CAPÍTULO 1

INTELIGÊNCIA ARTIFICIAL

"Musical meaning arises when an antecedent situation requiring an estimate as to the probable modes of pattern continuation produces uncertainty as to the temporal-tonal nature of the expected consequent. A system which produces a sequence of symbols according to certain probabilities is called a stochastic process, and the special case of a stochastic process in which the probabilities depend on the previous events, is called a Markoff chain. Since in a Markoff chain, the uncertainty decreases as the distance from the starting point increases, in order to heighten the meaning (information) of the musical discourse the composer will have to introduce some uncertainty at every step."

Umberto Eco, The Open Work.

1 Inteligência artificial.

1.1 Apresentação.

Neste primeiro capítulo faremos uma descrição de caráter filosófico da relação interdisciplinar da ciência cognitiva, pelo viés da inteligência artificial (IA). Iniciaremos, pois, com uma descrição do chamado paradigma cognitivista, e sua aplicação no estudo da mente. Neste ponto pretendemos clarificar os principais conceitos e noções envolvidos nas modelagens cognitivistas, tais como o mecanicismo, o funcionalismo, o computacionalismo, máquina de Turing, e representações mentais simbólicas. Em seguida, descreveremos mais detalhadamente o método utilizado em IA para a modelagem da mente, que tem como pressupostos os conceitos logo acima citados, e iremos nos remeter a algumas das implementações clássicas da inteligência artificial.

1.2 O estudo da mente no enfoque da IA – o paradigma cognitivista.

1.2.1 Contexto histórico.

Estabelecer um limite preciso no surgimento de uma nova área científica é uma tarefa, normalmente, controversa. Assim também ocorre com a ciência cognitiva. Gardner (1995) atribui este início ao simpósio Hixon de 1948 sobre Mecanismos Cerebrais do Comportamento. Por outro lado, Dupuy (1995) atribui o surgimento desta área às Conferências Macy, que ocorreram de 1946 a 1953, num total de dez conferências. Certamente, os maiores nomes do início ciência cognitiva estavam presentes em ambos os eventos realizados nos Estados Unidos, como John von Neumann, Nobert Wiener, Hebert Simon, Alan Newell, Warren McCulloch, Walter Pitts, Karl Lashley, entre inúmeros outros, o que dificulta uma demarcação precisa sobre seu surgimento. Podemos dizer, que por volta do final da década de 1940 começaram a surgir eventos e encontros entre uma comunidade

científica interdisciplinar que fundou o que ficou conhecido como *cibernética*. Esse grupo interdisciplinar envolvia pesquisadores das áreas de matemática, lógica, engenharia, fisiologia e neurofisiologia, psicologia e antropologia (DUPUY, 1995, p.9). O grupo cibernético tinha um interesse em comum: um estudo da mente de uma forma empírica e que ao mesmo tempo considerasse estados mentais em sua abordagem. Tal afirmação nos leva a dois contrapontos teóricos: um com *o dualismo cartesiano*, e outro com o *behaviorismo*.

O dualismo cartesiano é a doutrina filosófica que postula uma dicotomia ontológica e substancial entre a mente e o corpo. Mente e corpo seriam, então, compostos de duas substâncias de natureza diferente, o corpo sendo de natureza física, e dessa forma sujeito à suas leis mecânicas, e a mente sendo não física, portando não sujeita a tais leis. Broens (1998, p.190) afirma que a natureza da mente "(…) consiste em ser puro pensamento: o exercício da atividade reflexionante independe, para Descartes, do corpo físico ao qual, no entanto, a razão (ou espírito) está intimamente ligada." Uma das grandes questões relacionadas ao dualismo substancial cartesiano é justamente explicar como ocorre a ligação entre mente e corpo, Como duas substâncias diferentes podem ter uma relação de causa eficiente. Além disso, por não ser física, a mente (ou alma, na terminologia de Descartes) não pode ser acessada objetivamente. Apenas através da introspecção o sujeito pode atingi-la. O dualismo substancial percorreu os séculos e de certa forma ainda é uma doutrina aceita, inclusive um dos temas centrais da atual filosofia da mente é relacionado ao problema mente-corpo.

Muitas das áreas que têm como interesse central o estudo da mente e do conhecimento, como a psicologia e a filosofia, por exemplo, aceitavam o dualismo cartesiano e desenvolveram suas teorias metafísicas sobre ele. Contudo, durante o século XX o combate filosófico e científico às idéias provenientes do dualismo cartesiano foi o *leitmotive* entre as linhas de pesquisa interessadas no estudo mente.

O outro contraponto teórico que queremos brevemente estabelecer com a ciência cognitiva é a psicologia behaviorista, que surgiu no final do século XIX e teve um importante papel na primeira metade do século XX, justamente por buscar um entendimento científico da mente humana. De maneira bastante simplificada pode-se considerar que o ponto de partida do behaviorismo é que a mente era vista como inacessível a um observador externo. Portanto, como estudar um objeto ao qual não se tem acesso direto? A resposta metodológica dos behavioristas foi estudar sua manifestação: o *comportamento*. Não se tem acesso à mente, seja ela o que for, mas pode-se de maneira objetiva analisar e estudar o comportamento humano, num sistema de estímulo e resposta. Com o behaviorismo surge a bem conhecida

metáfora da caixa-preta. A mente é vista como uma caixa-preta, e dessa forma, só podem ser analisados os estímulos sensoriais que entram nela e as respostas comportamentais que são geradas. O método behaviorista de investigação é baseado na apresentação de estímulos e análise de respostas, utilizando dados precisamente mensuráveis em situações controladas de ambiente laboratorial. Podemos afirmar que o behaviorismo iniciou uma das primeiras investigações científicas de aspectos da mente humana, pelo viés do estudo do comportamento.

Será em oposição aos pressupostos behavioristas que irá surgir no final da década de 1940 a agenda cognitivista, ainda no chamado movimento cibernético. O objetivo principal, então, desta nova ciência, era o estudo do que se encontra entre estímulo e resposta, abrir a caixa-preta. Antes, porém, de nos aventurarmos mais profundamente nos métodos investigativos da inteligência artificial a respeito da mente, vamos passar por uma breve descrição de três conceitos que serão fundamentais e indispensáveis posteriormente.

1.2.2 O mecanicismo.

Para entendermos a noção de mecanicismo temos que nos remeter novamente ao dualismo, e sua contrapartida, o monismo. Já vimos que o dualismo postula uma dicotomia entre mente e corpo. O monismo, ao contrário, diz que mente e corpo são a mesma coisa, ou seja, a mente se reduz ao cérebro, não existe uma substância mental ontologicamente diferente da substância física. O próprio Descartes admitia o mecanicismo do corpo, apenas não o admitia no caso da "alma", (por exemplo, no *Discurso do Método*, quinta parte). Em contrapartida, na mesma época, Pascal já imaginava a possibilidade de máquinas executarem atividades consideradas exclusivamente mentais (BROENS, 1998). O mecanicismo entende que a mente é a realização de processos ordenados de maneira legisforme, na perspectiva de um universo mecânico. Conhecendo-se as leis que ordenam tais processos podemos explicar o que a mente é, ou pelo menos como ela funciona, e mesmo reproduzi-la mecanicamente. Numa postura reducionista, se a natureza da mente é a mesma que a do cérebro; se estas duas coisas são na verdade uma só, e sendo o cérebro físico e, portanto, sujeito às leis da física, logo a mente também o é. O mecanicismo mental é, normalmente, associado a uma perspectiva monista, onde mente e cérebro se identificam (ou a primeira se reduz ao

segundo), mas existe uma terceira via entre o dualismo substancial e o monismo, chamada de *dualismo de propriedades* ou *monismo anômalo*. Nesta perspectiva, a mente é vista como uma *propriedade emergente* da atividade física do cérebro, mas não redutível a ela. Dessa forma, não podemos entender nem descrever a natureza e o funcionamento da mente olhando apenas para o nível local, cerebral.

Monistas ou dualistas de propriedades, ambos admitem que a mente é física, e está ligada diretamente, de maneira causal ao cérebro, ou mesmo que a noção de mente se reduz ao cérebro e, portanto, pode ser descrita em termos mecânicos e é realizada (total ou parcialmente) por aparatos mecânicos. O que o mecanicismo afirma então é: a mente pode ser explicada total ou parcialmente por leis, e em termos físicos (mecânicos).

Por sua vez, procedimentos mecânicos podem ser análogos uns aos outros, como numa máquina que execute a mesma tarefa que outra. Um coração artificial executa a mesma tarefa que um biológico, por exemplo. Da mesma forma, hipoteticamente, funções mentais podem ser realizadas por outros tipos de máquinas diferentes do cérebro humano. Nisto liga-se outra noção importante para a IA: o funcionalismo.

1.2.3 O funcionalismo.

O funcionalismo é uma teoria filosófica normalmente creditada a Hilary Putnam (1980a, 1980b). Basicamente o que ela nos diz é que um sistema mental pode ser realizado por diversos tipos de sistemas físicos diferentes desde que apresentem a complexidade suficiente, independentemente de sua natureza ou funcionamento. Inclusive essa múltipla realização, talvez, possa ocorrer independentemente das atualmente conhecidas leis da física. Podemos dizer, baseados em seus divertidos exemplos, que hipoteticamente pode haver uma mente em outros planetas onde, quem sabe, as leis da mecânica (pelos menos as que nós conhecemos) não são aplicáveis. Para o funcionalismo, portanto, não importa nem o material nem as leis a que tal material está sujeito para que possa existir um sistema que apresente estados mentais. *"We could be made of Swiss cheese and it wouldn't matter"* (PUTNAM, 1980b, p.134). Há, dessa forma, dois níveis isomórficos distintos, um físico e outro funcional, cuja relação parece puramente acidental:

> "*(...) to identify the [mental] state in question with its physical or chemical realization would be quite absurd, given that the realization is in a sense quite accidental, from the point of view of psychology, anyway. (...) It is as if we met Martians and discovered that they were in all functional respects isomorphic to us, but we refused to admit that they could feel pain because their C fibers were differen.*" *(PUTNAM, 1980b, p.136)*

O funcionalismo é uma teoria que foi bastante questionada na filosofia da mente e na psicologia (BLOCK, 1980). Inclusive, como é possível observar, o funcionalismo foi acusado de ser uma teoria dualista, já que postula a existência de dois planos distintos, cuja ligação entre eles é "bastante acidental", nas palavras de Putnam. No mínimo podemos coloca-lo na categoria de dualismo de propriedades, visto que diversos sistemas diferentes podem apresentar os mesmos estados mentais, se forem complexos o suficiente; nesse caso, a mente seria uma propriedade emergente da complexidade física. De qualquer forma, explicar a conexão entre mente e cérebro é o antigo problema enfrentado pela filosofia cartesiana, como já vimos. Essa acusação de dualismo é uma conseqüência vinda da IA, com seu método top-down. A teoria funcionalista foi uma decorrência do surgimento e desenvolvimento das implementações da Inteligência Artificial, desde Turing. Dupuy (1995, p.26) afirma:

> "Essa teoria que se nos tornou familiar tão familiar [o funcionalismo], pela qual distinguimos o "programa" (*software*) do "material" (*hardware*) é um produto da revolução conceptual que assinala o advento das ciências cognitivas, e não sua origem." (aspas e grifo do autor)

1.2.4 Representações mentais.

Antes de prosseguirmos em direção a uma descrição dos modelos computacionais do início da ciência cognitiva, faz-se necessária uma incursão a uma outra noção filosófica, recorrente nos estudos acerca da mente: a representação mental. Independente da postura ontológica (dualista, dualista de propriedades ou monista) e mesmo de um posicionamento favorável ou não ao funcionalismo, a maioria dos filósofos sempre defendeu a existência de representações mentais, pelo menos até o século XX. A epistemologia sempre as tomou como um pressuposto. Afirma-se que entre o sujeito cognitivo e o mundo externo a ele existe uma mediação, as representações mentais. Nessa perspectiva, a experiência do sujeito não ocorre

em sua interação direta com o mundo e seus objetos, mas sobre as representações mentais que estão entre um lado e outro. Podemos defini-las como objetos internos que estão no lugar de objetos externos, representando-os.

Conforme veremos ao longo desse trabalho, a noção de representação mental foi inúmeras vezes revisitada e reformulada ao longo da ciência cognitiva, chegando até a negação de sua existência nos processos mentais (PORT e VAN GELDER, 1998). A discussão sobre representações é extensa e importantíssima para a ciência cognitiva, mas por agora, vamos nos concentrar em um tipo específico de representação mental. No caso específico da IA, é assumida uma postura que considera a representação mental como de natureza simbólica. Sendo simbólica, existe uma relação arbitrária entre a representação e o que ela representa, dessa forma, sem nenhum tipo de correlato a não ser uma correspondência convencionalizada entre a representação e o objeto representado. Isso será adequado no caso da IA, como veremos.

1.2.5 Máquina de Turing e o modelo cognitivista.

Nas conferências anteriormente citadas (na parte 1.2.1), vimos que pesquisadores de várias áreas buscavam estabelecer um método comum para a investigação da mente, fundado na elaboração de modelos que a simulem. Tal método só foi possível de ser concebido após a década de 1930, depois das teorias de Turing sobre autômatos finitos para a resolução de problemas lógicos e, principalmente, a criação de sua máquina. Dupuy (1995, p.26) afirma que as modelagens da mente "começaram antes de existir o computador – ou, mais precisamente, quando ele existia como objeto material técnico, mas ainda não se dispunha de uma teoria funcionalista desse objeto".

O método de investigação cognitivista (na época chamado de cibernético) era, basicamente, o desenvolvimento de modelos lógicos da mente, que depois puderam ser implementados em máquinas computacionais, e.g., o computador serial digital que von Neumann criou nos anos 40. Buscava-se que a máquina desempenhasse algum tipo de atividade considerada consensualmente inteligente. Sendo o desempenho da máquina igualmente eficiente (ou pelo menos semelhante) ao desempenho humano em tais atividades, teria-se, pois, em mãos um bom modelo (lógico e abstrato) da mente humana. Esse

procedimento de comparação é conhecido como *teste de Turing* (TURING, 1950). O método de se estabelecer processos lógicos e depois implementá-los numa máquina física é conhecido como abordagem *top-down*.

O primeiro computador serial digital foi batizado de EDVAC, e baseado no ENIAC (a primeira calculadora eletrônica) desenvolvido na universidade da Pensilvânia. Na máquina ENIAC ainda não se encontrava a distinção entre *hardware* e *software*, toda a operação lógica realizada era inseparável da concepção de seus circuitos eletrônicos (uma concepção mais próxima ao monismo). Após examinar tal máquina, von Neumann concebeu a nova geração de máquinas computacionais onde existem dois níveis: o lógico e o físico (lembremos do dualismo e do funcionalismo; voltaremos a isso com mais detalhes). Dupuy (1995, p.77) afirma que von Neumann, na formulação desse sistema, baseou-se Alan Turing. A teoria da máquina de Turing forneceu a descrição formal para a construção do nível lógico da máquina de von Neumann.

Vamos então, brevemente, descrever as noções tanto criadas quanto desenvolvidas por Turing, por volta da década de 1930, que forneceram a base lógico-formal da IA.

1.2.5.1 Funções Turing-computáveis.

Em 1931, Gödel publica num artigo (Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme) a *teoria da incomplitude*. Tal teoria diz que se tivermos um sistema formal rico o suficiente para englobar a aritmética, teremos a seguinte propriedade: ou o sistema é inconsistente (inconsistência na lógica significa que o sistema possui contradições em seus teoremas), ou que existe pelo menos uma proposição verdadeira que não é demonstrável nesse sistema. Gödel chegou a essa formulação através de uma classe de funções recursivas (DUPUY, 1995 p.29).

A teoria de Gödel era uma resposta a Hilbert, que havia formulado um problema conhecido como *problema da decisão*: dada uma fórmula qualquer do cálculo dos predicados, existe um procedimento sistemático, geral, efetivo, que permita determinar se essa fórmula é demonstrável ou não? (DUPUY, 1995, p.28).

Nesta época havia um turbilhão teórico na área da lógica, e os avanços concretizados em seu *metier* foram extremamente influentes na vindoura ciência cognitiva. A lógica era

uma ferramenta que pode ser entendida como uma modelagem da razão, mesmo porque desde a filosofia aristotélica já há uma formalização de certos tipos de pensamento, como o presente em silogismos, um pensamento dedutivo. Conforme Dupuy afirma, Gödel havia em sua teoria demonstrado ser possível aritmetizar (CUTLAND, 1980, p.72) qualquer proposição numa linguagem formal, o que implica afirmar que "essa aritmetização da lógica oferece um fundamento rigoroso à máxima: raciocinar é calcular (sobre números inteiros)" (1995, p.29).

Turing ao publicar seu famoso artigo *On computable numbers, with an application to the Entscheidungsproblem* em 1936, não tinha intenção de resolver problemas de natureza filosófica relacionados à mente. Seu artigo se dedicava aos problemas lógicos expostos acima. Turing caracterizou a noção de calculabilidade efetiva, ou procedimento finito, em termos da noção de máquina, um conjunto finito de funções, que pudesse calcular qualquer operação lógica dentro de um sistema formal. Essa noção, a da *máquina de Turing*, ainda não estava ligada a de uma máquina física, um artefato; estava sim ligada a formalização da noção de máquina, um procedimento servil: "a formalização da noção intuitiva de um procedimento automático, submetido a regras fixas, totalmente controlável pela razão, que não implique nem sentido, nem interpretação, nem criatividade." (DUPUY, 1995, p.30).

Conforme aponta Benante (2001, p.17), o problema da decisão de Hilbert pode agora ser definido em termos da máquina desenvolvida por Turing: dada uma formula qualquer do cálculo dos predicados, existe uma máquina de Turing que permita determinar se essa fórmula é demonstrável ou não? Podemos entender o problema da decisão como um problema de previsibilidade, ou seja, se a demonstração ou não de uma fórmula pode ser prevista. Não havia meio de se verificar a demonstrabilidade de um teorema sem uma tentativa propriamente dita. De fato, a máquina de Turing chegou ao resultado já conhecido, uma resposta negativa ao problema de Hilbert: não existe uma máquina de Turing capaz de prever se um teorema é demonstrável sem efetivamente tentar demonstrá-lo. Ou seja, aparentemente a máquina de Turing tinha apresentado ser equivalente a humanos no tratamento de certos problemas, ou pelo menos tinha demonstrado a mesma limitação na impossibilidade de se prever a demonstração de um teorema, o que corrobora a hipótese de Gödel, e assim, inclui as funções recursivas na classe de funções calculáveis pela máquina de Turing.

Ao mesmo tempo, Alonso Church (1936) define uma classe de funções chamadas de *lambda-definíveis*, que abarcam todas as funções calculáveis mecanicamente. Church propõe uma tese, onde diz que toda função calculável é uma função recursiva. Já vimos que a classe

de funções recursivas é equivalente às funções calculáveis pela máquina de Turing. Portanto, o que se verifica é que a classe de funções lambda-definíveis é equivalente à classe de funções calculáveis pela máquina de Turing. Sendo as duas classes de funções equivalentes, a máquina de Turing, então, calcula qualquer função calculável mecanicamente. Chama-se essa classe de funções de funções Turing-computáveis (ou tese de Turing-Church).

Existe, até nossos dias, uma imensa discussão sobre o que é Turing-computável, ou se existem funções que não estariam na classe de funções Turing-computáveis. Biraben (1996, p.50) afirma: "O fato de todas as funções efetivamente calculáveis submetidas a teste serem lambda definíveis [ou turing-computáveis] não é evidência suficiente para pensar que toda função efetivamente calculável é lambda definível" (grifo do autor). Benante (2001, p.72) comentando a afirmação acima diz que se aceitarmos definições por indução a máquina de Turing é capaz de calcular todas as funções mecânicas.

1.2.5.2 Descrição formal da máquina de Turing.

Podemos explicar o modelo formal de Turing através de uma descrição ilustrativa, como muitos apontam. Segundo Turing (1936) a máquina é constituída de três partes:

1) uma fita potencialmente infinita para ambos os lados onde se grava e lê símbolos em espaços discretos, ou quadrados (na terminologia de TURING, 1936).

2) uma cabeça de gravação e leitura que se movimenta sobre essa fita, andando um espaço de cada vez para um dos dois lados.

3) uma unidade de controle da cabeça de gravação que opera sobre estados também discretos (no domínio do tempo). Esse controle é determinado por um conjunto finito de regras sobre o conjunto também finito de estados e símbolos.

Além disso, é necessário uma formalização desses estados, símbolos e regras:

a) um conjunto finito de estados internos ou condições (na terminologia de TURING, 1936) $C=\{c_1, c_2, \ldots, c_n\}$

b) um conjunto finito de símbolos, ou alfabeto $S=\{s_1, s_2, \ldots, s_n)$

c) um conjunto finito de instruções, regras, ou configurações (terminologia de TURING, 1936) $M=\{m_1, m_2, \ldots, m_n\}$

Temos ainda, que o subconjunto $\{c_i, c_f\} \in C$, onde c_i é o estado inicial e c_f é o estado final. Além disso, cada instrução $m_i \in M$ é dada na seguinte forma: $m_i=\{c_n, s_n, s_k, d_e, c_j\}$, onde c_n é o estado atual da máquina, s_n é o símbolo que se encontra no quadrado sob a cabeça de gravação-leitura, s_k é o símbolo que a máquina deve gravar no mesmo quadrado, d_e é a direção que a máquina deve mover a cabeça de gravação-leitura (existem apenas duas opções: direita e esquerda), e c_j que é o novo estado a que a máquina deve passar.

Então, a operação da máquina é determinada pelo conjunto M, que por sua vez pode ser descrito na forma de uma tabela. O que o programador faz, portanto, é criar esta tabela determinando o comportamento da máquina em cada situação discreta, estabelecendo anteriormente o conjunto de estados internos C e o alfabeto S. Quando a máquina atingir o estado final c_f diz-se que ela chegou a solução do problema proposto, podendo-se ler nos caracteres gravados na fita da máquina a solução. Se o problema não for solucionável em *n* passos finitos, a máquina não parará. Por exemplo, se programarmos a máquina para calcular o maior número primo possível, ela nunca parará de calcular, visto a memória (fita) ser potencialmente infinita.

Para cada tipo de problema deve-se determinar uma tabela de máquina específica. Chamemos de MT_n tal máquina. Podem existir inúmeras máquinas do tipo MT para inúmeros tipos de problemas. Turing afirma (1936, p.241) que tal máquina MT_n pode ser simulada numa máquina de Turing Universal. Uma MT Universal pode computar qualquer função que uma máquina do tipo MT_n computa, simulando-a. Para isso colocam-se os dados, estados internos, alfabeto e instruções de MT_n justapostos aos dados que devem ser computados na fita da MT Universal. A MT Universal seguindo suas instruções (a sua tabela) simulará uma maquina do tipo MT_n contida em sua fita e apresentará o resultado final.

Com esse tipo de máquina pode-se, como já vimos, computar-se todas as funções mecanicamente calculáveis.

1.2.6 A prática da modelagem na IA.

A proposta apresentada por Turing, a de que existe uma máquina, ainda que no plano abstrato ou puramente formal, que pode calcular todos os tipos de funções efetivamente calculáveis foi uma das descobertas de maior impacto numa nova geração de cientistas que estava surgindo nessa nova área com o nome de cibernética. Estes cientistas buscavam sistemas mecânicos capazes de auto-regulação, a partir da noção de *feedback*. Tratavam as máquinas com *feedback* como se elas tivessem metas e corrigissem seu desempenho para atingir tais metas. Wiener definiu cibernética como: "(…) *the entire field of control and communcication theory, whether in the machine or in the animal* (…)." (WIENER, 1948, p.11).

Pela citação acima podemos reparar que não havia a preocupação em distinguir entre máquinas artificiais e seres vivos, muito pelo contrário. Sem esse caráter não *biomórfico*, talvez a IA nem chegasse a ser fundada, pois o objetivo era habilitar as máquinas com propriedades típicas de sistemas biológicos, dirigir seu comportamento para a realização de um fim. Nessa época, pesquisadores como Lashley, McCulloch e Pitts estavam estabelecendo pela primeira vez uma correlação entre sistemas lógico formais e o sistema nervoso central, influenciados sem dúvida pelas idéias de Turing (GARDNER, 1995; DUPUY, 1995). Ao mesmo tempo, as idéias de Turing estavam guiando outros pesquisadores em uma direção alternativa: o desenvolvimento de um computador que pudesse calcular qualquer função como uma MT Universal.

A ciência moderna e contemporânea se caracteriza como instituição da modelagem, isto é, uma área onde modelos abstratos são gerados para explicar fenômenos naturais. Os modelos gerados podem ser intercambiados entre diversas teorias, do sistema solar ao átomo (DUPUY, 1995, p.24). A noção de funcionalismo científico é trazida aqui, pois um modelo científico pode ser sustentando em diversas realizações materiais. Dupuy afirma que:

> "(…) conhecer é produzir um modelo do fenômeno e efetuar sobre ele manipulações ordenadas. (…) As ciências cognitivas fazem desse modo o modo único de todo conhecimento. (…) Seja um sistema cognitivo material: cientista, homem, animal, organismo, órgão, máquina. O que faz que esse sistema conheça por modelos e representações deve ele próprio ser modelizado, abstraindo-se do substrato material, diferente a cada vez, o sistema de relações funcionais responsável pela faculdade de conhecer." (1995, p.27)

Nesse contexto de modelagem, a ciência cognitiva procura estabelecer um modelo da habilidade de modelar, isto é, da mente. Se é possível extrair da mente um modelo, este por sua vez pode ser implementado em outro sistema material, desde que mantenha suas propriedades funcionais. Se entendermos que as propriedades da mente, como o raciocínio matemático, podem ser formalizadas através de sistemas formais lógicos, aí está nosso modelo. Um modelo abstrato que funciona seguindo procedimentos formais e resolve teoremas como nós, humanos, às vezes até mais rapidamente, e que pode ser implementado em outro sistema, mesmo que este seja lápis e papel. Nesse sentido entra em cena o computador serial digital proposto por von Neumann.

A MT pode ser considerada, então como o modelo básico abstrato da mente, mas o modelo físico surge quando von Neumann, ao desenvolver o EDVAC em 1943, cria o computador serial digital (CSD). Como vimos brevemente, no EDVAC temos pela primeira vez um sistema que possui dois níveis de análise distintos: o *software* e o *hardware*. A analogia entre a MT e o CSD é clara: temos inicialmente a memória física do CSD que equivale à fita da MT, porém de capacidade finita; uma unidade lógica para realizar a computação; e uma unidade de controle que coordena o funcionamento interno.

Porém a grande diferença entre a máquina de Turing e o CSD relaciona-se a memória; von Neumann percebeu que poderia trabalhar apenas com uma memória. Na máquina de Turing temos a fita infinita e, além dela, a tabela da máquina onde ficam armazenadas as instruções, o alfabeto e conjunto de estados. Essa tabela, pois, também é uma memória. Existe a memória onde estão os dados (fita) e outra onde estão o código de operação e controle (tabela). Von Neumann, em sua máquina, colocou tanto dados quanto códigos numa mesma memória, apenas dividida em duas partes, como uma MT Universal simulando uma MT específica. Assim, a máquina de von Neumann é uma implementação física de uma MT Universal.

Porque, então, tomou-se como modelo da mente o CSD (ou uma MT Universal física)? Ora, com uma máquina de Turing específica tinha-se apenas um modelo específico (e lógico) de alguma propriedade racional humana. Agora, com o CSD temos uma máquina real (física) que pode operacionalizar qualquer procedimento específico, sendo necessário apenas alimentar sua memória com dados e códigos adequados para tal procedimento. A comparação com o cérebro parecia inevitável. O que faz um pesquisador quando quer solucionar um problema matemático? Ele define os dados envolvidos em seu problema, e estabelece uma

série de operações matemáticas, normalmente predefinidas, que manipulam os dados. Através de um processo formal, mecânico, ele manipula os dados inicias e chega ao resultado (CUTLAND, 1980, p.7), as vezes utilizando artefatos externos a ele, como um ábaco, por exemplo. O funcionamento de um computador serial digital é da mesma natureza, ele manipula os dados de acordo como um procedimento formal predefinido. No caso de um ábaco, temos a necessidade de um operador externo, no caso do computador, não. Ele é um sistema autômato, que quando inicia um processo só irá parar quando chegar ao estado final, o resultado, ou por uma falha de programação ou energia.

Tendo no CSD o modelo oficial da IA, os pesquisadores daquela primeira geração cibernética foram desenvolvendo vários programas computacionais na tentativa de se investigar aspectos normalmente específicos da mente humana, como a realização de cálculos, prova de teoremas lógicos e matemáticos, manipulação da linguagem natural, e assim por diante. Um exemplo é o GPS *(General Problem Solver*) desenvolvido por Newell e Simon (1972), um sistema que era capaz de solucionar vários tipos de problemas complexos, da demonstração de teoremas e solução de enigmas ao jogo de xadrez.

Programas como o GPS tentavam simular procedimentos humanos através de um inventário de operações lógicas codificado na forma de regras. O programa seguia às regras para a manipulação dos dados de entrada, gerando uma saída como resultado. A eficiência era uma questão de se ter problemas solucionáveis logicamente e um inventário rico o suficiente.

Uma das aplicações mais controversas na IA esteve relacionada à manipulação da linguagem natural. Vamos ao exemplo de Block (1980), referente a uma máquina hipotética capaz de reproduzir sentenças em linguagem natural. O conjunto de sentenças possíveis em uma língua é finito, visto termos quantidades finitas de fonemas e letras, porém muito grande. Existindo um fantástico time de pesquisadores, que dentre esse conjunto finito de possibilidades, separem as frases que fazem sentido num subconjunto (*smart speakable strings*), tal subconjunto pode ser armazenado numa máquina que pode gerar uma conversação através de procedimentos randômicos baseados em palavras-chave (um método clássico na IA). Block diz a esse respeito:

> "*Now, if the team has been thorough and imaginative in listing the smart speakable strings, this machine would simulate human conversational abilities. Indeed, if the team did a brilliantly creative job, the machine's conversational abilities might be superhuman (though if it is to "keep up" with current events, the job would have to be redone often). But this machine clearly has no mental states at all. It is just a huge list-searcher plus a tape recorder.*" (1980, p.282) (aspas do autor)

Um dos argumentos que levam Block, criticando o funcionalismo, a afirmar a não existência de estados mentais nesta máquina é que se nossos estados mentais são causalmente dependentes de nossos estados psicológicos e/ou neurofisiológicos, tal máquina não pode realmente apresentar estados mentais pois não é equivalente a nós nestes dois últimos níveis (1980, pp.282-283). Nesse mesmo contexto, Searle (1980) irá questionar as molelagens computacionais, principalmente algumas relacionadas à linguagem natural, por outro aspecto (simulação versus realização). Vamos antes ilustrar nossa discussão com outra famosa implementação.

Por volta de 1970 é criado por Winograd um programa batizado de SHRDLU (o nome decorre das letras, da sétima à décima segunda, mais recorrentes de uma impressora) que pode compreender linguagem natural, ainda que num domínio muito limitado (GARDNER, 1995, p.173). O programa demonstra a compreensão de expressões pelo fato de que, recebendo um comando ambíguo, pede esclarecimentos. SHRDLU é um programa que age sobre um universo limitado onde estão situados alguns objetos geométricos. O interlocutor deve pedir para o programa executar algumas ações com tais objetos através da linguagem natural. Para compreender e estabelecer diálogos coerentes o programa é construído sobre vários módulos (em linguagem lisp) que atuam sob a supervisão de uma unidade de controle. Ele possui um dicionário constituído de duas partes, uma sintática e uma semântica, onde cada uma delas é processada por módulos diferentes. Esse sistema permite ao SHRDLU segmentar as frases determinando os substantivos, adjetivos e verbos para compreender que ação deve executar e sobre qual objeto. Existe um módulo responsável pela organização espacial dos objetos uns em relação aos outros, e, além disso, existem outros módulos responsáveis por: resolução de problemas de maneira dedutiva, determinação de frases como declarativas, imperativas ou interrogativas, geração de respostas. Abaixo uma ilustração de sua arquitetura:

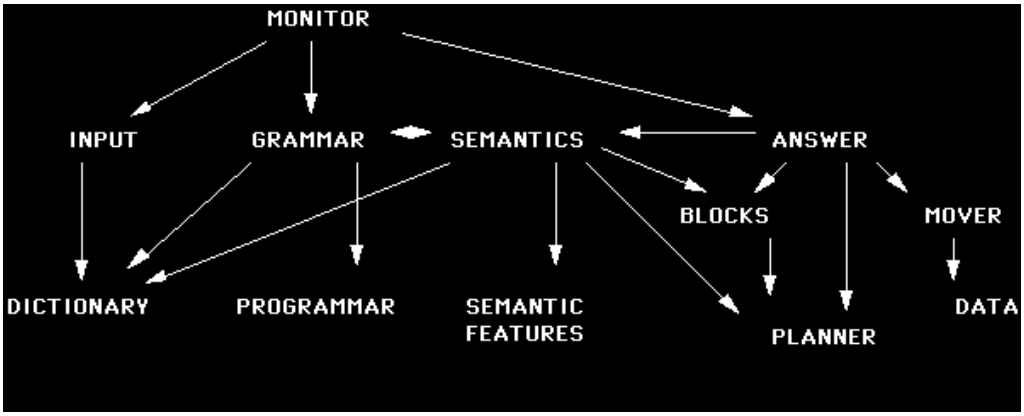

Fig. 1.1. Arquitetura do SHRDLU.

O display para o usuário é mostrado abaixo:

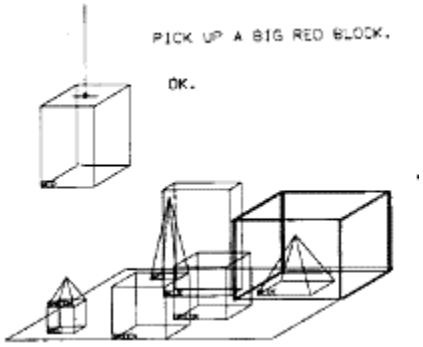

Fig. 1.2. *Display* do usuário.

Agora, transcrevemos aqui trechos de uma execução do programa (passos numerados, com computador em caixa-alta) (In Winograd website):

p1 *pick up a red block*
OK
p2 *grasp the pyramid*
I DON'T UNDERSTAND WHICH PYRAMID YOU MEAN
p 3 *find a block which is taller than the one you are holding and put it in the box*
BY "IT" I ASSUME YOU MEAN THE BLOCK WHICH IS TALLER THAN THE ONE I AM HOLDING OK
p4 *what does the box contain?*
THE BLUE PYRAMID AND THE BLUE BLOCK
p5 *what is the pyramid supported by?*
THE BOX
p6 *how many blocks are not in the box?*
FOUR OF THEM
p7 *is at least one of them narrower than the one which I told you to pick up?*
YES THE RED CUBE
[…]
p10 *can a pyramid be supported by a block?*
YES
p11 *can a pyramid support a pyramid?*

I DON'T KNOW

[…]

p22 *how many things are on top of green cubes*
I'M NOT SURE WHAT YOU MEAN BY " ON TOP OF" IN THE PHRASE "ON TOP OF GREEN CUBES"
DO YOU MEAN
1 DIRECTLY ON THE SURFACE
2 ANYWHERE ON TOP OF?

[…]

p26 *why did you do that?*
TO CLEAN OFF THE RED CUBE
p27 *why did you clear off that cube?*
TO PUT IT ON A LARGE GREEN CUBE
p28 *why did you do that?*
TO STACK UP THE RED CUBE A LARGE RED BLOCK AND A LARGE GREEN CUBE
p29 *why did you do that?*
BECAUSE YOU ASKED ME TO DO IT
p30 *how did you do it?*
BY PUTTING A LARGE RED BLOCK ON THE TABLE THEN LETTING GO OF IT THEN PUTTING A LARGE GREEN CUBE ON IT THEN LETTING GO OF THAT CUBE THEN PUTTING THE RED CUBE ON THAT CUBE THEN LETTING GO OF THAT CUBE
p31 *how many objects did you touch while you were doing it?*
FOUR OF THEM
p32 *what did the red cube support before you started to clean it off?*
THE GREEN PYRAMID
p33 *there were five blocks to the left of the box then*
NO ONLY FOUR OF THEM THE RED CUBE TWO LARGE GREEN CUBES AND A LARGE RED BLOCK

[…]

p36 *does a steeple*
SORRY I DON'T KNOW THE WORD STEEPLE
p37 *a "steeple" is a stack which contains two green cubes and a pyramid*
I UNDERSTAND
p38 *are there any steeples now?*
NO
p39 *build one*
OK

[…]

Certamente o diálogo acima é um exemplo interessante, e algumas características podem nos surpreender, dando realmente a impressão que SHRDLU compreende o que está fazendo. Vamos analisar algumas passagens dessa transcrição: no passo 10, por exemplo o sistema dedutivo é responsável pela resposta de que um bloco pode suportar uma pirâmide; no passo 11 o sistema dedutivo não encontra uma solução e o *output* "eu não sei" é gerado; em 22, existindo ambigüidade, questões são geradas para esclarecimento. No passo 36, a entrada contém uma palavra desconhecida (fora do banco de dados) e o sistema pede uma definição, podendo então executar a ação. Agora cabe a recorrente pergunta: Passaria o SHRDLU no teste de Turing? Ou, mesmo passando, o programa tem alguma compreensão dos *inputs* e *outputs*, ou do que são objetos, relações entre objetos e suas ações sobre eles?

Vamos ver o que Searle tem a nos dizer sobre estes tópicos. Em seu artigo de 1980 Searle elenca sérias críticas às modelagens da IA, inclusive apontando SHRDLU (na verdade Winograd) como um dos criticados. Mas, inicialmente, vamos estabelecer uma diferença postulada pelo mesmo Searle: IA fraca e IA forte. Por IA fraca entendemos o computador como uma *ferramenta* no estudo da mente, utilizada para formular e comprovar hipóteses de modo mais rigoroso e preciso, nas palavras do autor. Por ouro lado, para a IA forte os computadores devidamente programados possuem estados mentais, são uma mente. Como diz Searle:

> "*En la IA fuerte, como la computadora programada cuenta con estados cognoscitivos, los programas no son meras herramientas que nos permiten probar las explicaciones psicológicas, sino que los programas constituyen por sí mismos las explicaciones.*" (1980, p.82)

A crítica que o autor dirige a IA forte é voltada principalmente aos programas desenvolvidos com o intuito de manipularem aspectos semânticos da linguagem natural. Esse foi realmente o ponto de maior impacto nas discussões sobre o alcance cognitivo das implementações computacionais, e por conseqüência, da capacidade e do alcance explicativos do modelo computacional da mente. Nos programas desenvolvidos para a manipulação de dados formais, isto é, sem implicação semântica, a questão da compreensão de significados (pelo programa) simplesmente não existia. No entanto, quando sistemas artificiais tinham supostamente a capacidade de compreender linguagem natural, interpretar estórias e gerar frases, a relação entre manipulações sintáticas e entendimento semântico veio à tona. Block negou a tais programas, mesmo àqueles que passassem no teste Turing, estados mentais pela ausência de correlatos psicológicos e/ou neurofisiológicos, mas Searle irá por outro caminho. Ele propõe um contra-exemplo ao teste de Turing: *o quarto chinês* (SEARLE, 1980).

Imagine uma pessoa, Johnny por exemplo, isolado num quarto, sem nenhum contato com o mundo exterior a não ser através de folhas de papel escritas em chinês. Porém Johnny não conhece a língua chinesa. Para ele os ideogramas são apenas símbolos gráficos sem nenhum significado impressos num suporte físico, nesse caso papel. Inicialmente, Johnny recebe um calhamaço de folhas escritas em chinês. Johnny também recebe um outro calhamaço escrito em inglês, língua esta que ele compreende totalmente, que contém regras que lhe permitem relacionar o segundo calhamaço com o primeiro. Essas regras apenas mostram a ele como relacionar símbolos pela sua forma. Suponha, agora, que Johnny recebe uma terceira resma com caracteres em chinês e regras novamente em inglês, que lhe

permitem correlacionar elementos desta terceira resma com as duas primeiras, e que ainda o instruam como responder com certos símbolos chineses a certos caracteres escritos nesta terceira resma. Mas sem que Johnny saiba, as pessoas que lhe enviam os calhamaços chamam de *alfabeto* o primeiro, de *dados* o segundo e de *perguntas* o terceiro. Além disso, chamam de *respostas* os papeis que ele envia respondendo ao terceiro calhamaço, e de *programa* o conjunto de regras em inglês que lhe foi enviado. Contudo, "somente para complicar um pouco a estória" (SEARLE, 1980, p.84), imagine ainda que estas pessoas enviam a Johnny relatos agora em inglês, e lhe fazem perguntas em inglês para serem respondidas em inglês, que sabemos que ele compreende. Suponhamos que, passado um tempo, Johnny se torne extremamente habilidoso para responder as perguntas em chinês, enquanto que os programadores se tornem extremamente habilidosos para escreverem os programas, a tal ponto que um observador externo (fora do quarto) que fale chinês e inglês não possa distinguir entre as respostas de Johnny e as de um nativo falante do chinês para os dados e perguntas que são enviadas. E também, que as respostas em inglês que Johhny envia para as perguntas, também em inglês, são indistinguíveis das de outro falante da língua inglesa. Para o observador externo, então, as respostas em chinês são igualmente boas perante as resposta em inglês. No entanto, a diferença é que as respostas em chinês são geradas mediante a manipulação de símbolos formais não interpretados, cujo conteúdo semântico não é apreendido e, nesse caso, Johnny se comporta simplesmente como um computador.

Searle relaciona seu exemplo a duas afirmações dos partidários da IA:

> "*1) puede decirse literalmente que la máquina comprende el relato y proporciona respuestas a las preguntas, y*
> *2) lo que la máquina y su programa hacen es explicar la capacidad humana de comprender el relato y responder preguntas acerca de él*." (1980, p.83)

Para o primeiro caso, Searle responde (1980, p.85) que apesar de suas respostas serem indistinguíveis para um falante de chinês, ele continua sem entender um só símbolo que manipula e, pelo mesmo motivo, computadores programados para lidarem com linguagem natural (vide SHRDLU acima) também não compreendem as entradas e nem as saídas que geram. Quanto ao segundo, responde que tomar o computador como modelo da mente é supor que seres humanos "são exemplos concretos de programas" (1980, p.85), operam sobre operações puramente formais. Qualquer tipo de procedimento formal que um computador realize também pode ser realizado por um ser humano, como Johnny, mas não garantem que ele tenha entendimento sobre os dados que manipula. Manipulações formais da linguagem

natural operam sobre aspectos sintáticos, deixando de lado informações semânticas, e podemos, mesmo que intuitivamente, postular que compreendemos este tipo de informação (SEARLE, 1980, p.86). Na verdade tudo depende de como entendemos a relação sintaxe-semântica.

Numa linguagem artificial é possível que características da estrutura sintática de fórmulas correspondam sistematicamente a suas características semânticas (*"the syntax of a formula encodes its meaning."* FODOR e PYLYSHYN, 1988, p.28) , e todas as linguagens artificiais, como as da lógica, podem ter esta propriedade, mas a linguagem natural não, ou pode apenas parcialmente (FODOR e PYLYSHYN, 1988, p.28). Os autores ainda afirmam:

> "*If, in principle, syntactic relations can be made to parallel semantic relations, and if, in principle, you can have a mechanism whose operations on formulas are sensitive to their syntax, then it may be possible to construct a syntactically driven machine would be just what's required for a mechanical model of the semantical coherence of thought; correspondingly, the idea that the brain is such a machine is the foundational hypothesis of classical cognitive science*." (FODOR e PYLYSHYN 1988, p.30)

Se aceitarmos tal relação entre um símbolo e seu significado podemos postular uma IA forte como sendo plausível. Mas se, por outro lado, a sintaxe não for, sozinha, suficiente para abarcar conteúdos semânticos, Searle parece ter razão. A semântica da linguagem natural parece ser, antes de mais nada, uma codificação determinada socio-históricamente de maneira não totalmente precisa. Nesse sentido, a grande questão parece ser a diferença na sistematicidade da convenção entre símbolo e significado para linguagens artificiais em relação às naturais.

Por fim, Searle (1980, p.87) postula que o computador está para a mente numa relação metafórica ou analógica, "mas que nada se prova com isto". Não existe um modelo computacional da mente, e sim uma metáfora computacional, e, na sua perspectiva, metáforas e analogias não têm poder explicativo. Mesmo hoje em dia este embate não está encerrado, porém, apesar de sérias e relevantes, as críticas de Searle (Fodor, Pylyshyn e outros) não impediram a IA (principalmente pelo viés fraco) de se desenvolver – e a deixaram menos ingênua, talvez.