

## Capítulo 2

# 2. Agentes Autónomos

A conceptualização, especificação e implementação de Agentes Autónomos é matéria constituinte de uma disciplina derivada essencialmente da Inteligência Artificial cuja investigação científica teve grandes desenvolvimentos nos últimos anos. Os agentes são aplicados nas mais diversas áreas que variam desde a interacção homem-máquina até complexos processos de controlo industrial. Devido ao elevado número de aplicações e à relativa abertura do conceito, existem diversas definições sobre o conceito de *Agente* e não existe um consenso alargado entre os autores da área sobre esta matéria. No entanto, baseando-nos no trabalho de Pattie Maes [Maes, 1996], uma primeira e simples definição de *agente* é: “*Um sistema computacional que habita um dado ambiente, sente e age nesse ambiente, e ao fazê-lo realiza um conjunto de objectivos ou tarefas para o qual foi projectado*”.

O crescimento da investigação realizada no campo dos agentes está ligado ao facto de ser uma convicção geral dos investigadores que os agentes constituem um paradigma de *software* apropriado para desenvolver aplicações para ambientes abertos, heterogéneos e distribuídos como, por exemplo, a Internet. Desta forma, o crescimento do número e dimensão de ambientes com estas características constitui uma forte motivação para a investigação em agentes. A adequação dos agentes a processos de resolução de problemas cuja perspectiva centralizada não demonstra ter capacidade de os resolver satisfatoriamente é outra razão para o crescimento da investigação realizada em agentes. Por outro lado, os sistemas compostos por múltiplos agentes (Sistemas Multi-Agente) constituem uma metáfora natural para perceber, simular ou construir um vasto conjunto de sociedades artificiais.

Neste capítulo é discutido o conceito de agente. Nesta perspectiva, inicialmente é analisada a controvérsia em torno da definição de agente e as principais definições de agente propostas por investigadores da área. São também discutidos os principais atributos que um agente deve possuir tais como a autonomia, reactividade, pró-actividade, habilidade social, mobilidade, comunicação, cooperação e aprendizagem. São ainda analisados os possíveis ambientes em que os agentes se podem incluir e os domínios que inspiraram o aparecimento da investigação na área dos agentes inteligentes. As diferentes

arquitecturas de agentes, propostas na literatura da especialidade, são finalmente analisadas e comparadas entre si.

## 2.1 Áreas Científicas que Inspiraram os Agentes

A área da investigação designada por Agentes Autónomos surgiu inspirada nas áreas científicas da Inteligência Artificial, Engenharia de Software, Sistemas Distribuídos e Redes de Computadores, Sociologia, Teoria dos Jogos e Economia. A influência destas áreas no campo dos agentes autónomos situa-se aos níveis de (figura 1):

- **Inteligência Artificial** – Micro-aspectos, como a resolução de problemas, raciocínio lógico, representação e utilização de conhecimento, planeamento, aprendizagem, etc.;
- **Engenharia de Software** – O agente como uma abstracção, programação orientada por agentes;
- **Sistemas Distribuídos e Redes de Computadores** – Arquitecturas de agentes, Sistemas Multi-Agente, comunicação e coordenação;
- **Sociologia** – Macro-aspectos como a formação de sociedades virtuais e a interacção entre agentes;
- **Teoria dos Jogos e Economia** – Negociação, resolução de conflitos, mecanismos de mercado.



Figura 1: Campos que inspiraram os Agentes e Sistemas Multi-Agente

Embora a Inteligência Artificial tenha exercido inicialmente uma influência muito forte sobre o campo dos Agentes Autónomos e Sistemas Multi-Agente, verifica-se hoje em dia que este campo já evoluiu, muito para além de poder ser considerado uma sub-área da IA.

## 2.2 Agentes

Ao longo dos últimos anos verificou-se uma grande explosão na investigação na área dos agentes inteligentes ou agentes autónomos. No entanto, a controvérsia que envolve esta área continua a ser elevada. Este facto deve-se essencialmente a:

- Inexistência de um paradigma de programação bem definido para sistemas distribuídos;
- O termo agente ser vulgarmente utilizado para descrever *software* em geral devido às definições vagas e contraditórias de que é objecto;
- O paradigma dos agentes tentar resolver o problema da assunção do “mundo fechado” na orientação a objectos;
- Ao interesse da comunicação social no assunto que resultou na extrapolação da área científica para o público em geral, sem que o seu significado fosse correctamente explicado.

Desta forma o termo agente é por vezes utilizado para descrever coisas com um nível de complexidade muito reduzido tais como um simples termostato (ou outro instrumento de controlo e/ou regulação) ou um *daemon* de *software* (como um servidor de impressão ou um servidor *http*).

Embora não exista uma definição consensual do conceito de agente, existe a noção de que a autonomia é essencial num agente. Baseados nesta noção central de autonomia e nos trabalhos de Wooldridge e Jennings [Wooldridge e Jennings, 1995] [Wooldridge, 2002], Pattie Maes [Maes, 1996] e Russel e Norvig [Russel e Norvig, 1995], iremos adoptar a seguinte definição de agente:

*“Um Agente é um sistema computacional, situado num dado ambiente, que tem a percepção desse ambiente através de sensores, tem capacidade de decisão, age de forma autónoma nesse ambiente através de actuadores, e possui capacidades de comunicação de alto-nível com outros agentes e/ou humanos, de forma a desempenhar uma dada função para a qual foi projectado.” [Reis, 2002b]*

A definição apresentada diz-nos que um agente é um sistema computacional que se encontra situado num dado ambiente. Este ambiente pode ser uma parte do mundo real (universidade, fábrica, hospital, campo de futebol, etc.), um ambiente simulado ou mesmo um computador. Os agentes mais vulgares são os agentes de *software*. No entanto, os agentes podem ter uma existência física (possuindo um corpo), designando-se nesse caso por agentes robóticos.

Independentemente do tipo de agente e ambiente, essencial na definição de agente que adoptamos, é a capacidade do mesmo se aperceber do ambiente e nele agir de forma autónoma. Desta forma, o agente deve possuir sensores e actuadores apropriados ao seu

ambiente e à execução das tarefas para as quais foi projectado. A Figura 2 apresenta um esquema típico de um agente.

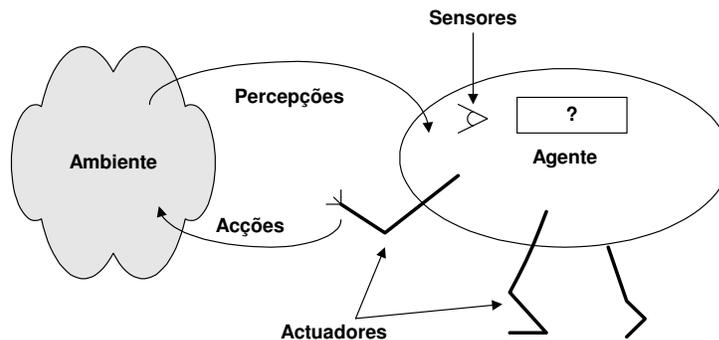


Figura 2: Esquema Típico de um Agente

Um humano apercebe-se do seu ambiente através dos seus olhos, ouvidos, do olfacto, do gosto e do tacto. Age nesse ambiente utilizando os seus actuadores: braços, pernas, cordas vocais, etc. Os sensores de um agente robótico podem incluir câmaras, microfones, sensores de proximidade (incluindo, entre outros, infravermelhos e ultra-sons), sensores de tacto e aceleração, etc. Usualmente os actuadores dos robôs são braços e pernas robóticas, motores e rodas, e altifalantes. Nos agentes de *software* é mais difícil definir o que são os sensores e actuadores do agente. Por exemplo, para um agente que joga uma partida de xadrez, os sensores permitir-lhe-ão determinar a posição das peças no tabuleiro e os actuadores serão capazes de agir no tabuleiro realizando jogadas. A definição exacta da forma de funcionamento dos sensores e actuadores do agente pode, no entanto, apresentar diversas alternativas.

### 2.2.1 Definições de Agente

Existem diversas definições de agente que, no entanto, são normalmente muito estreitas relativamente ao que definem ou, por oposição, extremamente gerais. Entre as mais conhecidas, encontram-se:

“... Algo que age ou tem o poder ou autoridade para agir ... ou representar alguém” [AHD, 2000]

“Um agente é algo que pode ser visto como tendo a percepção de um ambiente através de sensores e agindo nesse ambiente através de actuadores” [Russel e Norvig, 1995]

“Um agente é uma peça de hardware ou (mais normalmente) um sistema computacional baseado em *software* que goza das seguintes propriedades: autonomia, reactividade, pró-actividade e habilidade social” [Wooldridge e Jennings, 1995]

“... Sistemas computacionais que vivem em ambientes complexos e dinâmicos,

sentem e agem autonomamente nesse ambiente e ao fazerem-no realizam um conjunto de objectivos e tarefas para o qual foram projectados” [Maes, 1996]

“... Um agente é uma entidade persistente de *software*, dedicada a um objectivo específico. A persistência distingue os agentes de subrotinas pois estes têm as suas próprias ideias sobre como realizar tarefas e têm as suas próprias agendas. O objectivo específico distingue-os de aplicações multifunção; os agentes são tipicamente muito mais pequenos.” [Smith et al., 1994]

“Os agentes inteligentes desempenham continuamente 3 funções: percepção das condições dinâmicas do ambiente; acção para afectar as condições desse ambiente; raciocínio para interpretar as percepções, resolver problemas, realizar inferências e determinar acções.” [Hayes-Roth, 1995]

“Agentes inteligentes são entidades de *software* que realizam um conjunto de operações em nome de um dado utilizador ou de outro programa com um dado grau de independência ou autonomia, e que fazendo-o, utilizam algum conhecimento ou representações dos objectivos e desejos do utilizador.” [IBM, 1997]

“Uma entidade é um agente de *software*, unicamente se for capaz de comunicar correctamente com outros numa linguagem de comunicação de agentes (ACL)” [Genesereth e Ketchpel, 1994]

“Um agente autónomo é um sistema situado e que faz parte de um dado ambiente, que sente esse ambiente e age nele ao longo do tempo, de forma a realizar a sua própria agenda e de forma a afectar o que sentirá no futuro“ [Franklin e Graesser, 1996]

“Um Agente é um tuplo  $\langle P, A, I, in, out \rangle$  onde  $P$  é um conjunto de objectos de entrada (percepções),  $A$  é um conjunto de objectos de saída (acções),  $I$  é um conjunto de estados internos,  $in$  é uma função de  $P \times I$  para  $I$  (a função de entrada do agente) e  $out$  é uma função de  $I$  para  $A$ . O ambiente do agente é um tuplo  $\langle A, P, W, see, do \rangle$  onde  $A$  é um conjunto de operações de saída do agente,  $P$  é o conjunto de operações de entrada do agente,  $W$  é o conjunto de estados possíveis,  $see$  é uma função de  $W$  para  $P$ , e  $do$  é uma função de  $A \times W$  para  $W$ ” [VanLehn, 1991]

Nestas definições destaca-se o facto de os agentes serem capazes de se aperceberem do seu ambiente e agirem nesse mesmo ambiente [Russel e Norvig, 1995] [Maes, 1996] [Hayes-Roth, 1995] [Franklin e Graesser, 1996] [VanLehn, 1991], representarem um dado utilizador ou realizarem um dado conjunto de tarefas [AHD, 2000] [Maes, 1996] [IBM, 1997] [Franklin e Graesser, 1996], serem persistentes [Smith et al., 1994] [Franklin e Graesser, 1996], capazes de comunicar utilizando uma linguagem de alto-nível [Genesereth e Ketchpel, 1994] e autónomos [Maes, 1996] [IBM, 1997] [Franklin e Graesser, 1996].

Uma das definições de agente mais conhecidas e aceites na comunidade é apresentada por Wooldridge e Jennings [Wooldridge e Jennings, 1995] que definem um agente como uma peça de hardware ou (mais usual) um sistema computacional baseado em *software* que goza das seguintes propriedades:

- **Autonomia.** Os agentes operam sem a intervenção directa de humanos ou outros agentes e possuem algum tipo de controlo sobre as suas acções e estado interno;
- **Reactividade.** Os agentes têm a percepção do seu ambiente e respondem rapidamente às alterações que nele ocorrem;
- **Pró-Actividade.** Os agentes não se limitam a agir em resposta ao seu ambiente. Eles são capazes de tomar a iniciativa e exibir comportamento direccionado por objectivos;
- **Habilidade Social.** Os agentes são capazes de interagir com outros agentes (e possivelmente com humanos) através de uma dada linguagem de comunicação de agentes<sup>3</sup>.

Estas propriedades são mais complexas de obter do que pode parecer à primeira vista. A autonomia, embora consensual na comunidade dos Agentes e essencial na nossa definição adoptada de agente, nunca pode ser totalmente obtida. É evidente que o agente tem de ser criado e colocado em funcionamento originalmente por um humano (ou por um outro agente). A assunção de que o funcionamento do agente não terá um término também não é totalmente válida. Evidentemente que, à luz da ciência actual, o agente terá um tempo de vida limitado e um final de operação. Por outro lado, embora seja essencial para poder ser considerado um agente, que o seu funcionamento seja efectuado sem operação directa de um humano, usualmente a interacção com o humano é desejável (ou mesmo essencial como no caso da definição apresentada em [IBM, 1997]). Assim, é usual a construção de agentes que embora se possam comportar de forma autónoma, são também capazes de acatar ordens ou instruções de humanos, se tal for desejável.

Construir agentes puramente reactivos é muito simples mas não é desejável<sup>4</sup>. Um agente puramente reactivo, reagiria consecutivamente às mudanças no ambiente sem procurar atingir os seus objectivos de médio ou longo prazo, isto é, de exibir comportamento orientado por objectivos. Interessa pois definir agentes capazes de balancear o comportamento reactivo com o comportamento pró-activo. No entanto, a dificuldade em

---

<sup>3</sup> Estas linguagens são vulgarmente designadas por *ACL – Agent Communication Language*, na literatura anglo-saxónica.

<sup>4</sup> No entanto, alguns autores afirmam que os agentes não necessitam de possuir objectivos explícitos e raciocinar em termos de coordenação pois a racionalidade emerge da interacção entre os agentes reactivos [Brooks, 1991].

balancear estes dois tipos de comportamento é elevada [Hannenbauer et al., 2001] [Reis et al., 2001].

A pró-actividade, isto é, a obtenção de comportamentos orientados por objectivos, é simples de conseguir em sistemas funcionais. No entanto, esta simplicidade só se verifica se for considerado que o ambiente é estático, isto é, não muda durante a execução de um dado procedimento ou função. Para além disso, o agente tem que dispor de toda a informação que necessita para executar esse procedimento ou função, sem incerteza no ambiente. No entanto, como será analisado na secção seguinte, estas assunções não são válidas para a maioria dos ambientes. Para ambientes dinâmicos e não totalmente acessíveis, os agentes têm de ser capazes de reagir às mudanças no ambiente e raciocinar se os objectivos originais ainda são válidos, em face das mudanças no ambiente enquanto executam um dado procedimento. Isto significa que os agentes têm de ser reactivos e consequentemente capazes de reagir rapidamente às mudanças no ambiente.

A capacidade social de um agente está relacionada com a sua capacidade de trocar mensagens de alto-nível (e não só *bytes* de dados sem um significado associado) e efectuar processos de interacção social com outros agentes (e/ou humanos) semelhantes aos utilizados pelos humanos no seu dia a dia. Estes processos incluem a coordenação, cooperação, competição e negociação. De forma a efectuá-los, é necessário raciocinar acerca dos objectivos dos outros agentes (eventualmente humanos) presentes no ambiente ou, pelo menos ter a noção da sua existência. É também necessário compreender que eles são, também, agentes autónomos e que não partilham necessariamente dos nossos objectivos. Desta forma pode ser necessário negociar e cooperar com os outros agentes eventualmente trocando informação e/ou bens. Por exemplo, de forma a convencer um agente a ser cooperativo pode ser necessário efectuar um pagamento ou oferecer um determinado bem ou serviço. Podem, inclusivamente, existir agentes no ambiente com objectivos totalmente opostos<sup>5</sup> e não ser possível, consequentemente, efectuar qualquer processo cooperativo que os inclua.

O balanceamento da capacidade social com as capacidades pró-activas e reactivas é também de enorme importância. Esta importância é ainda maior no caso da realização de trabalho cooperativo por um conjunto de agentes que partilham de um objectivo comum. Neste caso, cada agente necessita de balancear a reacção a eventos no ambiente com a decisão individual necessária para desempenhar tarefas individuais e com o comportamento social necessário para desempenhar tarefas colectivas [Reis et al., 2001].

---

<sup>5</sup> Um exemplo típico consiste no futebol robótico em que os agentes de uma equipa têm como objectivo marcar golos numa baliza e evitar que sejam marcados golos na outra baliza. Os agentes adversários têm exactamente os objectivos opostos. Desta forma é impossível qualquer tipo de actividade cooperativa com esses agentes sendo necessário definir formas de competição entre eles.

Para além das propriedades essenciais mencionadas, Wooldridge e Jennings definem ainda uma *noção forte de agente*, que deriva essencialmente da área da Inteligência Artificial, e envereda por uma visão antropomórfica, onde um agente é visto como uma entidade cognitiva e com consciência, que é capaz de exibir sentimentos, percepções e emoções, à semelhança dos humanos. Assim, nesta noção, as seguintes propriedades podem também ser desejáveis nos agentes [Wooldridge e Jennings, 1995]:

- **Mobilidade.** Capacidade de um agente se movimentar de um local para outro. Usualmente esta capacidade é mencionada no contexto de agentes de *software* e como tal a movimentação verifica-se no interior de uma rede de computadores;
- **Verdade.** Um agente deve sempre ser verdadeiro e não comunicar informação falsa propositadamente;
- **Benevolência.** Os agentes não devem assumir um comportamento contra-productivo e devem sempre tentar fazer aquilo que lhes é solicitado;
- **Conhecimento e Crença.** Possuir *conhecimento* consiste em, para além de possuir uma colecção de informação dinâmica, possuir também capacidade de raciocínio sobre essa informação. É necessário definir qual a melhor estratégia de raciocínio a aplicar numa determinada situação, e porquê (meta-conhecimento). Uma *crença* representa a noção actual que o agente possui sobre determinado facto. As crenças são geralmente dinâmicas, isto é, podem alterar o seu valor de verdade com o tempo.
- **Intenções e Obrigações.** *Intenções* são objectivos de longo prazo do agente. Resultam em padrões de comportamento que levam à execução de um determinado conjunto de acções individuais. As *obrigações* estão relacionadas com compromissos que o agente assumiu anteriormente. A partir do momento em que o agente expressou a sua disponibilidade para executar determinada tarefa, é responsável por realizar as acções necessárias para essa execução.
- **Racionalidade.** Um agente agirá de forma a atingir os seus objectivos e não agirá de forma a impedir que esses mesmos objectivos sejam atingidos. Em cada instante, em face do seu conhecimento e de acordo com as suas capacidades, tentará executar a melhor acção para cumprir esses objectivos;

Usualmente, para além das propriedades apresentadas, são também referidas como propriedades comuns dos agentes:

- **Inteligência.** O estado de um agente é formalizado por conhecimento (i.e. crenças, objectivos, planos e assunções) e ele interage com outros agentes utilizando uma linguagem simbólica. Possui capacidade de raciocínio abstracto e de resolução de novos problemas e adaptação a novas situações;
- **Continuidade Temporal.** O agente é um processo que é executado continuamente ao longo do tempo. Usualmente são utilizados também os termos

persistente ou “com uma vida longa” para designar a continuidade temporal dos agentes;

- **Carácter.** O agente possui uma personalidade credível e eventualmente possui também um “estado emocional”;
- **Capacidade de Aprendizagem.** Capacidades de aprendizagem que fazem com que o agente adquira novo conhecimento e altere o seu comportamento baseado na sua experiência prévia.

É visível que existem diversas definições do conceito de agente nem sempre concordantes. A autonomia é consensual como uma capacidade necessária para um agente. A reactividade, pró-actividade e exibição de comportamento social são também importantes e como tal consideradas na maioria das definições de agente. Outras capacidades como a inteligência e a racionalidade, embora importantes, estão já um pouco contidas nas anteriores. A aprendizagem, embora importante em alguns sistemas, pode não ser desejável noutros. Exemplos disto incluem sistemas em que embora a tarefa a desempenhar seja complexa, é crítico o agente não falhar. Como tal, não é particularmente útil aprender com as suas falhas pois é suposto ele não falhar o que resultaria em consequências muito graves. Por exemplo um agente concebido para efectuar a aterragem de um avião. Neste caso, pretende-se que o agente seja capaz de realizar a sua tarefa de forma autónoma e inteligente mas pode ser argumentado que não será desejável que ele possua aprendizagem pois interessa ser capaz de prever as acções que o agente irá executar na realização desta tarefa. Pelo que se conclui que será bastante mais difícil aos humanos confiarem no agente se este efectuar acções imprevisíveis.

## 2.2.2 Atributos dos Agentes

Foi discutido anteriormente que um agente pode ser definido a partir de um conjunto de características básicas. O conjunto destas características pode ainda ser utilizado como forma de agrupar os agentes em classes ou tipologias. Deve ser referido que um agente não precisa possuir todas estas características ou atributos, embora as suas capacidades estejam directamente associadas à presença delas. A escolha de quais os atributos que devem estar presentes num dado agente depende da funcionalidade que o projectista pretende dar a esse mesmo agente.

### 2.2.2.1 Autonomia

A autonomia é provavelmente a característica mais consensual na definição do conceito de agente pelos investigadores da área dos agentes inteligentes. Segundo Nwana [Nwana, 1996], autonomia refere-se ao princípio de que os agentes podem agir baseados nas suas próprias regras de decisão, sem existir a necessidade de serem guiados por humanos. Os agentes possuem estados e metas internas, agindo de forma a atingir estas metas em prol

dos seus utilizadores. Alguns autores defendem que a autonomia aumenta com o aumento da pró-actividade, i.e. da capacidade dos agentes agirem por iniciativa própria, sem a necessidade de agirem em virtude de mudanças no ambiente ou a pedido de um humano ou outro agente [Wooldridge e Jennings, 1994].

Huhns e Singh [Huhns e Singh, 1997a] classificam a autonomia dos agentes em cinco tipos distintos:

- **Autonomia Absoluta.** O agente possui controlo completo sobre as suas percepções, raciocínio e acções, e é muito pouco previsível;
- **Autonomia Social.** Um agente conhece os outros agentes presentes no sistema e é sociável, exercendo no entanto a sua autonomia em certas circunstâncias. A autonomia está em tensão natural com a coordenação ou com outras noções de alto-nível tais como compromissos;
- **Autonomia de Interface.** Em grande parte dos sistemas, onde a autonomia absoluta é impossível, a autonomia máxima possível para o agente é a autonomia respeitante à sua forma de interface com o exterior;
- **Autonomia de Execução.** Liberdade que o agente possui na execução de acções no ambiente;
- **Autonomia de Projecto.** Grau de autonomia dos projectistas do agente na sua construção. Quanto maior for a autonomia de projecto, maior poderá ser a heterogeneidade dos agentes.

De acordo com Foner [Foner, 1993], os agentes detentores de um elevado grau de autonomia podem manter a sua agenda independente das agendas dos seus utilizadores. Para tal, os agentes apresentam três requisitos básicos: capacidade de execução de acções periódicas, execução espontânea e iniciativa própria. Estes requisitos habilitam o agente autónomo a realizar acções preemptivas, independentes, que não são despoletadas por qualquer evento exterior, e que poderão eventualmente beneficiar o seu utilizador.

### 2.2.2.2 Mobilidade

A mobilidade é a capacidade de um agente se poder movimentar de um local para outro. No caso dos agentes de *software* refere-se à capacidade de estes se poderem movimentar através de uma rede de computadores. Este atributo é particularmente interessante para agentes que auxiliam os seus utilizadores na pesquisa de informações, principalmente na Internet. No entanto, este atributo pode causar sérios problemas de sobrecarga na rede, uma vez que implica a necessidade dos agentes transitarem entre diferentes máquinas através dessa mesma rede. Dependendo do balanceamento entre a dimensão do código do agente e a dimensão dos dados que este troca com o exterior, a mobilidade pode até ser vantajosa em termos de sobrecarga da rede. Suponha-se por exemplo o caso em que o agente é de dimensão reduzida (em termos do seu código) e a sua função consiste em

realizar pesquisas de informação que implicam uma análise de largas quantidades de informação até à selecção da informação mais adequada. Neste caso, seria desejável o agente movimentar-se para junto da fonte dos dados, de forma a que, durante a pesquisa, os dados atravessassem uma reduzida parcela da rede. No regresso, somente necessitaria de trazer o seu código fonte e os dados seleccionados, diminuindo claramente o tráfego total na rede.

Knapik e Johnson [Knapik e Johnson, 1998] apresentam outro problema que está relacionado com a segurança: um agente móvel pode conter defeitos sérios no seu código ou mesmo transportar um vírus. Sendo assim, a implantação de agentes móveis deve ser acompanhada de processos de autorização e validação, além da garantia de que a memória e os recursos da máquina estarão sempre protegidos. Isto implica o desenvolvimento de plataformas especiais para agentes móveis que permitam a navegação controlada destes agentes na rede.

A mobilidade pode também ser considerada numa outra perspectiva, mais abrangente, no caso dos agentes físicos (agentes que possuem um corpo). Neste caso, a mobilidade refere-se à capacidade do agente físico (robô) se deslocar fisicamente no seu ambiente, movimentando-se de uma localização para outra. Esta mobilidade física coloca um vasto conjunto de questões e desafios de investigação, incluindo: integração sensorial, percepção activa, determinação de trajectórias, navegação robusta, metodologias de localização própria e mapeamento do ambiente.

### 2.2.2.3 Reactividade

A reactividade é a capacidade de um agente reagir rapidamente a mudanças no seu ambiente. Para tal, o agente deve ser capaz de se aperceber do seu ambiente e actuar sobre ele. Este atributo está também presente directa ou implicitamente em quase todas as definições de agente [Wooldridge e Jennings, 1994], [Russel e Norvig, 1995], [Franklin e Graesser, 1996].

O termo reactividade é, no entanto, utilizado na literatura com pelo menos dois significados distintos. O primeiro refere-se à capacidade de um agente reagir às mudanças no seu ambiente rapidamente. O segundo refere-se à capacidade dos agentes decidirem as suas acções sem consultarem um modelo interno do mundo. Wooldridge e Jennings [Wooldridge e Jennings, 1995] usam o primeiro significado de reactividade e referem agentes que se apercebem do seu ambiente e respondem de forma rápida às mudanças que ocorrem nesse ambiente. Russel e Norvig [Russel e Norvig, 1995], por oposição, referem-se a reactividade quando os agentes utilizam planeamento reactivo e regras de condição-acção para definir o seu comportamento.

Embora a reactividade (no sentido de reacção atempada às mudanças do ambiente) seja desejável, um agente puramente reactivo, reage consecutivamente às mudanças no ambiente sem procurar atingir os seus objectivos de médio ou longo prazo. Se estes

objectivos não forem triviais, ou seja a tarefa a desempenhar pelo agente for complexa, será muito difícil a um agente reactivo realizá-la. De forma a que o agente possa exibir um comportamento orientado por objectivos interessa definir metodologias que permitam balancear o comportamento reactivo com as restantes capacidades do agente.

O interesse pela reactividade e pelos agentes reactivos levou ao aparecimento de uma tendência de investigação designada por “escola reactiva” [Ferber, 1999]. Os investigadores desta “escola” afirmam que não é necessário que os agentes individuais sejam deliberativos e possuam inteligência para que um Sistema Multi-Agente demonstre inteligência como um todo [Brooks, 1991] [Ferber e Drogoul, 1992] [Ferber, 1999]. Exemplos de sociedades biológicas tais como colónias de formigas foram utilizados de forma a provar que agentes reactivos muito simples podem, em conjunto, comportar-se de forma inteligente, resolvendo tarefas que podem ser classificadas como complexas.

#### **2.2.2.4 Pró-Actividade**

Este atributo pode ser também denominado de capacidade de iniciativa, uma vez que representa um comportamento independente. As acções são seleccionadas de acordo com os objectivos gerais do agente e não simplesmente devido a mudanças que ocorrem no ambiente do agente. A pró-actividade é essencial para que os agentes possam exibir comportamento orientado por objectivos, os quais podem ser gerados internamente.

#### **2.2.2.5 Comunicação**

Quando existe mais do que um agente presente no ambiente, há uma necessidade óbvia de dispor de capacidade de comunicação entre estes. Entretanto, o conceito de capacidade de comunicação não se restringe apenas à capacidade de troca de informações entre agentes. Franklin e Graesser [Franklin e Graesser, 1996] afirmam que agentes podem comunicar com outras entidades além dos agentes, incluindo humanos e o seu próprio ambiente.

De forma a dispor de capacidade de comunicação com o exterior, os agentes necessitam de dispor de actuadores apropriados ao envio dessas mensagens (placa de rede, colunas de som, etc.) e conhecimento que lhes permita utilizar protocolos de transporte dessas mensagens. Por outro lado, e a um nível de análise mais elevado, os agentes necessitam de dispor de uma linguagem que possa também ser entendida pelos outros agentes presentes no ambiente. Precisam ainda de partilhar a semântica do que é veiculado.

A disponibilidade de actuadores apropriados, protocolos e linguagem para efectuar a comunicação é essencial para dispor de outras capacidades mais elaboradas, construídas com base na capacidade de comunicação, como sejam a habilidade social e a capacidade de cooperação. Este tema será objecto de uma análise mais aprofundada na secção 9.7, em que se analisa a forma de definição de protocolos de comunicação que facilitem a cooperação de agentes.

### 2.2.2.6 Habilidade Social

A habilidade social está directamente associada com a característica de capacidade de comunicação, uma vez que representa a capacidade de interagir com outros agentes. Genesereth e Ketchpel [Genesereth e Ketchpel, 1994] afirmam que um agente tem de ser capaz de comunicar numa linguagem de alto-nível, e indicam a necessidade da existência de uma linguagem comum para a comunicação de agentes, de forma a que estes possam trocar informação. No entanto, embora a utilidade da existência desta linguagem seja reconhecida por todos os investigadores da área, o problema é definir qual será a linguagem ideal. Actualmente, não existe uma linguagem “standard” aceite mundialmente, para a comunicação entre agentes.

A habilidade social implica ainda mais do que a existência de uma linguagem comum. Os agentes devem ser capazes de partilhar a semântica das mensagens e utilizar o mesmo vocabulário (ontologia) de acordo com a aplicação em causa [Uschold e Grüninger, 1996]. Por exemplo, um “rato” terá um significado completamente diferente no contexto de uma loja de computadores ou de um jardim zoológico. Desta forma é necessário que ambos os agentes interpretem o conceito da mesma forma, ou seja, utilizem a mesma ontologia no decurso da sua comunicação.

### 2.2.2.7 Cooperação

Cooperação pode ser entendida como a capacidade que os agentes têm de trabalhar em conjunto de forma a concluírem tarefas de interesse comum. Nwana [Nwana, 1996] acredita que a cooperação entre agentes é fundamental, sendo a razão principal para a existência de um ambiente multi-agente. Tal como os humanos, os agentes têm de combinar os seus esforços de forma a atingir objectivos comuns que não podem ser atingidos individualmente.

Para permitir esta cooperação, o agente deve ser dotado de uma certa habilidade social, capacitando-o a interagir com outros agentes e possivelmente humanos, através de alguma linguagem de comunicação [Wooldridge e Jennings, 1994]. Deve ainda dispor de metodologias apropriadas para realizar essa cooperação, de acordo com a tarefa cooperativa que irá ser executada pelo conjunto de agentes. As estratégias propostas na literatura para efectuar a cooperação entre agentes incluem estratégias fixas, aprendizagem de trabalho de equipa, organizações hierárquicas, normas e leis sociais, modelização mútua, etc. Usualmente a comunicação é vital para a realização da cooperação, embora em alguns casos seja possível a cooperação sem comunicação [Genesereth et al., 1986]. Esta comunicação é usualmente realizada a um nível abstracto e elevado, utilizando linguagens e protocolos apropriados para a comunicação inter-agente.

### 2.2.2.8 Aprendizagem

Uma característica que é frequentemente confundida com a inteligência é a capacidade de aprender. Um agente só pode possuir uma autonomia completa quando possuir a capacidade de avaliar as variações de seu ambiente externo, e propor a acção mais correcta. No mundo real, o ambiente muda e não se pode esperar que, se o agente possuir continuidade temporal, o ambiente que ele vá encontrar dentro de alguns anos seja semelhante ao ambiente para o qual foi inicialmente projectado. Acresce a este facto que será muito difícil ao projectista prever, na fase de projecto do agente, a variação futura do ambiente onde este irá operar. Assim, mesmo quando um agente não reconhece a existência de uma acção com elevada probabilidade de sucesso para ser executada, é esperado que ele procure, encontre e experimente alternativas. A questão não é acertar sempre, mas aprender continuamente por experiência, seja através de sucessos ou de fracassos. No entanto, a inclusão desta capacidade de aprendizagem não é trivial em agentes. Por vezes é difícil identificar a falha ou sucesso da acção e mais complexo ainda identificar o motivo que levou a essa falha ou sucesso. Acresce a este facto que, mesmo após uma correcta identificação dos motivos de uma dada falha, por vezes é também difícil utilizar esta informação de forma a encontrar planos alternativos com maior probabilidade de sucesso.

A aprendizagem pode ser também um processo interactivo em que um treinador ensina um dado agente (ou conjunto de agentes) a executar uma determinada tarefa individual ou cooperativa. Nestes casos, o treinador pode fornecer ao agente o conhecimento, através de uma sequência de instruções ou informá-lo apenas quando o agente não possuir o conhecimento necessário para executar a sua tarefa [Lemon et al., 1996]. O treinador pode ainda possuir uma visão mais geral sobre a tarefa, e capacidade de percepção e conhecimento específico que lhe permita detectar situações de falha ou sucesso com maior facilidade do que o agente que executa a tarefa. No caso da aprendizagem multi-agente, os agentes aprendem a realizar uma tarefa que envolve mais do que um agente na sua execução. Neste caso, a figura do treinador é novamente bastante importante pois uma visão global da situação permite detectar falhas e construir planos de execução da tarefa com maior facilidade.

### 2.2.3 Agentes vs. Objectos

Um dos grandes desafios que se coloca à comunidade de investigadores em Agentes Inteligentes consiste em estabelecer claramente a distinção entre os conceitos de Agente e Objecto, mostrando a inovação que existe por detrás do conceito de Agente [Wooldridge, 2002].

O conceito de Objecto para a comunidade de Programação Orientada por Objectos, consiste numa entidade computacional que encapsula um estado e é capaz de executar um conjunto de acções (métodos) sobre este estado e comunicar através de mensagens.

Embora as semelhanças entre os conceitos sejam evidentes, as diferenças são também visíveis:

- Os Agentes possuem capacidade de decisão autónoma;
- Os Agentes são capazes de comportamento reactivo, pró-activo e social;
- Os Agentes possuem um fluxo de controlo próprio e distinto dos restantes agentes que compõe um dado Sistema Multi-Agente.

Embora os objectos possuam autonomia sobre o seu estado (dados) não possuem autonomia sobre o seu comportamento, ou seja, se disponibilizarem um determinado método, sempre que outro objecto o invoque este estará disponível. Pelo contrário, os agentes possuem controlo sobre o seu comportamento e, como tal, um outro agente terá de solicitar ao agente que execute uma dada acção. Este pode decidir, em cada situação concreta, efectuar ou não a acção solicitada. No modelo orientado a objectos a decisão é efectuada pelo objecto que solicita a execução do método. Esta distinção entre objectos e agentes é vulgarmente sumariada pelo seguinte slogan:

*“Objects do it for free; agents do it for money”<sup>6</sup>*

Relativamente à pró-actividade, reactividade e comportamento social, o modelo orientado a objectos não contém qualquer referência a estes tipos de comportamentos.

A distinção relativa ao fluxo de controlo próprio baseia-se essencialmente no fluxo de controlo único do modelo orientado a objectos “standard”. Neste modelo o fluxo único de controlo invalida a noção de entidade autónoma. No entanto, os recentes objectos activos [Booch, 1994], diminuem esta distinção. Os objectos activos são supostos terem o seu fluxo de controlo próprio sem necessariamente terem a capacidade de exhibir um comportamento autónomo.

Embora existam distinções importantes entre agentes e objectos, é claro que o modelo orientado a objectos é bastante adequado à implementação de agentes. Desta forma, linguagens tais como Java [Deitel e Deitel, 1999], o C++ [Bronson, 1997] e suas extensões, encontram-se entre as mais utilizadas na implementação de agentes autónomos. Numa extensão a este raciocínio [Shoham, 1993] os agentes podem ser vistos como um novo paradigma de programação: a Programação Orientada por Agentes.

Ao longo da história da *Engenharia de Software* as novas metodologias têm vindo a surgir no sentido de incrementar a localização e encapsulamento das unidades básicas de programação. O surgimento do paradigma da Programação Orientada a Agentes confirma

---

<sup>6</sup> Este slogan pode ser traduzido para Português como: “Os objectos fazem-no de graça; os agentes fazem-no por dinheiro.”. Outra versão do slogan [Wooldridge 2002] é: “Objects do it for free; agents do it because they want to”, ou seja, “Os objectos fazem-no de graça; os agentes fazem-no porque querem.”

esta tendência. A tabela 1 demonstra os diversos paradigmas de programação e suas características principais, ao longo dos anos.

	<i>Programação Monolítica</i>	<i>Programação Estruturada</i>	<i>Programação Orientada a Objectos</i>	<i>Programação Orientada a Agentes</i>
<b>Comportamento das Unidades</b>	Externa	Local	Local	Local
<b>Estado das Unidades</b>	Externa	Externa	Local	Local
<b>Invocação/Comportamento das Unidades</b>	Externa	Externa (chamada)	Externa (mensagem)	Local (regras e objectivos)

Tabela 1: Evolução dos Paradigmas de Programação.

A tendência verificada na evolução da engenharia de *software* (tabela 1) vai no sentido de um cada vez maior encapsulamento. As unidades de programação têm-se tornado cada vez mais autónomas e a programação orientada por agentes incrementa ainda mais esta autonomia dando às unidades de programação, agora designadas por agentes, controlo sobre o seu próprio comportamento.

## 2.3 Ambientes

Um agente é um sistema computacional que “vive” num determinado ambiente, apercebendo-se e agindo nesse ambiente. As características do ambiente em que o agente “habita” são determinantes na definição da arquitectura do agente e da sua forma de operação. Antes de projectar um agente é necessário uma análise cuidada das percepções e acções possíveis e da complexidade da tarefa a executar pelo agente. Neste contexto a distinção entre ambientes reais e simulados não implica que um ambiente real seja mais complexo do que um ambiente simulado. Por vezes, agentes de *software* construídos para ambientes simulados possuem um conjunto possível de percepções e acções mais rico do que agentes robóticos construídos para operar em determinado ambiente real.

Seguindo uma classificação baseada no trabalho de Russel e Norvig [Russel e Norvig 1995], podemos classificar as propriedades dos ambientes segundo as seguintes classes:

- **Acessível vs. Inacessível.** Um ambiente acessível é aquele onde um agente consegue obter, através dos seus sensores, informação actualizada, precisa e completa sobre o ambiente. Grande parte dos ambientes típicos não são acessíveis neste sentido. Entre os ambientes inacessíveis destacam-se a Internet e todos os ambientes físicos reais.
- **Determinístico vs. Não determinístico.** Num ambiente determinístico cada acção tem um efeito único garantido, não existindo qualquer incerteza quanto ao

resultado da sua execução.

- **Estático vs. Dinâmico.** Um ambiente estático é suposto permanecer inalterado enquanto o agente decide a próxima acção a executar. Em contraste, num ambiente dinâmico outros agentes encontram-se a agir no ambiente ao mesmo tempo. Todos os ambientes físicos do mundo real e a Internet são dinâmicos neste sentido.
- **Discreto vs. Contínuo.** Um ambiente diz-se discreto se existe um número finito de percepções e acções possíveis para o agente e contínuo caso contrário. Um ambiente pode ser contínuo no que diz respeito às percepções do agente e discreto no que diz respeito às acções (e vice versa).

Quanto mais acessível um ambiente for, mais simples será projectar e construir agentes para nele operarem<sup>7</sup>. A qualidade das decisões efectuadas por um agente está obviamente ligada à quantidade e qualidade da informação que este possui, sendo desta forma muito mais simples tomar as decisões correctas em ambientes acessíveis. É importante também referir que a escala de acessibilidade não pode ser considerada uma escala discreta (acessível / não acessível). Em ambientes não acessíveis é importante considerar o grau de acessibilidade do ambiente ao agente. Por exemplo, para um agente robótico colocado num ambiente físico, será muito mais simples tomar decisões correctas se possuir câmaras, sensores de infravermelhos e outros equipamentos sensoriais avançados, de forma a ter uma razoável acessibilidade ao estado do seu mundo. Neste contexto é também importante considerar a precisão da informação recolhida pelos sensores do agente. Os sensores do agente podem ter acesso à grande maioria da informação necessária para a operação desse agente, mas não serem capazes de obter essa informação de forma precisa. A existência de erros na percepção do agente torna as suas tarefas bastante mais complexas de executar.

O não determinismo de um ambiente é uma grande fonte de complexidade para o mesmo, impondo sérias restrições à capacidade de um agente para efectuar previsões quanto ao estado futuro do mundo. Esta característica captura no entanto características importantes do mundo real, como sejam a possibilidade de as acções executadas por um dado agente falharem ou não terem um efeito único, bem determinado e garantido à partida. Desta forma é muito mais complexo construir agentes para ambientes não determinísticos, pois é necessário providenciar capacidades de recuperação e tolerância a falhas. No entanto, mesmo ambientes determinísticos tais como os jogos de tabuleiro<sup>8</sup> podem ser complexos.

---

<sup>7</sup> É no entanto de referir que, mesmo num ambiente acessível, se a tarefa a realizar pelo agente for complexa, não será trivial projectá-lo e construí-lo.

<sup>8</sup> Exemplos típicos são os jogos como o Xadrez e o Go. Para este último jogo, devido à enorme quantidade de acções possíveis em cada instante (cerca de 200) e à elevada complexidade envolvida no raciocínio

O dinamismo de um ambiente é uma das fontes mais significativas de complexidade. Grande parte da investigação realizada no início da inteligência artificial, nomeadamente a realizada no âmbito do desenvolvimento de algoritmos de planeamento, estava preocupada em determinar a sequência de acções que permitia a um dado agente, partindo de um estado inicial especificado, atingir um dado estado objectivo também especificado [Allen et al. 1990]. O desenvolvimento deste tipo de algoritmos assumia implicitamente que o ambiente era estático, ou seja, que não seria alterado por qualquer outro agente no decurso da execução do plano [Wooldridge 2002]. No entanto, claramente que os ambientes físicos do mundo real e grande parte dos ambientes computacionais não são estáticos e este tipo de planeamento rígido não pode neles ser aplicado.

Em ambientes dinâmicos, os agentes não podem assumir que se não executarem alguma acção durante um dado intervalo de tempo, o ambiente vai permanecer inalterado. Acresce a este facto que os agentes têm inclusivamente de entrar em consideração com o tempo de raciocínio na decisão das acções a executar. O efeito de uma dada acção no instante  $t_i+t_a$  poderá ser totalmente diferente do efeito da mesma acção no instante  $t_i+t_b$ . Desta forma, supondo que  $t_i$  é o instante de tempo actual, processos de raciocínio que tenham como resultado a execução da mesma acção mas que demorem respectivamente  $t_a$  e  $t_b$  poderão conduzir a resultados totalmente distintos.

O dinamismo implica também que os agentes têm de efectuar acções apropriadas para recolherem informação do ambiente [Moore, 1990]. Neste contexto é importante o agente possuir sensores flexíveis (como por exemplo câmaras direccionáveis). Outra característica dos ambientes dinâmicos é a presença de outros agentes que agem no ambiente e que podem interferir (de forma positiva ou negativa) nas acções efectuadas pelo agente. Aqui, o agente deixa de estar isolado e surge a necessidade de se coordenar, comunicar, cooperar, competir ou negociar com estes agentes adicionais. Um ambiente estático é inerentemente muito mais simples, pois o agente pode recolher informação sobre uma dada característica do ambiente unicamente uma vez e assumir que este ambiente não será modificado, a não ser pelas suas próprias acções. Nos ambientes estáticos, os agentes não têm também necessidade de se sincronizarem ou coordenarem as suas acções com as de outros agentes no ambiente [Bond e Gasser, 1988b] nem de interagirem de qualquer outra forma com outros agentes.

A última característica dos ambientes diz respeito a estes serem discretos ou contínuos. Um ambiente discreto é aquele que só pode assumir um conjunto predefinido de estados discretos. Por oposição, num ambiente contínuo é impossível enumerar todos os estados possíveis. Desta forma, o Xadrez e outros jogos de tabuleiro, constituem ambientes discretos pois existe um número finito (embora muito grande) de estados possíveis do

---

necessário para o jogar, não foi ainda possível construir um agente capaz de vencer os melhores jogadores humanos mundiais.

tabuleiro. Ambientes físicos do mundo real são por natureza intrinsecamente contínuos. A continuidade de um ambiente pode-se verificar a três níveis:

- **Estado do Mundo (Ambiente).** O estado do mundo pode ser discreto ou contínuo. Por exemplo, no caso de um ambiente para um jogo de tabuleiro, como o Xadrez ou as Damas, o ambiente será discreto. Em diversos ambientes simulados, o estado do mundo é também discreto, no sentido em que é representado por variáveis discretas e existe um número finito de estados possíveis para esse ambiente;
- **Percepções do Agente.** Embora um dado ambiente possa ser contínuo, as percepções de um agente inserido nesse ambiente podem ser discretas. Este é o caso, por exemplo, de um robô com sensores *booleanos* de proximidade que se encontra num ambiente físico;
- **Acções do Agente.** Tal como no caso das percepções, as capacidades de acção de um agente que se encontra num dado ambiente contínuo podem ser unicamente discretas, i.e. o agente dispor de um conjunto limitado de acções.

Os ambientes discretos são bastante mais simples para os agentes por diversos motivos. O motivo principal está directamente relacionado com o facto dos sistemas computacionais serem por natureza discretos e, como tal, para tratarem ambientes contínuos, terem de efectuar algum tipo de discretização e aproximação. É este o caso, por exemplo, na percepção através de uma câmara, em que é necessário efectuar a digitalização da imagem previamente ao seu processamento através do sistema computacional. Embora a precisão na aproximação digital possa ser elevada (estando limitada unicamente pela capacidade do hardware do sistema computacional), em qualquer tipo de digitalização, existe sempre informação que é perdida ou impossível de ser tratada em tempo útil. Desta forma, num ambiente contínuo os agentes irão sempre raciocinar com informação aproximada. Acresce a este facto a complexidade da dimensão do espaço-de-estados infinito do ambiente<sup>9</sup>.

Analisando as propriedades dos ambientes é fácil concluir que os ambientes mais complexos para os agentes são os ambientes inacessíveis, não determinísticos, dinâmicos e contínuos. Ambientes com estas propriedades são por vezes designados por ambientes abertos [Hewitt, 1986]. Embora estes ambientes sejam os mais complexos, verifica-se que os ambientes físicos do mundo real são intrinsecamente abertos. Como tal, e numa perspectiva de futuro desenvolvimento de agentes para operarem no mundo real, interessa ser capaz de definir metodologias que permitam o desenvolvimento de agentes para

---

<sup>9</sup> Embora um ambiente discreto possa ter um número muito elevado e virtualmente infinito de estados possíveis.

operarem em domínios que têm a maioria ou mesmo todas as características dos ambientes abertos.

## 2.4 Classificação dos Agentes

Devido ao largo espectro de aplicações de agentes, à dificuldade em definir o que é realmente um “agente inteligente” e à enorme dinâmica que esta área de investigação tem tido ao longo dos últimos anos, existem diversos sinónimos do termo “agente inteligente” criados por diferentes investigadores na tentativa de melhor caracterizar o seu próprio trabalho. Assim, é vulgar encontrar na literatura da especialidade designações de agentes tais como: robôs, agentes de *software* (ou *softbots*), *knowbots*, *taskbots* ou *userbots*, assistentes pessoais, personagens virtuais, etc. Embora o aparecimento destes sinónimos seja perfeitamente compreensível, por vezes obscurece o conceito tornando mais difícil analisar o que é a investigação em agentes.

De forma a caracterizar melhor a área científica dos agentes, para além de analisar os domínios que inspiraram a área, é útil dividir os agentes em classes, analisando as diversas tipologias de agentes propostas na literatura. O elevado número de atributos discutido nas secções anteriores, permite perceber como é difícil implementar um agente que incorpore todos aqueles atributos. Este facto está também relacionado com o facto das características de um agente serem idealmente dependentes do tipo de aplicação a que se propõe. A análise dos atributos que estão presentes nos agentes tem sido utilizada pelos investigadores da área para classificar os agentes em tipologias. Uma tipologia é basicamente uma classificação por tipos de agentes que possuem atributos em comum.

Nwana [Nwana, 1996] propõe uma tipologia de agentes que identifica sete dimensões de classificação distintas:

- **Mobilidade.** Agentes estáticos ou móveis. Os agentes móveis, podem estar residentes na sua máquina de origem ou temporariamente numa outra máquina;
- **Modelo de Raciocínio.** Presença ou não de um modelo de raciocínio simbólico, ou seja, um agente pode ser deliberativo ou puramente reactivo;
- **Função do Agente.** A função principal assumida pelo agente, como por exemplo um agente de pesquisa de informação (que procura informação adequada a um dado utilizador na Internet) ou de interface (que facilita a interacção homem-máquina de uma determinada aplicação);
- **Autonomia.** Grau de autonomia do agente;
- **Cooperação.** Realização ou não de acções cooperativas com outros agentes;
- **Aprendizagem.** Inclusão ou não de capacidades de aprendizagem no agente;
- **Características Híbridas.** Estas combinam duas ou mais filosofias diferentes

num mesmo agente.

Combinando as características de autonomia, cooperação e aprendizagem, Nwana [Nwana, 1996] deriva quatro tipos de agentes principais: agentes colaborativos, agentes colaborativos com capacidade de aprendizagem, agentes de interface e agentes verdadeiramente inteligentes (figura 3). É importante notar que os limites desta classificação não devem ser interpretados como linhas rígidas bem definidas. Na verdade, o facto de agentes colaborativos terem mais ênfase na cooperação e autonomia que os agentes com capacidade de aprendizagem, evidentemente não exclui a possibilidade de os primeiros desenvolverem capacidades de aprendizagem.

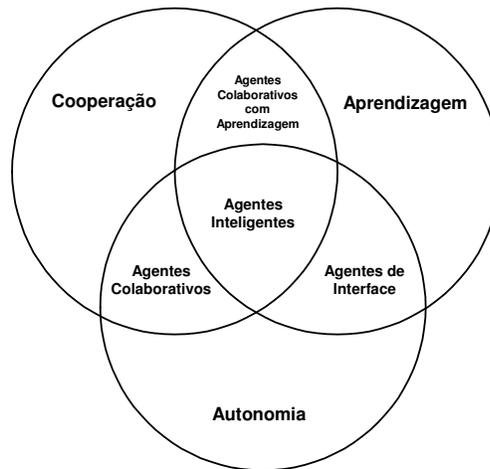


Figura 3: Categorias de Agentes definidas por Nwana.

Após estabelecer a sua tipologia, Nwana estabeleceu 7 categorias de agentes de acordo com a sua arquitectura e função: agentes colaborativos, agentes de interface, agentes móveis, agentes de informação, agentes reactivos, agentes híbridos e agentes inteligentes.

Franklin e Graesser [Franklin e Graesser, 1996] acreditam que um agente, por definição, deve ser um processo em contínua execução e apresentaram a taxonomia representada esquematicamente na figura 4, que divide os agentes autónomos em três grupos principais: agentes biológicos, agentes robóticos e agentes computacionais.



Figura 4: Categorias de Agentes definidas por Franklin e Graesser.

Outras classificações de agentes baseiam-se essencialmente nas aplicações dos agentes inteligentes. Estas incluem, entre outras, a pesquisa de informação, assistentes pessoais, gestão do correio electrónico, controle de energia eléctrica, gestão de redes de telecomunicações, gestão do trânsito, exploração submarina, controle de veículos e naves espaciais, manufactura integrada por computador, gestão de tráfego aéreo, gestão de transportes, negociação, comércio electrónico, transacções financeiras e bolsas de valores, formação de equipas, jogos, entretenimento e personagens virtuais.

## 2.5 Arquitecturas de Agentes

Uma arquitectura de *software* pode ser descrita como sendo a configuração dos componentes que constituem um sistema e das conexões que coordenam as actividades entre estes componentes [Abowd et al., 1996]. Quando se fala de arquitecturas de agentes podemos-nos referir não só à arquitectura interna de cada agente mas também à arquitectura do próprio sistema multi-agente, ou seja ao modo de organização dos agentes dentro de um sistema e como estão estruturados os seus relacionamentos e interacções.

Assim como existem diversas arquitecturas de *software*, o mesmo ocorre em relação às arquitecturas de agentes (de *software*), as quais possuem certas características que permitem a avaliação da sua qualidade e eficácia. Relativamente aos agentes robóticos, a sua arquitectura compreende não só a arquitectura de *software* como também a sua arquitectura de hardware (componentes físicos e sua interligação).

Wooldridge e Jennings [Wooldridge e Jennings, 1994] afirmam que a arquitectura de um agente pode ser estruturada através de uma metodologia específica para definir agentes, abrangendo técnicas e algoritmos para suportar essa mesma metodologia. Knapik e Johnson [Knapik e Johnson, 1998] destacam que a discussão sobre a qualidade de uma arquitectura de agentes é extremamente subjectiva, pois os detalhes desta discussão estão intimamente relacionados com aspectos específicos da aplicação de agentes que se pretende desenvolver. Uma boa arquitectura de agentes para um dado domínio poderá ser completamente desadequada para outro domínio. No entanto, Mowbray [Mowbray, 1995] estabelece alguns conceitos que podem ser úteis para o desenvolvimento de uma arquitectura promissora:

- **Simplicidade.** Idealizar a arquitectura e seus componentes de forma que sejam fáceis de entender, implementar e manter.
- **Funcionalidade.** Seleccionar uma arquitectura e ferramentas de desenvolvimento que focalizem os aspectos específicos do problema a ser abordado.
- **Expansibilidade.** A arquitectura deve poder ser facilmente ampliada, uma vez que nem todas as necessidades futuras podem ser previstas inicialmente.

- **Isolamento ou Portabilidade.** Uma arquitectura, para poder ser expandida, deve possuir uma implementação portátil, evitando-se soluções não padronizadas.

Uma classificação muito simples para as arquitecturas de agentes foi proposta por Russel e Norvig [Russel e Norvig, 1995]. Esta classificação divide os agentes nas seguintes classes:

- **Agentes Reflexos Simples:** Utilizam um conjunto de regras condição-acção preestabelecido. Perante a percepção de uma dada situação, seleccionam e executam uma dada acção pré-especificada. Os agentes deste tipo podem ser considerados os mais simples pois o seu comportamento é puramente reactivo;
- **Agentes com Representação do Estado do Mundo:** Estes agentes possuem uma representação do estado do mundo que é actualizada dinamicamente com a percepção do agente. Desta forma, a reactividade destes agentes é condicionada pelas experiências anteriores que se encontra reflectida nesse estado do mundo. Este modo de funcionamento implica que a mesma percepção ocorrida em momentos distintos e consequentemente estados diferentes do mundo, tenha como resultado acções diferentes;
- **Agentes Baseados em Objectivos:** Os agentes baseados em objectivos, para além da descrição do estado corrente do mundo, utilizam a informação sobre os objectivos que pretendem atingir. Neste caso, a decisão pode implicar a pesquisa e planeamento prévio, antes da selecção da acção a executar em cada instante. Este tipo de agentes é mais flexível pois diferentes comportamentos podem ser obtidos para o mesmo estado do mundo, dependendo dos objectivos do agente;
- **Agentes Baseados em Utilidade:** A utilidade corresponde a uma medida de satisfação do agente relativa ao cumprimento dos seus objectivos. A utilidade será mais elevada se o estado do mundo actual do agente estiver próximo de atingir os seus objectivos. As utilidades podem ser usadas para decidir entre objectivos em conflito ou ainda, quando existe incerteza nas acções, para medir a proximidade no alcance dos objectivos. Os agentes baseados em utilidade conseguem ser mais racionais pois possuem capacidades para avaliar a utilidade da execução de uma determinada acção;

A descrição que apresentam Russel e Norvig [Russel e Norvig, 1995], baseia-se na crescente complexidade da estrutura interna de cada agente. Um agente que se baseia, sucessivamente, em mais factores e capacidades para tomar uma decisão e/ou aprende com as suas interacções com o mundo exterior, pode tornar-se mais eficaz na persecução dos seus objectivos. No entanto, a divisão apresentada baseia-se essencialmente no tipo de raciocínio efectuado pelos agentes e considera unicamente agentes com arquitecturas muito simples. No seu trabalho, Russel e Norvig [Russel e Norvig, 1995] apresentam também outras arquitecturas de agentes mais avançadas em que se incluem: agentes com capacidades de aprendizagem, agentes de decisão e agentes com capacidades de planeamento.

Em oposição, Wooldridge e Jennings [Wooldridge e Jennings, 1994] apresentam uma descrição mais geral das arquitecturas de agentes. Baseiam-se na forma de construção dos agentes envolvidos para dividir as arquitecturas dos agentes computacionais em três grandes categorias:

- **Arquitecturas Deliberativas:** Seguem a abordagem clássica da Inteligência Artificial, onde os agentes actuam com pouca autonomia e possuem modelos simbólicos explícitos dos seus ambientes.
- **Arquitecturas Reactivas:** Procuram não utilizar algum tipo de modelo ou raciocínio simbólico complexo e tomar decisões “em tempo real”. Correspondem aos agentes reflexos simples na classificação de Russel e Norvig [Russel e Norvig, 1995].
- **Arquitecturas Híbridas:** Combinam as características das duas abordagens anteriores, nomeadamente as capacidades deliberativa e reactiva.

Para além destas arquitecturas gerais, Wooldridge e Jennings [Wooldridge e Jennings, 1995], apresentam ainda o conceito de arquitectura por camadas (horizontais ou verticais). Neste tipo de arquitecturas, os vários subsistemas dos agentes organizam-se em hierarquias e interagem por níveis.

Uma outra arquitectura que assume particular destaque na literatura da especialidade é a arquitectura BDI (“*Belief-Desire-Intention*”) [Rao e Georgeff, 1991]. Esta é uma arquitectura essencialmente deliberativa em que o estado interno de processamento de um agente é descrito através de um conjunto de “estados mentais”, usualmente: *crenças* (“*beliefs*”), *desejos* (“*desires*”) e *intenções* (“*intentions*”).

Alguns autores propuseram arquitecturas mais complexas que procuram incluir todas ou pelo menos a grande maioria das capacidades referidas na secção 2.1. De entre estas arquitecturas destaca-se a arquitectura de um agente social que procura fornecer ao agente capacidades deliberativas e cooperativas de forma a poder interagir com os restantes agentes presentes no sistema [Bond e Gasser, 1988b] [Moulin e Chaib-Draa, 1996].

As secções seguintes apresentam uma análise mais detalhada e comparativa das arquitecturas referidas anteriormente.

### 2.5.1 Arquitecturas Deliberativas

As *Arquitecturas Deliberativas* seguem a abordagem clássica da Inteligência Artificial, onde os agentes actuam com pouca autonomia e possuem modelos simbólicos explícitos dos seus ambientes. Estas arquitecturas interpretam em grande parte os agentes como parte de um sistema baseado em conhecimento. As decisões dos agentes (selecção de qual a próxima acção a executar) são realizadas através de raciocínio lógico. O agente possui

uma representação interna do mundo, e um estado mental explícito que pode ser modificado por alguma forma de raciocínio simbólico (figura 5).

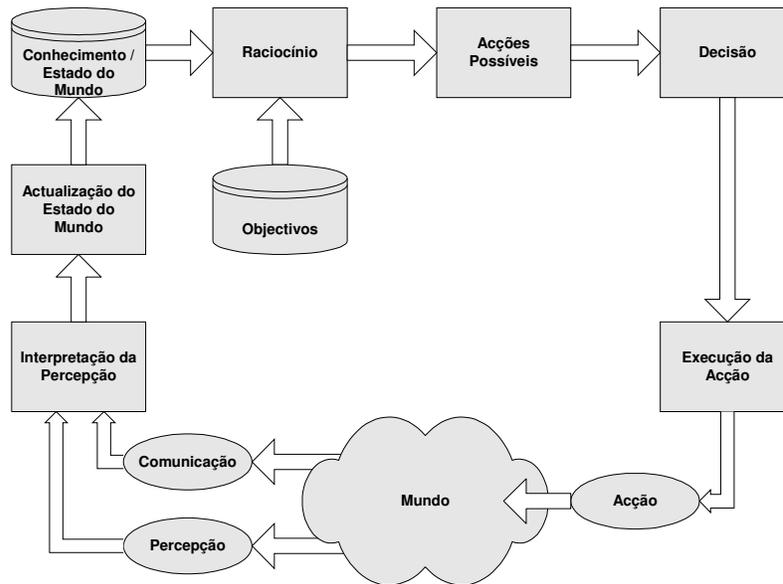


Figura 5: Esquema genérico de uma Arquitetura Deliberativa

Na construção de agentes seguindo a arquitectura deliberativa, colocam-se dois importantes problemas: como traduzir o mundo real para uma descrição simbólica e utilizar a percepção para manter essa estrutura simbólica actualizada? Como raciocinar utilizando essa informação simbólica de forma a decidir as acções a executar em cada instante?

A figura 5 apresenta um esquema típico de uma arquitectura deliberativa. Após a interpretação da percepção proveniente do ambiente, o agente utiliza esta informação para manter actualizada uma representação interna, usualmente simbólica, do estado do mundo. Este estado do mundo em conjunto com os objectivos do agente são utilizados como forma de gerar as possíveis acções a executar pelo agente e seleccionar as mais apropriadas a serem executadas por esse mesmo agente.

## 2.5.2 Arquitecturas Reactivas

As *Arquitecturas Reactivas* procuram não utilizar um tipo de modelo ou raciocínio simbólico complexo. Este tipo de arquitectura baseia-se na proposta de que um agente pode desenvolver inteligência a partir de interações com o seu ambiente, não necessitando de um modelo pré-estabelecido [Brooks, 1991]. O agente reactivo habitualmente toma as suas decisões “em tempo real”, com base num conjunto de informação muito limitado, e regras simples de situação/acção que permitem seleccionar um dado comportamento. A informação proveniente dos sensores é usualmente utilizada

directamente no processo de decisão não sendo criada qualquer representação simbólica do mundo (figura 6).

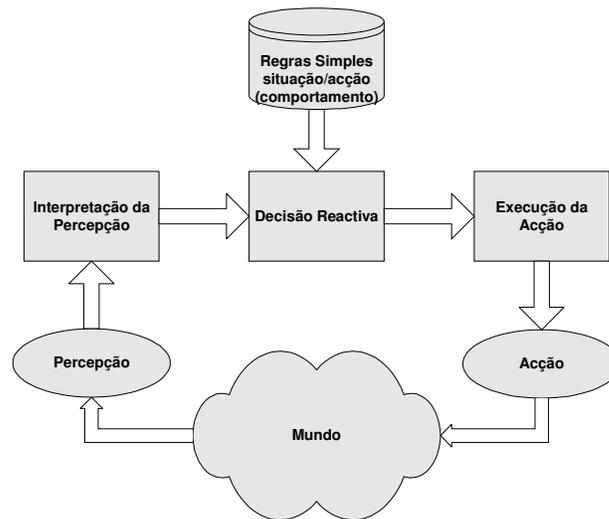


Figura 6: Esquema genérico de uma Arquitectura Reactiva

A arquitectura reactiva melhor conhecida é a *Arquitectura de Subordinação* (“*Subsumption Architecture*”), de Rodney Brooks [Brooks, 1986]. Existem duas características principais na arquitectura de subordinação, que são:

- A tomada de decisão do agente deve ser efectuada através de um conjunto de comportamentos direccionados para a execução de tarefas, isto é, cada comportamento deve ser pensado como uma função individual de acção, que vai continuamente tomando conhecimento do ambiente e traduz essa percepção para uma acção a executar;
- Vários comportamentos podem ser accionados simultaneamente. Assim, deverá existir um mecanismo para seleccionar de entre as diferentes acções possíveis, aquela que será a acção óptima em cada momento. Desenvolve-se então uma Hierarquia de Subordinação, que organiza os comportamentos por níveis. Os níveis mais baixos da hierarquia são prioritários em relação aos comportamentos de níveis superiores, isto porque, as acções de níveis mais altos representam comportamentos mais abstractos.

A arquitectura reactiva, nomeadamente a arquitectura de subordinação, tem grandes vantagens, como sendo, a simplicidade, economia, bom trato computacional e boa robustez contra falhas [Wooldridge e Jennings, 1995], mas também tem desvantagens evidentes, nomeadamente no que se refere à característica de os agentes decidirem unicamente baseados em informação da percepção actual, possuem uma hierarquia de comportamentos pré-fixada e serem incapazes de realizar acções que impliquem a execução de planos de longo-prazo.

### 2.5.3 Arquitecturas Híbridas

Como o próprio nome sugere, as *Arquitecturas Híbridas* combinam as características das duas abordagens anteriores. Os agentes puramente reactivos têm uma limitação significativa que é a sua dificuldade em implementar comportamento orientado por objectivos. Os agentes puramente deliberativos são baseados em mecanismos de raciocínio simbólico muito complexos e tornam-se, por vezes, incapazes de uma reacção imediata a estímulos do exterior. Um agente híbrido combina estas duas componentes, e caracteriza-se por uma arquitectura composta por níveis ou camadas. A possibilidade de realizar uma disposição por níveis é uma ferramenta poderosa de estruturar as funcionalidades e controlo do agente. A ideia principal é categorizar as funcionalidades do agente em camadas dispostas hierarquicamente onde geralmente a camada reactiva tem alguma prioridade sobre a camada deliberativa, de modo a permitir uma resposta rápida aos eventos mais importantes detectados no ambiente (figura 7).

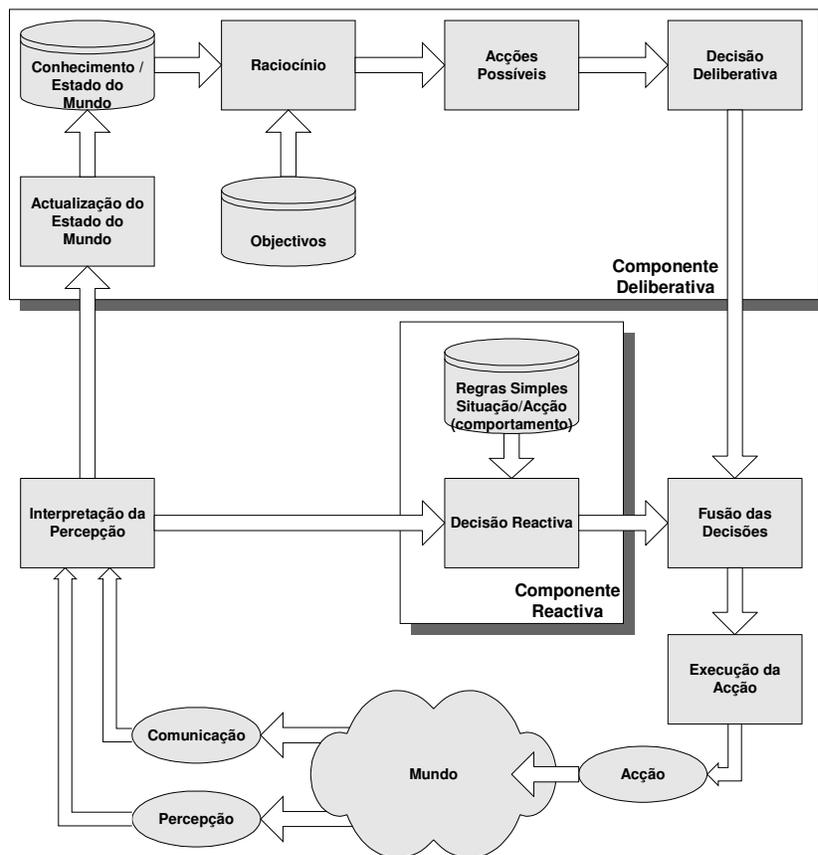


Figura 7: Esquema genérico de uma Arquitectura Híbrida

Na figura 7 é visível o esquema genérico de uma arquitectura híbrida. Após a interpretação da percepção e da comunicação provenientes do ambiente, o agente possui duas componentes: uma componente reactiva e uma componente deliberativa. O funcionamento da componente reactiva é extremamente simples e baseia-se num conjunto

de regras situação/acção que associam directamente a determinados estímulos, determinadas decisões. A componente deliberativa implica a construção de um estado simbólico do mundo e a utilização de raciocínio simbólico de forma a decidir em cada instante as acções a executar de forma a atingir os objectivos. Uma parte complexa e sujeita a um elevado número de trabalhos de investigação, é a forma de efectuar a fusão das decisões deliberativa e reactiva como forma de seleccionar a acção final a executar.

Diversos investigadores da área utilizam arquitecturas híbridas para agentes capazes de executar diferentes tarefas. Neves e Oliveira [Neves e Oliveira, 1997] [Neves, 1999] propõem a utilização desta solução para controlar um robô móvel deslocando-se num ambiente não totalmente previsível. A componente deliberativa tem aqui como função guiar o comportamento da componente reactiva, através de sugestões de planos a realizar. Uma das inovações deste trabalho está contida no facto de existir uma interacção entre as duas componentes, podendo a componente deliberativa alterar as regras de situação/acção que a componente reactiva utiliza. A componente reactiva, em certas situações de emergência tais como a iminência de uma colisão, sobrepõe-se à deliberativa, assumindo o controlo do robô.

#### 2.5.4 Arquitectura por Camadas

A arquitectura por camadas, como o próprio nome indica, compreende a existência de vários subsistemas que se organizam em hierarquias que interagem por níveis [Wooldridge e Jennings, 1995]. Aliás, o esquema típico de uma arquitectura híbrida é a arquitectura por camadas. Podem-se considerar dois tipos de camadas, de acordo com o fluxo de controlo:

- **Camadas Horizontais.** Onde as camadas de *software* estão directamente ligados aos sensores de *input* e às acções de *output*, ou seja, cada camada actua como um agente, uma vez que apresenta sugestões no que diz respeito à acção a efectuar pelo agente (Figura 8); e
- **Camadas Verticais.** Onde os sensores de *input* e as acções de *output* têm, no mínimo uma camada entre eles, isto é, têm de ser trabalhados, pelo menos, por uma camada (ver Figura 9).

A grande vantagem da arquitectura horizontal sobre a arquitectura vertical, é a sua simplicidade conceptual. Isto porque, se é pedido ao agente para ter  $n$  tipos diferentes de comportamentos, então apenas será necessário implementar  $n$  camadas diferentes. Por outro lado, uma vez que as camadas competem umas com as outras, poderá ser necessária

a introdução de um mediador<sup>10</sup> que, embora venha para resolver o problema pode fazer com que haja um estrangulamento no processo de decisão do agente.

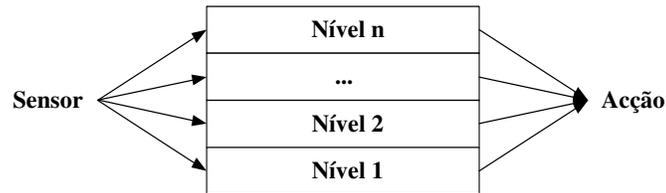


Figura 8: Representação da Arquitectura por Camadas Horizontais

Este problema praticamente não acontece com a arquitectura vertical, uma vez que o fluxo de controlo passa sequencialmente por cada camada, até que a última camada gere a acção a executar. No entanto, nesta arquitectura existe pouca flexibilidade, já que para o agente tomar uma decisão é necessário que o fluxo de controlo atravessasse cada uma das diferentes camadas, ou seja, caso haja uma falha numa das camadas, o agente terá problemas para gerar e executar uma acção.

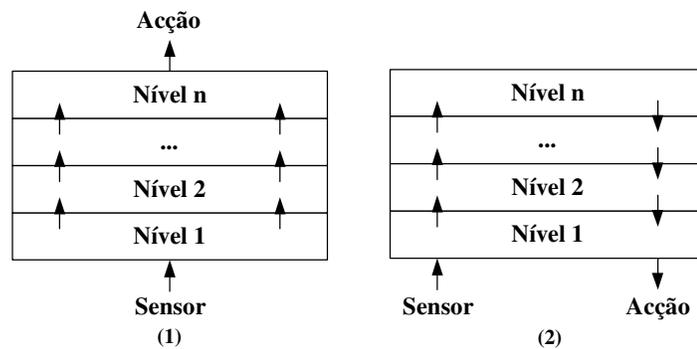


Figura 9: Representação da Arquitectura por Camadas Verticais

Conforme demonstra a Figura 9, a arquitectura por camadas verticais pode incluir unicamente uma passagem de controlo (Figura 9 (1)) ou duas passagens de controlo (Figura 9 (2)). Na arquitectura com uma passagem, o fluxo de controlo passa sequencialmente pelas diferentes camadas da arquitectura, até ao último nível, onde é gerada a acção a executar. Na arquitectura com duas passagens, a informação sobe as várias camadas até atingir o nível superior, sendo esta a primeira passagem. Em sentido inverso, o fluxo de controlo desce e atravessa as várias camadas, dando origem à acção a executar, sendo esta a segunda passagem.

<sup>10</sup> Este mediador irá funcionar de forma semelhante ao módulo de fusão das decisões apresentado para a arquitectura híbrida.

### 2.5.5 Arquitectura BDI

A arquitectura BDI (“*Belief-Desire-Intention*”) [Rao e Georgeff, 1991] é uma arquitectura essencialmente deliberativa, onde o estado interno de processamento de um agente é descrito através de um conjunto de “estados mentais”. Estes “estados mentais” adoptados na definição das arquitecturas BDI, diferem ligeiramente de autor para autor. No entanto, as noções de *crença* (“*belief*”), *desejo* (“*desire*”) e *intenção* (“*intention*”) são aceites unanimemente, até porque estão na origem do nome atribuído a esta classe de arquitecturas de agentes. As intenções numa arquitectura BDI podem ser definidas a diferentes níveis de detalhe. Por exemplo, um agente pode ter como intenção comprar um dado livro mas não ter ainda decidido em que loja o comprar. Desta forma a estrutura de intenções assemelha-se a um plano hierárquico onde os operadores aplicáveis nos vários níveis de abstracção do plano e as intenções na estrutura de intenções possuem propriedades semelhantes.

As *crenças* de um agente referem-se ao que o agente acredita em cada instante, e descrevem o estado do mundo do agente (o seu conhecimento sobre o ambiente). As crenças representam, desta forma, informação.

Os *desejos* de um agente referem-se ao que o agente deseja obter. No entanto, a forma de alcançar esses desejos pode não ser conhecida num dado instante. Os desejos podem ser, num dado momento, inconsistentes. Os objectivos dos agentes resultam de um processo de raciocínio, por parte do agente, que consiste numa escolha de um subconjunto dos desejos que são consistentes e atingíveis.

As *intenções* referem-se a um conjunto de acções ou tarefas que o agente seleccionou, comprometendo-se assim na realização dos seus objectivos. As intenções devem ser consistentes internamente e representam o resultado do processo de deliberação.

O comportamento de um agente “BDI” é ditado pelas relações existentes entre os três “estados mentais” básicos, sendo identificados, na literatura, três tipos de comportamento para estes agentes [Rao e Georgeff, 1995]:

- **Realismo Forte.** O conjunto de intenções é um subconjunto do conjunto de desejos, que é, por sua vez, influenciado por um subconjunto das crenças. Se um agente não acredita em algo, nunca irá desejar ou intentar esse algo.
- **Realismo.** O conjunto de crenças é um subconjunto do conjunto dos desejos que é, por sua vez, influenciado por um subconjunto do conjunto de intenções. Um agente realista, se acredita em algo, então deseja e intenta esse algo.
- **Realismo Fraco.** O agente não deseja algo se a negação desse algo é acreditada; não intenta algo se a negação desse algo é desejada; e não intenta algo se a negação desse algo é acreditada.

A arquitectura genérica de um agente BDI [Wooldridge, 1999] é constituída por sete componentes: conjunto de crenças, função de revisão de crenças, função de geração de opções, conjunto de opções, função de filtragem, conjunto de intenções e função de selecção de acção (figura 10).

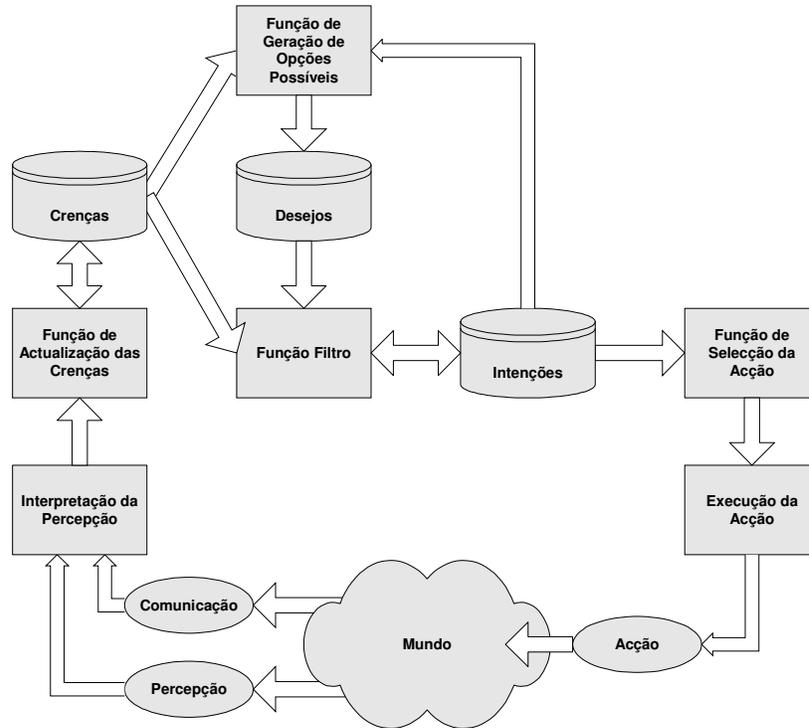


Figura 10: Esquema genérico de uma Arquitectura BDI

Estes componentes podem ser descritos da seguinte forma:

- **Conjunto de crenças.** Representa a informação que o agente possui em cada instante sobre o ambiente.
- **Função de revisão de crenças.** Função que actualiza o conjunto de crenças, de acordo com novas percepções e comunicações recebidas pelo agente e tendo em conta as crenças actuais do agente.
- **Função de geração de opções.** Função que determina as opções (desejos) disponíveis para o agente, com base nas suas crenças e intenções actuais. É o processo que permite ao agente decidir a linha de acção a seleccionar em cada instante, de forma a satisfazer as suas intenções. Algumas das opções resultantes da aplicação desta função tornar-se-ão intenções, e servirão de “feedback” para geração de novas opções, resultando na geração de opções mais concretas. O agente irá considerar assim comprometer-se a intenções progressivamente mais específicas, até que o conjunto das suas intenções corresponda a acções imediatamente executáveis. Esta função deve apresentar-se consistente - as opções geradas devem ser consistentes com crenças e intenções actuais - e

oportunista - deve ser capaz de reconhecer quando o ambiente se altera de modo a oferecer ao agente novas formas de satisfação das suas intenções, ou a possibilidade de atingir intenções que eram inatingíveis.

- **Desejos.** Representam os possíveis cursos de acção (opções) disponíveis para o agente.
- **Função de filtro.** Esta função representa o processo de deliberação do agente, que determina as intenções deste com base nas suas crenças e desejos actuais e nas suas intenções prévias. É o processo onde o agente decide o que fazer e actualiza em consonância as suas intenções. Esta actualização implica, por vezes, a desistência de intenções que deixam de ser atingíveis ou apresentam um custo demasiado elevado; e ainda a adopção de novas intenções que resultaram da adopção de novas opções.
- **Intenções.** Representa o conjunto de intenções, i.e., os estados que o agente está actualmente comprometido a tentar obter.
- **Função de selecção de acção.** Esta função determina qual a acção a realizar em cada momento com base nas intenções actuais do agente.

Refere-se ainda uma outra categorização dos “estados mentais” de um agente, realizada a um nível superior, de acordo com o conceito de atitude [Wooldridge e Jennings, 1995]. Segundo estes autores, os “estados mentais” de um agente podem ser incluídos em duas classes: atitude de informação e pró-atitude. A atitude de informação está relacionada com a informação que o agente possui sobre o mundo no qual está inserido. Nesta categoria inclui-se a *crença* e o *conhecimento*. A pró-atitude diz respeito às atitudes que guiam as acções do agente. Nesta categoria, incluem-se o *desejo*, a *intenção*, a *obrigação*, o *compromisso* e a *escolha*.

Para melhor compreender a distinção entre os conceitos de *desejos* e *intenções* surge a noção de *compromisso*. Pode afirmar-se que um agente está sempre comprometido com as suas *intenções*, e não, obrigatoriamente, com os seus *desejos* (objectivos). Este compromisso define o grau de persistência que o agente terá com as suas intenções. Determina ainda o modo como as intenções são revistas e actualizadas. As condições de término de um compromisso, classificam estes em três tipos diferentes: compromisso cego (o agente mantém a intenção até que acredite que atingiu essa intenção), compromisso honesto (o agente mantém a intenção enquanto acreditar que esta ainda pode ser atingida) e compromisso aberto (o agente desiste da intenção, não só se já não acredita que esta possa ser atingida, mas também se a intenção já não for um objectivo).

A existência de outros agentes no sistema levanta problemas adicionais ao processo de deliberação de um qualquer agente, pois é necessário raciocinar sobre as intenções dos outros, e possivelmente, sobre as intenções de conjuntos de agentes que se juntam em

actividades de grupo<sup>11</sup>. Norman [Norman *et al.*, 1998] introduz a noção de *acordos* entre agentes, que se baseia em *direitos* e *acções* a serem realizadas. As acções que um agente pode realizar para atingir os seus objectivos sem incorrer em penalidades impostas pelos outros membros da sociedade, são determinadas pelas acções que é capaz de executar e tem direito a realizar. Os agentes podem ainda *delegar* acções, isto é, dar permissão a outros agentes para executar acções que apenas eles têm direito a executar. Os agentes envolvidos num acordo têm de se comprometer a agir em conformidade com esse acordo. Para assegurar a estabilidade da actividade do grupo, aos agentes que estão envolvidos num acordo, é requerida uma noção de *compromisso de grupo* para agir em concordância com o acordo. Tal compromisso de grupo assegura que cada agente inserido no grupo está obrigado a fazer cumprir a sua parte no acordo. Neste contexto, Castelfranchi [Castelfranchi, 1995] faz uma distinção entre *compromisso individual*, *compromisso social* e *compromisso colectivo*. Um *compromisso social* é mais do que um *compromisso individual* partilhado por um grupo de agentes e envolve a relação entre pelo menos três agentes e uma acção. Um *compromisso colectivo* é o compromisso individual de um conjunto de agentes a uma determinada intenção, existindo conhecimento da sociedade sobre esse facto.

### 2.5.6 Arquitectura Genérica de um Agente Social

Uma arquitectura extremamente completa mas também bastante complexa é a de um agente deliberativo social. Este tipo de agente deverá possuir modelos explícitos dos outros agentes e capacidades para manter estes modelos actualizados em face da sua percepção e das comunicações recebidas. Estes modelos incluem crenças, objectivos e planos dos outros agentes. O agente deverá ser capaz de raciocinar utilizando o conhecimento introduzido neste módulos (intenções, expectativas, compromissos, predições, reacções antecipadas, etc.), realizando decisões e criando planos com respeito aos outros agentes. O nível de complexidade das tarefas executadas pelo agente ditará o nível de complexidade das suas estruturas de conhecimento e raciocínios efectuados. Na figura 11 é apresentada uma arquitectura genérica para um agente social baseada nos trabalhos de Bond e Gasser [Bond e Gasser, 1988b] e Moulin e Chaib-Draa [Moulin e Chaib-Draa, 1996].

---

<sup>11</sup> No futebol robótico este raciocínio é essencial, uma vez que existe uma equipa de agentes com objectivos opostos à equipa do agente.

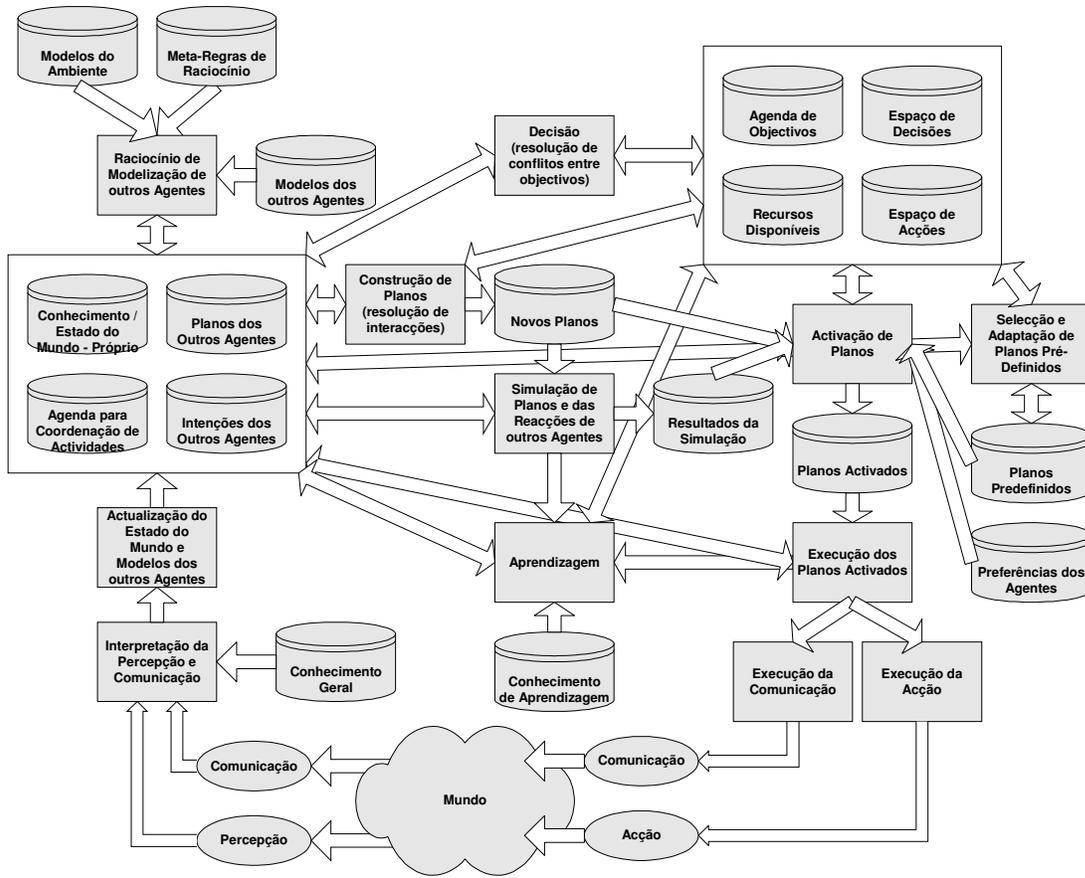


Figura 11: Esquema Genérico de uma Arquitectura de um Agente Social

Como é visível analisando a figura 11, a arquitectura proposta por Bond e Gasser é extremamente complexa, incluindo uma grande variedade de módulos que lidam com conhecimentos distintos.

A percepção e a comunicação provenientes do ambiente são interpretadas pelo agente à luz do seu conhecimento. Utilizando esta informação devidamente tratada, o agente mantém actualizada uma estrutura de informação que contém o estado do mundo actual, a agenda do agente que é responsável por activar as suas actividades activando oportunisticamente os módulos necessários, conhecimento próprio e sobre os outros agentes, nomeadamente no que diz respeito aos seus possíveis planos e intenções. As informações relativas aos outros agentes são derivadas por um módulo de raciocínio e modelização dos outros agentes. Este módulo utiliza modelos pré-definidos dos agentes e do ambiente e meta-regras de raciocínio. O módulo de modelização dos outros agentes pode ser activado de forma a inferir as intenções e planos dos restantes agentes.

O módulo de decisão é responsável por seleccionar do espaço de decisão, quais os objectivos a atingir dadas as crenças actuais do agente e as intenções dos outros agentes. Os objectivos seleccionados são colocados na agenda de objectivos que vai ser utilizada pelas funções de planeamento.

O agente possui um conjunto de planos pré-definidos e um módulo que procura seleccionar e adaptar esses planos à situação corrente. Por vezes, é impossível utilizar planos pré-definidos e como tal novos planos têm de ser construídos tendo em conta as crenças actuais do agente e os recursos que dispõe. O agente pode também ter a capacidade de simular o efeito dos novos planos tendo em conta o estado do mundo actual e as suas crenças sobre as intenções e planos dos outros agentes. Os resultados desta simulação podem ser armazenados e utilizados no processo de decisão.

A aprendizagem pode também ser incluída no agente de forma a que este utilize os resultados da execução das suas acções e conhecimento específico de aprendizagem no sentido de melhorar as suas decisões futuras. Utilizando aprendizagem, novos planos e novos objectivos podem ser incluídos respectivamente no espaço de acções e espaço de decisões do agente.

A decisão de quais os planos que vão ser activados tem em consideração as preferências do agente, os resultados da simulação, os planos disponíveis (novos e pré-definidos) e os objectivos do agente. Os planos seleccionados são então executados através da execução de acções no ambiente e comunicações com os outros agentes. As crenças do agente são actualizadas tendo em conta a informação relativa aos planos em execução.

Esta arquitectura, embora bastante complexa, contém uma base para a discussão de arquitecturas de agentes sociais envolvidos num sistema multi-agente.

### **2.5.7 Seleção da Arquitectura Mais Adequada**

Habitualmente os diferentes investigadores na área dos agentes autónomos definem arquitecturas específicas para os agentes que projectam no âmbito de cada aplicação. Embora baseadas nas arquitecturas básicas apresentadas nesta secção, usualmente, as arquitecturas utilizadas em aplicações práticas incluem módulos específicos como por exemplo, módulos de: aprendizagem, raciocínio estratégico, escalonamento, planeamento, modelização, comunicação, cooperação, negociação, revisão de crenças, simulação, manutenção de conhecimento, etc.

A selecção da arquitectura a utilizar depende em grande medida da complexidade da tarefa que o agente irá executar e do ambiente em que ele irá operar, nomeadamente no que diz respeito às suas características (acessibilidade, dinamismo, continuidade, determinismo, etc.). Depende ainda dos outros agentes que irão coexistir com o agente no ambiente e das suas características: competitivos, cooperativos, desconhecidos à partida, etc.

## **2.6 Aplicações de Agentes Autónomos**

O número de aplicações potenciais da tecnologia dos agentes a nível industrial e comercial é enorme. De entre estas aplicações destacam-se:

- **Aplicações Industriais.** Embora a tecnologia dos agentes seja relativamente recente, diversas aplicações industriais foram projectadas e implementadas. Aliás estas aplicações estiveram entre as primeiras a ser desenvolvidas [Jennings et al. 1998]. De entre as aplicações industriais mais relevantes da tecnologia dos agentes autónomos destacam-se: aplicações na área das telecomunicações, controlo de fabrico e controlo de processos, distribuição de energia eléctrica [Wittig et al., 1994], controlo de tráfego aéreo e sistemas de transportes [Jennings et al., 1998].
- **Agentes de Pesquisa de Informação.** Com o enorme crescimento verificado na Internet ao longo dos últimos anos e devido à natureza dinâmica e heterogénea da informação e da Internet, o paradigma dos agentes revela-se adequado à implementação de pesquisadores de informação. Desta forma as aplicações de agentes de pesquisa, recuperação e filtragem de informação na Internet constituem uma área em franca expansão no âmbito dos agentes autónomos. Neste contexto, têm também ganho relevância aplicações de agentes vulgarmente designadas por assistentes pessoais e que possuem o objectivo de auxiliar um dado utilizador a realizar as suas tarefas mais repetitivas. De entre as aplicações mais conhecidas destacam-se a gestão de correio electrónico, filtragem de notícias e marcação distribuída de reuniões.
- **Comércio Electrónico.** O enorme crescimento do comércio electrónico nos últimos anos e a evidente aplicabilidade dos agentes a este domínio levaram ao aparecimento de inúmeras aplicações de mercados electrónicos e agentes para comércio electrónico. De entre as aplicações mais relevantes destacam-se os mercados electrónicos para B2C (“business-to-consumer”) tais como o Kasbah [Chavez e Maes, 1996], SMACE [Cardoso e Oliveira, 2000] [Cardoso, 1999] ou o AuctionBot [Wurman et al., 1998] e para B2B (“business-to-business”) como o FOREV [Macedo, 2001] [Rocha e Oliveira, 2001a].
- **Aplicações de Entretenimento.** Os jogos e outros sistemas destinados ao lazer, sempre foram uma das aplicações mais relevantes de grande parte da tecnologia humana. Embora por vezes as aplicações relacionadas com o entretenimento não sejam consideradas investigação a sério, na realidade é frequente encontrar avanços científicos aplicados a áreas socialmente úteis que derivam directamente do desenvolvimento de aplicações na área dos jogos e lazer. De entre as aplicações de agentes neste domínio destacam-se os jogos, o desenvolvimento de personagens virtuais e histórias interactivas.
- **Aplicações Médicas.** A área da saúde sempre foi uma das áreas por excelência de aplicação da Inteligência Artificial. Sistemas Periciais tais como o sistema de diagnóstico médico MYCIN, encontram-se entre as mais mediáticas e bem sucedidas realizações da IA. Este tipo de sistemas periciais foram substituídos ao

longo dos últimos anos por aplicações baseadas no conceito de agente. Os agentes são também aplicados no domínio da saúde e as suas principais áreas de utilização situam-se ao nível de tarefas tais como o controlo de robôs e outros equipamentos hospitalares e sistemas de tratamento de pacientes distribuídos.

- **Agentes para Simulação.** A utilização de agentes em simuladores encontra-se cada vez mais generalizada. As áreas de aplicação neste contexto incluem desde simuladores de condução, voo, combate aéreo, futebol, produção e manufactura e diversas áreas da robótica.
- **Controlo de Robôs.** A utilização de agentes no controlo de robôs destina-se a obter uma navegação segura e eficiente e um funcionamento global versátil do robô. Na maioria das aplicações, o controlo de robôs efectuado directamente por humanos é, na prática, impossível e como tal o recurso a agentes autónomos é a solução mais aconselhável. Arquitecturas baseadas em agentes para o controlo de robôs móveis, incluem [Neves e Oliveira, 1997] [Reis, 2002a] e [Lopes et al., 1999] [Lopes et al., 2000].
- **Competições de Investigação Científica.** Uma área em grande expansão actualmente é a área das competições científicas. Neste tipo de competições, a investigação realizada por investigadores de uma dada área, é comparada utilizando um problema ou plataforma comum. O paradigma dos agentes autónomos é bastante adequado para permitir a realização deste tipo de competições mesmo no caso de a investigação a comparar não estar directamente relacionada com os agentes. De facto, se cada universidade ou empresa projectar o seu agente para resolver determinado tipo de problemas e operar num dado ambiente, será muito simples comparar esses agentes posteriormente. De entre as competições mais significativas, merecem destaque o RoboCup – Futebol Robótico [Kitano, 1997] [Kitano et al., 1997], o TAC – *Trading Agent Competition* [TAC, 2001] e a competição portuguesa *Micro-Rato* [Reis, 2002a].

## 2.7 Conclusões

Embora a construção de agentes individuais seja muito importante, na grande maioria das aplicações de agentes autónomos, estes não vão trabalhar isolados mas sim em conjunto com outros agentes e/ou humanos. Desta forma interessa estudar as formas de colocar os agentes a trabalhar em conjunto com outros agentes, ou seja, construir sociedades de agentes designadas vulgarmente por Sistemas Multi-Agente. Um dos pontos essenciais para permitir a construção dessas sociedades de agentes consiste em conseguir gerir as interacções e as dependências das actividades dos diferentes agentes no contexto do Sistema Multi-Agente, de forma a que estes possam trabalhar em conjunto de forma harmoniosa, i.e., coordenar esses agentes. Desta forma os próximos capítulos analisam a

investigação realizada na área dos Sistemas Multi-Agente e na coordenação de agentes competitivos (interessados no seu bem próprio) e cooperativos (interessados no bem da comunidade).

Este capítulo não teve como objectivo fornecer um “*overview*” detalhado da área dos agentes autónomos mas sim introduzir os principais conceitos necessários para a compreensão dos restantes capítulos da tese. Para uma análise mais profunda e detalhada do campo dos agentes autónomos e do campo dos Sistemas Multi-Agente analisado no capítulo seguinte, sugere-se a consulta de [O’Hare e Jennings, 1996], [Huhns e Singh, 1997a], [Weiss, 1999] ou [Wooldridge, 2002].