relação retorica (RR) do nó pai e relação retórica). O DTD DPML é a seguinte:

<!ELEMENT d-plan (node+)>

<!ATTLIST d-plan

name CDATA #REQUIRED

>

<!ELEMENT node (node\*, info\*)>

<!ATTLIST node

name CDATA #REQUIRED

goal CDATA #REQUIRED

role (root | nucleus | sat) #REQUIRED

RR CDATA #REQUIRED

focus CDATA #REQUIRED

>

Os planos do discurso são representados como estruturas baseadas em XML pelas seguintes razões. Primeiro, ela nos permite construir uma biblioteca de planos de conversação padrão no domínio médico que pode ser instanciado quando necessário, para ser utilizado em qualquer contexto de aplicação (texto, hipertexto, voz, e assim por diante). Em segundo lugar, fornece uma interface XML padrão entre os módulos geradores, para favorecer a distribuição de recursos e reutilização.

**O Gerente de Diálogo** é construído em cima da arquitetura TRINDI [18], que fornece um motor para movimentação de diálogos computacionais e de um espaço no qual informações relevantes para a seleção de efeitos e movimentos podem ser armazenados. Tais informações podem ser, por exemplo, o estado mental do agente e o plano atual. Depois que um plano tenha sido escolhido a partir da biblioteca de receitas do plano, o primeiro movimento do agente é gerado de acordo com a primeira etapa deste plano. No caso o agente está dialogando com um usuário, o diálogo inicia e o DM controla seu fluxo através da iteração dos passos seguintes, até que a conversa termina [25]:

1. A iniciativa é passada para o usuário, que pode fazer perguntas sobre qualquer um dos temas em discussão;

2. O movimento do usuário é traduzido em um ato comunicativo simbólico (através um processo de interpretação simplificada) e é passado para o DM;

3. O DM decide "o que dizer", selecionando um plano / sub-plano, conseguindo o objetivo comunicativo selecionado, para executar.

**O Plano Enricher** traduz a representação simbólica de um movimento de diálogo em especificação de comportamento de um agente no nível do significado. Um movimento de diálogo pode ser um ato comunicativo "primitivo" (por exemplo: uma "saudação", um 'obrigado', uma 'informar', um 'pedido') ou um plano mais complexo (por exemplo: "Descreva um objeto com as suas propriedades"), anotados de acordo com DPML. A escolha do ato comunicativo apropriado (um implore ao invés de uma ordem; mostrando ou não uma emoção) baseia-se no contexto de conversação [11]. Um algoritmo traduz esta árvore-estrutura baseada em DPML em outro língua baseada em XML (APML), através de um conjunto de regras de transformação que dependem na informação ligada a nós no plano do discurso: relação retórico Nome e tipo, o objetivo comunicativo, o foco do discurso e assim por diante.

**A Animação de Rosto e Corpo** interpreta o movimento de diálogo APML-marcado e decide como transmitir cada sentido (pela combinação de sinais). Como mencionado anteriormente, o corpo que utilizamos é na verdade uma combinação de um modelo de rosto em 3D compatível com o padrão MPEG-4 [24] e um sintetizador de discurso [4].

**4 Uma linguagem de marcação para Especificação de Comportamento: APML**

Agentes de conversação críveis do tipo descrito acima têm motivado uma série de linguagens de marcação usada para fornecer informações meta tais como controle e intenção. Estas linguagens diferem principalmente no nível de abstração da representação e especificação que eles fornecem. As maiorias das linguagens existentes permitem a especificação do agente, no nível de sinal, dependendo mais do tempo do tipo de "corpo" que apoia a expressão de um comportamento.

Pelas razões expostas no Sect. 2 deste capítulo, nós desenvolvemos um conjunto de linguagens baseadas em XML, que incluem primitivas de alto nível para especificar atos comportamentais, semelhantes às exercidas pelos seres humanos, a fim de expressar o comportamento do agente em diferentes níveis de abstração e controlar facilmente o comportamento de ACE de forma independente a partir do corpo. Nesta seção iremos descrever APML (Affective Presentation Markup Language), cujo objetivo consiste em especificar o comportamento do agente no nível do significado. Em particular, como o próprio nome sugere, a ênfase em APML é no aspecto afetivo da comunicação entre Agente e do Usuário. Antes de olhar para as etiquetas APML em detalhes, vamos olhar brevemente outras linguagens de marcação relevantes visando descrever e especificar comportamentos semelhantes às humanas.