

Anexo D

Glossário

“Quando eu uso uma palavra”, replicou Humpty Dumpty com superioridade, ‘ela significa o que eu quero que signifique... e nada mais’.”

Lewis Carroll (1832-98), *Through the Looking-Glass* (1872)

Obs.: No final de cada definição adotada, encontra-se o capítulo (C) ou anexo (A) no qual esta definição aparece pela primeira vez indicada em negrito.

- ✓ **abordagem conexionista** – abordagem que usa redes neurais artificiais [C1].
- ✓ **abordagem simbólica** – abordagem que usa símbolos para representação de conhecimento. É baseada na lógica. Sistemas simbólicos usam paradigma oposto ao conexionismo. São sistemas baseados em regras, casos ou princípios (Pereira e Grosz, 1993) – Veja apêndice B [C1].
- ✓ **ambigüidade** (lexical) – a mesma palavra, ou expressão básica, denota entidades diferentes no mundo. Por exemplo, *manga* pode ser a fruta ou a parte da camisa [C2].
- ✓ **antecedente** – elemento a esquerda da implicação lógica. Por exemplo, o A em $A \rightarrow B$ [C5].
- ✓ **antecedente ponderado** – antecedente de uma regra de produção com um peso associado, que na abordagem híbrida, significa o peso de conexão associado a essa regra [C3].
- ✓ **arquitetura conexionista** – rede de unidades processadoras (processadores elementares) altamente interconectada. Também conhecida como rede neural artificial [C3].
- ✓ **árvore de parsing** – árvore na qual os nós interiores representam frases, arcos representam aplicações de regras gramaticais e nós folha representam palavras. É uma forma de mostrar o resultado de uma análise sintática [C2].
- ✓ **ativo** – elemento da lógica de três valores, correspondente ao 1 (verdadeiro) [C2].
- ✓ **autômato de estados finitos** – dispositivo de computação que julga a legalidade de expressões lineares a direita sobre um alfabeto finito. Este tipo de linguagem é gerada pelas gramáticas lineares a direita. Uma gramática é *linear a direita* se toda produção é da forma $A \rightarrow bC$ ou $A \rightarrow b$, onde A e C são símbolos não terminais (variáveis) e b é um símbolo terminal ou a cadeia vazia. Uma linguagem gerada por tal gramática é chamada de *linguagem linear a direita*. (Moll *et al.*, 1988) [C5].
- ✓ **backpropagation** (retropropagação) – algoritmo conexionista (procedimento computacional de simulação de uma rede neural artificial) supervisionado, onde ocorre a propagação do padrão da camada de entrada em direção à camada de saída e depois há a comparação com a saída desejada e o erro é propagado de volta à camada escondida e à camada de entrada (*error backpropagation*, Rumelhart *et al.*, 1986a) [C2].

- ✓ **backward** – sistema de dedução baseado em regras, cujo encadeamento inferencial inicia-se com a expressão meta (conclusão), aplica-se as regras e chega-se na expressão fato (base de conhecimento) [C2].
- ✓ **bit** – dígito binário. No vetor de bits, cada elemento pode representar apenas dois valores [C2].
- ✓ **bottom-up** – técnica de análise, que consiste em partir dos símbolos terminais (que numa sentença corresponde às palavras), e continuar construindo os símbolos não terminais (por exemplo, SN, SV, etc.) até chegar no símbolo inicial da sentença (S) [C2].
- ✓ **camada de entrada** – camada inicial da rede neural artificial, onde os padrões de entrada são apresentados à rede [C2].
- ✓ **camada de saída** – camada final da rede neural artificial, de onde os padrões da saída são fornecidos pela rede [C2].
- ✓ **camada escondida** – também chamada de camada oculta, camada intermediária entre a entrada e a saída em uma rede neural artificial. Pode haver mais de uma camada escondida [C2].
- ✓ **ciclo de treinamento (ou de ativação)** – intervalo de tempo entre a apresentação do padrão à camada de entrada da rede neural artificial, a sua propagação através da camada escondida com direção à camada de saída da rede, e nos algoritmos supervisionados, a retropropagação (*backpropagation*) do erro de volta a camada de entrada. O treinamento de uma rede neural usualmente requer muitos ciclos de ativação [C2].
- ✓ **cluster** – grupo de unidades [AA].
- ✓ **complementaridade** – dois ou mais valores são complementares se somente um destes valores tem um sinal e todos os outros têm sinais opostos a este [C5].
- ✓ **computador (máquina) de von Neumann** – o computador atual, baseado no conceito de programa armazenado de von Neumann, composto de Unidade Central de Processamento, Unidades de Entrada e Saída e Memória [C3].
- ✓ **conexionismo** – veja abordagem conexionista [C1].
- ✓ **conseqüente** – elemento a direita da implicação lógica. Por exemplo, o B em $A \rightarrow B$ [C3].
- ✓ **desativado** – elemento da lógica de três valores, correspondente ao 0 (falso) [C2].

- ✓ **estrutura-D** – do inglês *Deep structure*, estrutura profunda. A gramática transformacional consiste de um *componente base* e de um *componente transformacional*. O componente base é uma gramática livre de contexto (ver hierarquia de Chomsky), que produz um conjunto de árvores de estruturas profundas (Grishman, 1986) [C4].
- ✓ **estrutura-S** – do inglês *Surface structure*, estrutura de superfície. A gramática transformacional consiste de um *componente base* e de um *componente transformacional*. O componente transformacional é um conjunto de regras de rescrita de árvore que, quando aplicado a uma árvore estrutura profunda, produz uma ou mais árvores de estrutura de superfície (Grishman, 1986) [C4].
- ✓ **feedforward** – rede onde a propagação do sinal caminha da camada de entrada para a camada de saída, passando pela(s) camada(s) escondida(s) [C2].
- ✓ **forward** – sistema de dedução baseado em regras, cujo encadeamento inferencial inicia-se com a expressão fato (base de conhecimento), aplica-se as regras e chega-se na expressão meta (conclusão) [C3].
- ✓ **frame** – coleção de atributos, em geral chamados de escaninhos (*slots*) e valores a eles associados (e possivelmente restrições a estes valores) que descrevem alguma entidade do mundo (Rich e Knight, 1994) [C2].
- ✓ **generalização** – característica notável das redes neurais artificiais, que consiste em permitir, mesmo com entradas não previamente vistas, mas próximas aos padrões conhecidos, que a rede neural forneça uma saída satisfatória [C1].
- ✓ **grade temática** – estrutura com todos os papéis temáticos atribuídos pelo predicado aos seus argumentos em uma sentença [C1].
- ✓ **gramática de cláusulas definidas** – formalismo computacional criado por Pereira e Warren (1980), para a tradução da gramática livre de contexto na lógica de predicados de primeira ordem [C2].
- ✓ **gramática livre de contexto** – gramática formal na qual o lado esquerdo de toda regra de produção contém apenas um símbolo não terminal (variável) [C1].

- ✓ **gramática sintagmática** – gramática que usa sintagmas como símbolos não terminais. Sintagmas são frases que compõem a sentença. Em *O menino quebrou o vaso*, *o menino* é o sintagma nominal sujeito, *o vaso* é o sintagma nominal objeto, e *quebrou o vaso* é o sintagma verbal. Exemplo de regra da gramática sintagmática: $S \rightarrow SN, SV$, ou seja, a sentença (S) é formada por um sintagma nominal (SN) seguido de um sintagma verbal (SV) [C2].
- ✓ **híbrida** – abordagem que usa rede neural baseada em conhecimento [C1].
- ✓ **hierarquia de Chomsky** – Noam Chomsky (1957) descreve quatro classes de formalismos gramaticais que diferem apenas na forma das regras de produção. As classes podem ser arranjadas numa hierarquia:
 - gramáticas regulares (tipo 3): as mais simples, equivalentes às máquinas de estados finitos, em poder expressivo. Toda regra tem um único símbolo não terminal do lado esquerdo e um terminal, seguido opcionalmente por um não terminal, por exemplo, $S \rightarrow a S$.
 - gramáticas livres de contexto – GLCs (tipo 2): o lado esquerdo consiste de um único símbolo não terminal. Portanto, cada regra licencia a rescrita do não terminal como o lado direito em *qualquer* contexto. As GLCs são populares para gramáticas da língua natural, apesar de se saber que pelo menos algumas línguas naturais não são livres de contexto (Pullum, 1991). Exemplo: $S \rightarrow a S b$.
 - gramáticas sensíveis ao contexto (tipo 1): o lado direito deve conter no mínimo a quantidade de símbolos do lado esquerdo. O nome *sensível ao contexto* vem do fato de que uma regra como $ASB \rightarrow AXB$ diz que um S pode ser rescrito como um X no contexto de um A precedente e um B seguinte.
 - gramáticas enumeráveis recursivamente (tipo 0): regras irrestritas. São equivalentes às máquinas de Turing em relação ao seu poder expressivo. Por exemplo, $AB \rightarrow C$ [C1].
- ✓ **indeterminado** – elemento da lógica de três valores, correspondente ao 0.5 [C5].
- ✓ **léxico** – vocabulário [C1].
- ✓ **localista** – tipo de representação em sistemas conexionistas, onde os conceitos (ou itens de informação) são representados por unidades da rede [C3].

- ✓ **lógica de predicados** – ou lógica de primeira ordem, ou ainda, cálculo de predicados de primeira ordem pode ser caracterizada como um sistema formal apropriado a definição de teorias do universo de discurso da matemática. Por exemplo, se H é um predicado que significa homem e M , mortal, para a sentença “Todos os homens são mortais”, teríamos a seguinte fórmula da lógica: $\forall x (H(x) \rightarrow M(x))$, que pode ser lida assim: “Para todo x , se x é homem, então x é mortal” [C2].
- ✓ **lógica de três valores** – lógica proposta por Lukasiewicz (1929), que admite três valores: *falso*, representado pelo 0, *verdadeiro*, representado pelo 1 e *indeterminado*, representado pelo 0.5. No caso do sistema HTRP, o valor *indeterminado* (0.5) é usado para representar palavras ambíguas [C4].
- ✓ **máquina de Turing** – controle de estado finito equipado com um dispositivo de armazenamento externo na forma de uma fita finita dividida em quadrados que pode ser estendida indefinidamente em ambas as direções. Cada quadrado da fita pode estar em branco ou pode carregar qualquer símbolo de um alfabeto de fita finita. O controle do estado finito é acoplado à fita através de uma cabeça de leitura/escrita. A qualquer instante dado, a cabeça estará percorrendo um quadrado da fita e o controle de estado finito estará em um estado. Dependendo deste estado e do símbolo no quadrado percorrido pela cabeça da fita, a máquina irá, em um passo, fazer o seguinte:
 - Entrar em um novo estado do controle de estado finito;
 - Sobrescrever um símbolo no quadrado percorrido (é possível sobrescrever o mesmo símbolo e portanto deixar o quadrado da fita sem mudar, ou sobrescrever um branco e “apagar” o símbolo da fita);
 - Deslocar a cabeça para esquerda ou para a direita um quadrado, ou não deslocar nada [C3].
- ✓ **microcaracterísticas semânticas** – tipo de representação que consiste no uso de vetores de elementos da lógica de três valores para representar as palavras (representação distribuída). Por exemplo, *homem é humano, grande, animado*, etc. Esta representação é usada por Waltz e Pollack (1985), McClelland e Kawamoto (1986), Rosa (1997) e outros [C1].
- ✓ **papel de caso** – termo usado por Fillmore (1968) para designar relacionamento semântico entre um predicado e seus argumentos. Também conhecido como *papel de caso temático* ou *papel temático* [C2].

- ✓ **papel temático** – na visão tradicional, trata-se do relacionamento semântico entre palavras de uma sentença. Normalmente o verbo (predicado) atribui papéis temáticos (AGENTE, PACIENTE, INSTRUMENTO, etc.) às outras palavras (argumentos) da sentença. Para Franchi e Cançado (1998), o papel temático representa uma função que o argumento tem em uma dada sentença, estabelecida através da relação semântica desse elemento com um predicado [C1].
- ✓ **parsing** – processo de atribuir categoria sintática (substantivo, verbo, etc.) a cada palavra de uma sentença e agrupá-las para formar a sentença. Derivado do latim, *pars orationis* ou parte do discurso [C3].
- ✓ **PDP (Processamento Distribuído Paralelo)** – abordagem conexionista [C1].
- ✓ **peso de conexão** – número real associado à ligação entre unidades em uma arquitetura conexionista [C1].
- ✓ **pesos finais** – pesos de conexão obtidos através da rede neural artificial no final do processo de treinamento [C5].
- ✓ **pesos iniciais** – pesos de conexão atribuídos a rede neural artificial no início do processo de treinamento [C1].
- ✓ **porta (gate)** – porta lógica equivalente, usada para avaliar tamanho de circuitos integrados, normalmente equivalente a uma porta *and* de duas entradas [AA].
- ✓ **processador elementar** – unidade de uma arquitetura conexionista que simula o neurônio do córtex cerebral. Tem muitas entradas, ponderadas pelos pesos de conexão, e uma saída. A saída só será ativada se a somatória ponderada das entradas atingir um certo limiar de ativação característico [C5].
- ✓ **processamento de linguagem natural (PLN)** – processamento (computacional) da língua natural, como o português, inglês, etc. É a habilidade de um computador em processar a mesma linguagem que os humanos usam no dia-a-dia. Os sistemas que fazem PLN devem incorporar a noção de armazenamento em computador e manipulação de dados lingüísticos [C1].

- ✓ **Prolog** – *programação em lógica* – linguagem de programação baseada na lógica de predicados de primeira ordem, onde uma implicação da lógica $A \rightarrow B$ se transforma na cláusula do Prolog $B :- A$. Na verdade, o Prolog puro trabalha apenas com cláusulas de Horn, que são cláusulas (disjunção de literais) com no máximo um literal positivo. *Disjunção* é o conectivo lógico binário que representa o *ou* (ou lógico): por exemplo, $(A \vee B)$ é verdadeiro se A verdadeiro ou B é verdadeiro, ou ambos. *Literal* é fórmula atômica, ou seja, uma fórmula não decomponível da lógica: por exemplo, a fórmula $(\neg A \wedge B) \rightarrow C$ é composta das seguintes fórmulas atômicas: $\neg A$, B e C. *Literal positivo* é a fórmula atômica sem o sinal de complemento, no exemplo acima, B e C são literais positivos [C2].
- ✓ **reconhecimento** – fase do algoritmo conexionista, onde é apresentada à camada de entrada da rede um padrão e espera-se que a rede responda em um único ciclo de ativação, qual é a classificação aprendida para este padrão [C2].
- ✓ **rede auto-associativa** – arquitetura conexionista na qual o padrão de entrada é associado com ele mesmo. Sua meta é completar o padrão. Quando uma porção do padrão de entrada é apresentada, o restante do padrão será preenchido ou completado. Isto é similar a associação simples de padrões, exceto pelo fato de que o padrão de entrada tem a dupla função de entrada mestre e do padrão a ser associado (Rumelhart *et al.*, 1986b) [C3].
- ✓ **rede conexionista** – rede neural artificial [C1].
- ✓ **rede neural artificial** – arquitetura conexionista [C1].
- ✓ **rede neural baseada em conhecimento** – arquitetura conexionista que permite a inclusão e a extração de conhecimento simbólico. Combinação de abordagens simbólica e conexionista, também conhecida como abordagem híbrida [C3].
- ✓ **rede recorrente** – rede neural artificial, onde há ligação entre a saída e a entrada. Este tipo de arquitetura é usada para tratamento de eventos temporais, veja por exemplo, a rede de Jordan (Jordan, 1986) no apêndice A – As Redes Neurais Artificiais [C2].
- ✓ **rede multicamadas** – rede neural artificial com várias camadas [C3].
- ✓ **regra de produção** – regra simbólica [C1].
- ✓ **regra de rescrita** – regra de produção [C1].
- ✓ **regra gramatical (simbólica)** – regra de produção (implicação lógica) baseada na lógica de predicados, como por exemplo, $(A \wedge B) \rightarrow C$, que se lê como “Se A e B são verdadeiros então C é verdadeiro” [C1].

- ✓ **regra simbólica com antecedentes ponderados** – regra de produção, onde cada elemento do antecedente da implicação lógica está ponderado, como por exemplo, $(0.8 A \wedge 1.3 B) \rightarrow C$, que se lê como “Se A, com peso 0.8, e B, com peso 1.3, são suficientes para ativar a unidade, então C é verdadeiro” [C5].
- ✓ **representação distribuída** – tipo de representação para as palavras em alguns sistemas de processamento de linguagem natural. Consiste na utilização de várias unidades para a representação de uma palavra. O termo ‘distribuído’ é usado aqui da mesma forma que em McClelland e Kawamoto (1986) e não como em van Gelder (1992), que redefine o conceito de ‘representação distribuída’. No caso do sistema HTRP, são usadas vinte unidades (microcaracterísticas semânticas) para representar uma palavra [C2].
- ✓ **restrições de seleção** – restrições que os predicados colocam em tipos de entidades que seus argumentos devem denotar. Pode-se dizer que uma sentença com *gostar* não apenas acarreta mas também pressupõe que o seu sujeito deva ser animado. A sentença “A raiz quadrada de dois não gosta de Ronaldinho” é portanto semanticamente anômala (Chierchia e McConnel-Ginet, 1990). Esta restrição pode ser vista como uma reflexão do papel temático atribuído por *gostar* ao seu sujeito. Portanto, a definição de papel temático pode também envolver significado pressuposicional. Segundo Katz e Fodor (1963), as restrições de seleção são mecanismos para tratar da má formação semântica. Estas restrições são associadas às palavras e listadas no léxico, informando que, para pares de palavras, as palavras podem combinar semanticamente uma com a outra em várias construções gramaticais (Lyons, 1995) [C2].
- ✓ **script** – descrição de evento genérica, particularmente na área de entendimento de linguagem natural. Também conhecido como esquema [C3].
- ✓ **supervisionado** – tipo de algoritmo conexionista, no qual são fornecidos à rede pares de entrada-saída. A cada ciclo de ativação, um padrão de entrada é fornecido, ocorre a propagação deste padrão em direção à camada de saída e então esta saída real é comparada com a saída fornecida. A diferença é propagada de volta, descontada dos pesos de conexão. O exemplo mais típico do algoritmo supervisionado é o algoritmo *backpropagation* [C5].

- ✓ **tolerância a falhas** – importante propriedade das redes neurais artificiais, que consiste em permitir, mesmo com entradas incompletas, que a rede neural forneça uma saída satisfatória. É a representação distribuída da sua arquitetura que possibilita esta propriedade [C1].
- ✓ **treinamento** – fase do algoritmo conexionista, na qual padrões são apresentados à rede. Esta fase pode ter muitos ciclos de ativação [C2].
- ✓ **unidade *and*** – unidade na arquitetura conexionista que só será ativada se todas as entradas estiverem presentes [C3].