Graduação em Engenharia da Computação

“Falsa Memória: Modelos Neurais baseados na Teoria do Rastro Difuso.”

Por

Julio Domingues Ferraz

Trabalho de Graduação

Universidade Federal de Pernambuco
secgrad@cin.ufpe.br
www.cin.ufpe.br/~secgrad

Recife, Fevereiro/2013

Universidade Federal de Pernambuco
Centro de Informática
Graduação em Engenharia da Computação

Julio Domingues Ferraz

"Falsa Memória: Modelos Neurais baseados na Teoria do Rastro Difuso."

Trabalho apresentado ao Programa de Graduação em Engenharia da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Engenharia da Computação.

Orientador: *Aluizio Fausto Ribeiro Araújo*

Recife, Fevereiro/2013

Eu dedico este trabalho aos meus pais, Bruno e Martha, como fruto do investimento em minha educação e de todo o amor a mim conferido.

Agradecimentos

Agradeço aos meus pais, Bruno e Martha, pela eterna confiança, pelo exemplo de integridade e humildade, pelo carinho e dedicação com que me investiram os mais importantes ensinamentos, pelos inabaláveis apoio e incentivo, e, especialmente, pelo incondicional e protetor amor.

A minha prezada irmã, Victória, pelo amor, apoio e amizade.

Aos meus avós, Marcos, Cezinha, Carlos Antônio (*in memorian*) e Judith, pela ternura e sabedoria que concederam à minha educação, pela integridade que me servirá eternamente de exemplo, e pelo imenso amor no qual sempre poderei abrigar-me.

Aos meus tios, pela paciência, gentileza e incentivo, e pela certeza de sua presença nos momentos mais importantes.

Aos meus primos, pela eterna amizade, entusiasmo e alegria que me trazem.

Aos meus amigos de vida e da jornada acadêmica, pelo compartilhamento de nossas angústias e incertezas, e, em especial, a Artur, Augusto, Breno, David, Héctor, Júlio, Luís, Marcelo e Pedro, pela cumplicidade de nossos sonhos.

Ao meu orientador, Aluízio Araújo, pela paciência, entusiasmo e dedicação com que me ensinou, guiou e apoiou, e pelo decisivo incentivo à pesquisa.

Ao avaliador, Germano Crispim Vasconcelos, pelo apoio e a atenção dedicada à análise deste trabalho.

Aos amigos de pesquisa, André, Hansenclever, Flávia, Orivaldo, Tiago e Daniel, pela alegria, gentileza e paciência com que me acolheram.

Aos professores do Centro de Informática, pela contribuíção à minha formação na graduação.

E aos funcionários do Centro de Informática, pela diligência com que mantiveram e mantêm o nosso prezado ambiente de estudo.

Resumo

A Falsa Memória consiste de um processo involuntário de adição, omissão ou troca de informações relativas a uma ou mais experiências passadas. O estudo deste fenômeno é fundamental para a compreensão do funcionamento da memória humana, suas capacidades e limitações. Nesse sentido, situado no campo da Inteligência Computacional, este trabalho visa implementar um sistema modular de redes neurais artificiais (ANNs) capaz de simular o comportamento humano em testes cognitivos que induzam a falsificação de memória. Para isso, propõe-se a revisão, adaptação e re-implementação do sistema concebido por Pacheco (2004), considerando-se fundamentos atuais da neurofisiologia e, principalmente, da psicologia, através da Teoria do Rastro Difuso (FTT). Para validação do modelo, foram simulados experimentos realizados com pessoas sob o paradigma DRM (Deese-Roediger-McDermott), que envolve testes de memorização e reconhecimento de listas de palavras. O sistema modular implementado produziu padrões de falsas memórias bastante próximos, sob os pontos de vista qualitativo e quantitativo, aos manifestados por indivíduos em laboratório. Comparado à primeira versão, o modelo revisto não só aproximou melhor os resultados dos testes reais como também ampliou sua corretude, principalmente no que se refere ao uso de uma representação confiável e não-tendenciosa dos dados de entrada do sistema.

Palavras-chave: Falsas Memórias, Redes Neurais, Teoria do Rastro Difuso, Mapas Auto-Organizáveis, Paradigma DRM, Reconhecimento de palavras.

Sumário

Lista de Figuras

Lista de Tabelas

Lista de Acrônimos

ANN Rede Neural Artificial (*Artificial Neural Network*).

ART Teoria de Ressonância Adaptativa (*Adaptive Resonance Theory*).

BIC Ligação Item-Contexto (*Binding of Item and Context*).

DRM Paradigma de testes psicológicos Deese-Roediger-McDermott.

FTT Teoria do Rastro Difuso (*Fuzzy-Trace Theory*).

LVQ Aprendizagem por Quantização Vetorial (*Learning Vector Quantization*).

RBF Rede de Função de Base Radial (*Radial Basis Function Network*).

SNSI Identificador Neural Estocástico de Sequências (*Stochastic Neural Sequence Identifier*).

SOM Mapa Auto-Organizável (*Self-Organizing Map*).

SVD Decomposição em Valores Singulares (*Singular Value Decomposition*).

WAS Espaço de Associação de Palavras (*Word Association Space*).

1 Introdução

A Falsa Memória representa o fenômeno ilusório de reconhecimento, de modo que fatos ou experiências que jamais ocorreram possam ser recordados e eventos que efetivamente foram experienciados, não despertar nenhum tipo de familiaridade.

Desde a publicação do artigo clássico de Roediger e McDermott (1995), a psicologia e neurofisiologia voltaram sua atenção para este fenômeno, formulando uma série de técnicas para replicação das falsas memórias em laboratório. Um dos primórdios deste campo, o paradigma de testes DRM (Deese-Roediger-McDermott) se utilizava da apresentação de listas de palavras semanticamente relacionadas para manifestar o fenômeno, permitindo que dados quantitativos fossem coletados. Foi desenvolvida também a Teoria do Rastro Difuso (FTT), apresentando hipóteses derivadas das teorias psicológicas acerca das fenomenologias atuantes durante a memorização e o reconhecimento, elicitando as possíveis causas da falsificação.

No campo da Inteligência Computacional, Pacheco (2004) propôs um sistema modular de redes neurais artificiais capaz de reproduzir de forma satisfatória o comportamento da memória humana quando sob circunstâncias favoráveis ao armazenamento ou recuperação de informações falsas. Essas condições foram simuladas de acordo com o paradigma DRM, em que listas de palavras semanticamente relacionadas a um mesmo tópico são apresentadas para memorização e, posteriormente, aplica-se um teste de reconhecimento contendo itens exibidos ou não anteriormente. O reconhecimento de palavras não-apresentadas relacionadas às memorizadas e a falta de familiaridade com certas palavras exibidas, resultados que indicam a presença de falsas memórias, foram demonstradas pelo sistema proposto. As taxas em que esses julgamentos ocorreram foram qualitativa e quantitativamente próximas às dos experimentos reais com humanos.

Neste trabalho, propomos a revisão e adaptação do modelo neural concebido por

Pacheco (2004), visando aproximá-lo dos fundamentos adquiridos pela Teoria do Rastro Difuso e neurofisiologia nos últimos anos. Ajustando a representação e a concepção desse sistema modular, pretendemos re-implementá-lo, tornando-o capaz de reproduzir de forma adequada os padrões de falsa memória observados nos experimentos com o paradigma DRM.

Como objetivos específicos, esse trabalho visa:

a) revisar o estado da arte nos campos da psicologia e neurofisiologia no que se refere ao fenômeno de falsas memórias;

b) estudar e re-implementar o sistema modular concebido por Pacheco (2004);

c) inserir um novo modelo de representação dos dados semânticos de entrada do sistema;

d) ajustar e inserir novas características ao sistema, visando a melhoria de seu desempenho;

e) e obter resultados de simulação de falsas memórias compatíveis com os níveis do fenômeno detectados em laboratório e as hipóteses da Teoria do Rastro Difuso.

O trabalho inicia-se por uma revisão dos conceitos básicos da psicologia acerca dos processos de memorização e reconhecimento, com foco no aparecimento de falsas memórias, no Capítulo 2. No Capítulo 3, serão descritos todos os módulos do sistema proposto, sendo esclarecidos detalhes de suas arquiteturas, algoritmos e regras de propagação. Os experimentos de validação da implementação de cada componente do modelo neural, assim como os resultados das simulações dos testes de reconhecimento com humanos serão apresentados e discutidos no Capítulo 4. Por fim, concluiremos o trabalho analisando suas contribuições e as oportunidades para futuras pesquisas.

2
Falsas Memórias

Ilusões de memória podem ser definidas como falsificações subjetivas na representação de experiências passadas, através da adição, omissão ou substituição de detalhes durante os processos de armazenamento e/ou recordação das informações memorizadas (Drever, 1964). Dois fenômenos ilusórios populares encaixam-se perfeitamente nesta descrição: o *déjà vu* e o *jamais vu*. Enquanto o primeiro corresponde a lembrar ou sentir que uma experiência de fato inédita já foi vivenciada no passado, o segundo representa a situação oposta, em que o indivíduo sente dificuldade em lembrar de algo pelo qual já passou.

Neste capítulo, apresentamos uma revisão bibliográfica dos estudos relacionados aos fenômenos ilusórios da memória nas áreas da psicologia e neurofisiologia. A Seção 2.1 aborda os principais conceitos introduzidos pela Teoria do Rastro Difuso, que visa analisar de forma qualitativa e quantitativa as fenomenologias por detrás dos processos de memorização e reconhecimento nos seres humanos. Já na Seção 2.2, é revisado o estado da arte do conhecimento neurofisiológico acerca das estruturas do encéfalo humano envolvidas nestes processos, de forma sucinta.

2.1 A Teoria do Rastro Difuso

A **Teoria do Rastro Difuso** (*Fuzzy-Trace Theory*) (Brainerd e Reyna, 1990), apesar de ser considerada uma derivação da teoria de Piaget (Piaget, 1968), é essencialmente distinta, caracterizando-se por valorizar a metáfora da intuição, concebida como um conceito *fuzzy*, no lugar da metáfora tradicional e lógica da "mente como um computador". Introduzida como uma explicação para os processos cognitivos de decisão, a FTT transformou-se numa das principais fundamentações teóricas do fenômeno de falsa memória (Brainerd e Reyna, 1998a).

O conceito mais básico da FTT prediz a existência de dois tipos de memória: a **memória literal** (*verbatim memory*), cuja função é a de armazenar representações de estímulos superficiais (sequências de sons, fonemas, letras ou cores, por exemplo), e a **memória de essência** (*gist memory*), responsável por gerenciar informações de significado. De acordo com a FTT, ambas as memórias são acessadas em paralelo, tanto durante o armazenamento (Reyna e Brainerd, 1995), quanto durante a recuperação de informação (Brainerd e Reyna, 1990).

2.1.1 Metodologia Experimental

Experimentos com falsas memórias geralmente começam com a apresentação de listas de palavras, imagens ou frases correlacionadas (ou não) a indivíduos voluntários (Brainerd e Reyna, 1998a). É comum na literatura referir-se a esta fase como a de **estudo** (*study phase*). Os elementos das listas estudadas são chamados de **alvos** (*targets*). Por sua vez, objetos não-estudados mas associáveis aos apresentados em termos de significado são referidos como **distraidores relacionados** (*related distractors*). Por último, objetos significativamente distintos dos estudados constituem os **distraidores não-relacionados** (*unrelated distractors*).

Após a fase de estudo dos experimentos, aplica-se um teste composto por alvos, distraidores relacionados e distraidores não-relacionados, em que os voluntários são instruídos a sinalizarem a ocorrência de reconhecimento de quaiquer informações que fizeram parte da lista de estudo. Esta é a fase de **reconhecimento** (*recognition phase*). **Acertos** (*hits*) são obtidos quando os itens estudados são considerados "antigos" e quando itens não-estudados são marcados como "novos". Analogamente, **alarmes falsos** (*false alarms*) ocorrem na identificação de um alvo como informação "nova" e de um distraidor como conteúdo "antigo".

É importante notar que tanto a fase de reconhecimento quanto a de estudo podem sofrer manipulações circunstanciais, para fins de análise da influência de fatores externos sobre a formação de falsas memórias. A título de exemplo, é comum testar-se o reconhecimento de alvos e distraidores após intervalos de tempo distintos (imediatamente, semanas ou até meses após o estudo) ou sob diferentes instruções de reconhecimento (em que se deve indicar apenas objetos "certamente" estudados, apenas distraidores semelhantes ao material estudado ou ambos). Como ilustrado em Brainerd et al. (1999), circunstâncias de testes distintas são cruciais para se estimar com robustez a taxa de ocorrência de processos ilusórios de memória.

No caso de variação do tempo decorrido entre a fase de estudo e a aplicação do

teste de reconhecimento, verificou-se uma taxa de decaimento visivelmente superior para a memória literal, frente à memória de essência (Brainerd e Poole, 1997). Isso significa que o conteúdo de significado é melhor retido ao longo do tempo; enquanto as informações fonéticas e de contexto espacial são mais facilmente acessadas logo após o estudo.

2.1.2 Recordação de Alvo e Familiaridade

A Teoria do Rastro Difuso possui sua raiz na abordagem dual do processo de reconhecimento, em que a decisão pode resultar de duas fenomenologias distintas: a recordação de alvo e a familiaridade (Atkinson e Juola, 1973, 1974). A **recordação de alvo** (*target recollection*) equivale à lembrança consciente da apresentação de um item de teste no material estudado, podendo ser fornecidos detalhes que variam desde o tom de voz do locutor durante o estudo a estímulos visuais percebidos no momento da memorização (Brainerd et al., 1999; Jacoby, 1996). De acordo com a FTT, essa fenomenologia estaria inerentemente associada à memória literal.

A segunda fenomenologia, **familiaridade** (*familiarity*), faz referência à sensação de semelhança entre o estímulo presente e as experiências passadas, mas não implica em recordação explícita. A Teoria do Rastro Difuso atribui essa fenomenologia, por sua vez, à memória de essência, responsável pela semântica dos dados (Brainerd e Reyna, 1998a).

Nesse sentido, de acordo com a FTT, um reconhecimento subsidiado pela memória literal provavelmente será caracterizado por uma forte presença da recordação de alvo, enquanto um reconhecimento através da memória de essência, por uma intensa manifestação de familiaridade. Nos dois casos, alvos serão corretamente reconhecidos.

Para distraidores relacionados, no entanto, os dois processos estimulam decisões distintas (Brainerd et al., 1995; Clark e Gronlund, 1996; Rotello e Heit, 1999). Enquanto a familiaridade favorece alarmes falsos, a recordação auxilia na correta rejeição do distraidor, através do contraste com um ou mais alvos recordados.

Identidade, Não-Identidade e Similaridade

Inicialmente, a definição das fenomenologias de recordação e familiaridade possibilitou mapeá-las em três tipos de julgamento durante o reconhecimento: **identidade** (*identity*), **não-identidade** (*nonidentity*) e **similaridade** (*similarity*) (Brainerd e Reyna, 1998b).

A decisão de identidade corresponde à ocorrência de recordação durante a apresentação de um alvo (item estudado), ou seja, à lembrança vívida do momento de seu

estudo. O resultado é o reconhecimento do estímulo externo, compatível com os traços literais extraídos da memória.

A não-identidade, por sua vez, consiste da presença da mesma fenomenologia de recordação de alvo, porém durante a apresentação de um distraidor. Como este processo implica na rejeição do distraidor, ele também é conhecido por **rejeição por recordação** (*recollection rejection*) (Brainerd et al., 2003). A teoria do rastro difuso explica que a extração de traços literais que contrastam com os estímulos sensoriais provenientes do distraidor é sobretudo estimulada pela ocorrência de familiaridade, isto é, fortes traços de essência extraídos da memória podem levar à lembrança da experiência dos alvos relacionados, e à consequente realização de que o item apresentado, apesar de familiar, não foi estudado.

Por último, o julgamento de similaridade resulta na aceitação de alvos e distraidores relacionados, sendo justificado pela extração de traços da memória de essência semelhantes ao estímulo recebido. Isto ocorre quando a semelhança semântica de um item de teste com o material de estudo ultrapassa um determinado limiar.

É intuitivo que a similaridade seja precedida pela identidade ou não-identidade, caso algum dos processos relativos a estes dois julgamentos venha a ocorrer. Isto é, a extração de traços da memória literal ocasiona aceitação ou rejeição incondicionais do item apresentado. O único caso em que a resposta da memória semântica é considerado é aquele em que não ocorre a fenomenologia de recordação literal.

Na ausência de resposta da memória literal e de semelhança suficiente com os traços semânticos memorizados, a Teoria do Rastro Difuso atribui o reconhecimento de um item de teste à propensão individual de aceitá-lo sem motivos claros para tal, a qual denominaremos de **tendência** (*bias*). Intuitivamente, a tendência pode ser estimada através da taxa de reconhecimento de distraidores não-relacionados durante experimentos com indivíduos (Brainerd et al., 1999).

É importante distinguir-se o conceito de fenomenologia (recordação e familiaridade) do de padrão de julgamento (identidade, não-identidade e similaridade). As fenomenologias são processos teóricos, o que torna possível a existência de outras dinâmicas além da recordação por detrás do julgamento de não-identidade e outros processos além da familiaridade por detrás da similaridade (Reyna e Lloyd, 1997).

2.1.3 Evidências da Falsificação de Memória

O Paradigma DRM

Os fundamentos dos experimentos descritos nesta seção remetem ao paradigma **DRM** (*Deese-Roediger-McDermott*), proposto por Deese (1959) e reavaliado por Roediger e McDermott (1995).

O estudo de Deese, em 1959, representou uma das primeiras evidências de ilusões de memória em testes de recordação livre, que, diferentemente dos testes de reconhecimento, contém uma prova em que se deve relembrar o maior número possível de itens estudados em cada lista. Deese agrupou alvos em 36 listas, cada qual nomeada a partir de um **distraidor crítico** (*critical distractor*) associável a todas as palavras do seu grupo. Os resultados mostraram que as pessoas costumam gerar memórias de experiências que não ocorreram, com os distraidores possuindo, em geral, uma alta probabilidade de serem erroneamente recordados.

Durante muito tempo negligenciada, a metodologia de Deese foi estendida ao falso reconhecimento através do clássico estudo de Roediger e McDermott (1995). Nele, testes de recordação livre com as 24 listas mais importantes foram seguidos pelo reconhecimento de alvos e distraidores. A evidência encontrada através dos resultados foi a de que o processo de recordação espontânea intensifica o reconhecimento de conteúdo não-estudado, ajudando a consolidar a memória falsificada.

Os Experimentos de Brainerd & Reyna

Utilizando o paradigma DRM, Brainerd e Reyna (1998b) realizaram um estudo da frequência com que indivíduos aceitam ou rejeitam alvos e distraidores após a memorização. O resultado foi bastante interessante, mostrando que, frequentemente, distraidores relacionados podem ser mais facilmente reconhecidos que alvos efetivamente estudados. A reprodução destes experimentos, na qual reside o foco deste trabalho, ocorrerá no Capítulo 4.

Nestes experimentos, foram adotadas as 24 listas de palavras DRM. Contendo 15 palavras, cada lista foi ordenada, propositalmente, do significado mais próximo ao mais distante do distraidor crítico. A lista *chair*, por exemplo, era formada por *table*, *sit*, *legs*, *seat*, *couch*, *desk*, *recliner*, *sofa*, *wood*, *cushion*, *swivel*, *stool*, *sitting*, *rocking* e *bench*.

Participaram dos experimentos 60 estudantes universitários. Cada indivíduo ouviu uma fita contendo listas de 12 a 15 palavras escolhidas aleatoriamente, com um intervalo de três segundos entre palavras e de dez segundos entre listas consecutivas. As

palavras mais associadas aos distraidores críticos sempre foram apresentadas, no início de cada lista. Depois de escutar 12 listas de palavras, dividiu-se os voluntários em dois grupos, para cada qual foi entregue uma página contendo instruções acerca do teste que iria ser aplicado logo em seguida.

Alertado acerca da presença de distraidores relacionados ao material estudado, o Grupo 1 foi instruído a reconhecer apenas as palavras apresentadas para memorização, isto é, os alvos. Já o Grupo 2 foi liberado para reconhecer também os distraidores, recebendo a instrução para se aceitar qualquer palavra cuja essência permitiria incluí-la em alguma das listas estudadas.

O teste de reconhecimento foi composto por 72 palavras:

- 36 alvos extraídos de cada uma das 12 listas estudadas (3 aleatoriamente escolhidos por lista);
- os 12 distraidores críticos das listas estudadas;
- os 12 distraidores críticos das 12 listas não-estudadas;
- 12 distraidores não-relacionados extraídos das listas não-estudadas (1 aleatoriamente selecionado por lista).

Os resultados deste experimento mostraram que indivíduos no Grupo 1 reconheceram em média 61% dos alvos (A), rejeitando erroneamente 39% dos mesmos (Tabela 2.1). Em termos de falso reconhecimento, 63% dos distraidores críticos relacionados (DCR), 19% dos distraidores críticos não-relacionados (DCNR) e 16% dos distraidores não-relacionados (DNR) foram aceitos. Já no Grupo 2, reconheceu-se 68% dos alvos A, 88% dos distraidores relacionados DCR, 28% de itens não-relacionados DCNRs e 25% de DNRs.

Percebe-se que o Grupo 2 aceitou distraidores críticos das listas estudadas a uma taxa substancialmente maior (88%) que alvos (68%). A Teoria do Rastro Difuso explica estes resultados partindo do princípio que o acesso aos dois tipos de memória, literal e de essência, ocorre de formas distintas. A maior facilidade de recuperação de traços semânticos armazenados durante o estudo justificaria o reconhecimento intenso de distraidores críticos, baseados na memória de essência. Em contrapartida, o lento acesso à memória literal dificultaria a extração das informações superficiais dos alvos estudados, que, diferentemente dos distraidores críticos, nem sempre são bons subsídios à recuperação do significado central de suas listas.

O baixo índice de reconhecimento de distraidores não-relacionados, críticos ou não, confirma a hipótese de que itens com significado distinto dos estudados serão

pobres manifestantes de recordação ou familiaridade. A diferença de aceitação destes distraidores entre os Grupos 1 e 2 (19% e 28% para DCNRs; 16% e 25% para DNRs, respectivamente) demonstra que instruções menos restringentes, como a do Grupo 2, elevam a tendência dos indivíduos a acatar uma palavra qualquer (*bias*).

Tabela 2.1: Resultados dos experimentos de Brainerd e Reyna (1998b) com estudantes. O Grupo 1 deveria reconhecer apenas alvos (A). O Grupo 2 deveria reconhecer alvos (A) e distraidores relacionados (DCR). Nenhum dos grupos deveria aceitar distraidores críticos não-relacionados (DCNR) ou distraidores não-relacionados (DNR).

Tipos de teste	Percentual de aceitação	
	Grupo 1(std)	Grupo 2(std)
A	61%(08)	68%(09)
DCR	63%(13)	88%(08)
DCNR	19%(08)	28%(14)
DNR	16%(11)	25%(15)

O índice de reconhecimento de distraidores críticos no Grupo 1 foi, se não ligeiramente superior, equivalente ao reconhecimento de alvos. Como veremos na próxima seção, esse resultado é melhor explicado pela terceira fenomenologia proposta pela Teoria do Rastro Difuso, a recoleta fantasma.

2.1.4 Recordação Fantasma

Nas Seções 2.1.2 e 2.1.3, foram explicadas as dintinções teóricas do modelo primordial da FTT. Envolvendo duas fenomenologias (recordação de alvo e familiaridade) e três tipos de julgamento (identidade, não-identidade e similaridade), este modelo adequou-se muito bem a experimentos caracterizados por um nível moderado de falsas memórias semânticas, tendo sido estatisticamente implementado em Brainerd et al. (1999). Os parâmetros calculados para cada uma das fenomenologias responderam como esperado diante de manipulações experimentais que deveriam afetar processos de memória relacionados, confirmando o sucesso inicial da teoria.

A expectativa era de que, em testes com alto falso reconhecimento de distraidores relacionados, a proporção de alarmes por familiaridade excedesse a de rejeição por recordação, por diversos motivos, como o acesso mais fácil e rápido a traços *gist* que a traços literais na memória. Caso isso fosse confirmado, os alarmes falsos deveriam ser acompanhados por sentimentos de similaridade com o material estudado.

No entanto, percebeu-se que certos paradigmas de teste resultavam em muitos alarmes falsos sendo acompanhados de relatos característicos da fenomenologia de

recordação, por vezes excedendo os de similaridade (Israel e Schacter, 1997; Payne et al., 1996; Roediger e McDermott, 1995). Ao tentar explicar este fato, o modelo da FTT falhou.

A recordação vívida e ilusória, que capacita o indivíduo a descrever claramente as circunstâncias em que foi apresentado a um distraidor, estimulou a evolução da Teoria do Rastro Difuso. Ao lado da familiaridade e da recordação de alvo, esta nova fenomenologia foi denominada **recordação fantasma** (*phantom recollection*) (Brainerd et al., 2001).

Os paradigmas que induzem este processo possuem no mínimo dois aspectos: vários alvos compartilham o mesmo significado e distraidores similares em essência a esses alvos são testados. O procedimento DRM é um exemplo claro dessas circunstâncias.

É importante diferenciar a recordação fantasma da recordação de alvo. Apesar de ambas fenomenologias resultarem no reconhecimento de um item de teste como tendo sido um objeto certamente apresentado durante o estudo, a recordação fantasma ocorre sem a extração de traços *verbatim* armazenados na memória literal, uma vez que distraidores são percebidos pela primeira vez apenas no momento de serem testados.

Nesse sentido, a Teoria do Rastro Difuso defende que a causa da recordação fantasma consiste de uma memória de essência inusitadamente forte. A consolidação de traços *gist*, durante o estudo, seria então capaz de ultrapassar a barreira da familiaridade, gerando sensações de reconhecimento semelhantes àquelas produzidas pela recuperação de informações literais.

Falsa Identidade e Recordação Errônea

Após sua adaptação, a FTT passou a incorporar o julgamento de **falsa identidade** (*false identity*), que, por ser um mapeamento direto da recordação fantasma, também pode ser referenciado pelo mesmo nome desta fenomenologia.

Em paradigmas como o DRM, o estudo de Brainerd et al. (2001) estimou uma taxa de alarmes falsos devido à falsa identidade em torno de 65%, sendo 30% em razão da similaridade. Além disso, quando o nível de alarmes falsos tornou-se superior ao nível de acertos de alvos, identificou-se aproximadamente 74% dos falsos reconhecimentos como sendo devidos à recordação fantasma, enquanto os alarmes devido à familiaridade descresceram para 17% do total. Esses resultados comprovaram a importância de se considerar o novo modelo da FTT em estudos futuros.

Em Brainerd et al. (2003), uma análise qualitativa e quantitativa estendida do processo de não-identidade (ou rejeição por recordação) estimulou a definição de

um novo tipo de julgamento: a rejeição por recordação errônea ou, simplesmente, **recordação errônea** (*erroneous recollection*). Essa decisão corresponde à rejeição de um alvo baseada na recordação de itens semelhantes também apresentados, criando uma situação em que indivíduos utilizam-se de processos conscientes de supressão de falsas memórias para, de forma inapropriada, rejeitar material estudado (Brainerd e Wright, 2005).

2.1.5 Modelo Atual da FTT

No fim, o modelo revisado da Teoria do Rastro Difuso, que adequou-se melhor aos experimentos psicológicos, congrega os seguintes processos (ignorando a tendência, ou *bias*):

- Identidade (*identity*): a extração de traços da memória literal produz a aceitação de alvos e induz a fenomenologia de recordação vívida da sua apresentação;
- Recordação Errônea (*erroneous recollection*): processo *verbatim* que leva à rejeição incorreta de alvos (suponha *nurse*, por exemplo) devido à extração de traços literais de alvos relacionados (como *doctor* e *medicine*) e ocorrência de recordação vívida dos mesmos, anulando a familiaridade com o item de teste;
- Não-identidade ou Rejeição por Recordação (*nonidentity* ou *recollection rejection*): julgamento composto pela sensação de familiaridade com um distraidor relacionado e extração dos traços literais dos alvos correspondentes, o que induz a recordação destes e a correta rejeição do distraidor;
- Falsa Identidade ou Recordação Fantasma (*false identity* ou *phantom recollection*): mapeamento direto da fenomenologia de mesmo nome, consiste do processo *gist* de aceitação de distraidores fortemente relacionados (críticos), através da indução à ilusão de lembrança de sua apresentação;
- Similaridade (*similarity*): processo *gist* em que a extração de traços da memória de essência induz apenas a fenomenologia de familiaridade, levando ao reconhecimento de alvos ou distraidores relacionados.

2.1.6 Premissas Incorporadas ao Modelo Proposto

Pacheco (2004) propôs a investigação de falsas memórias por meio de simulações computacionais do fenômeno, baseando-se nos postulados primordiais da Teoria do

Rastro Difuso. As premissas que orientaram a construção de seu modelo, no entanto, serão revistas neste trabalho, à luz dos avanços teóricos sofridos pela FTT.

De acordo com a análise das seções anteriores, espera-se que o modelo neural apresentado neste trabalho (Capítulo 3) satisfaça as seguintes proposições:

a) O armazenamento de informações ocorre em paralelo nas memórias literal e de essência;

b) A recuperação de informação é processada em paralelo pelas memórias literal e de essência;

c) O decaimento da memória literal é mais rápido que o da memória de essência;

d) Os alvos são reconhecidos mais frequentemente por identidade (através da memória literal), enquanto os distraidores relacionados costumam ativar fortemente a memória de essência, sendo reconhecidos por similaridade;

e) O julgamento de identidade produz a aceitação correta de alvos através da memória literal;

f) O julgamento de não-identidade produz a rejeição correta de distraidores relacionados através da memória literal;

g) O julgamento de similaridade produz a aceitação correta de alvos e incorreta de distraidores relacionados através da memória de essência;

h) O julgamento de recordação errônea produz a rejeição incorreta de alvos através da recordação de alvos relacionados;

i) O julgamento de recordação fantasma produz a aceitação incorreta de distraidores relacionados;

j) O reconhecimento de distraidores não-relacionados deve-se principalmente ao grau subjetivo de tendência à aceitação (*bias*).

Com o intuito de validar o modelo proposto e analisar a incorporação das premissas enumeradas acima, simulações computacionais baseadas nos experimentos de Brainerd e Reyna (1998b) serão realizadas, como veremos no Capítulo 4. Espera-se que o sistema implementado alcance resultados, em especial, qualitativamente compatíveis com os observados nestes experimentos com humanos.

2.2 Dinâmica Neural

Nesta seção, serão discutidos os aspectos anatômicos e funcionais das estruturas do encéfalo humano envolvidas na memorização e reconhecimento. Analisaremos a memorização, o reconhecimento e as regiões ativadas durante estes processos, apresentando por fim um conjunto de premissas que guiarão o modelo proposto no Capítulo 3.

2.2.1 Processo de Memorização

O modelo de memorização proposto por Mishkin e Appenzeller (1987) postula que os estímulos são recebidos inicialmente por regiões corticais sensoriais primárias, de onde são enviados às seções de associação dos lobos parietal, occipital e temporal do cérebro. Nelas, os sinais são convertidos numa espécie de "representação neural" dos estímulos de entrada, originando a experiência de percepção.

Segundo o modelo BIC (*binding of item and context*) (Diana et al., 2007), a extração e codificação de informação contextual é papel do córtex parahipocampal (parte do giro hipocampal posterior), enquanto o córtex perirrinal (localizado no giro hipocampal anterior) lida com a representação do conteúdo específico do item apresentado.

Em seguida, essa nova representação é enviada ao hipocampo e à amígdala, que fazem parte do sistema límbico (conjunto das regiões encefálicas relacionadas à manifestação de emoções). O hipocampo associa a informação recebida com o contexto espacial e temporal em que a mesma está sendo aprendida e avalia sua relevância em termos de novidade (Fletcher et al., 1997; Knight, 1996). A partir de então, os estímulos são repetidamente transmitidos às regiões talâmicas do sistema límbico (tálamo e hipotálamo) conforme sua relevância.

Paralelamente, a amígdala recebe a representação neural produzida pelas regiões de associação corticais e atribui valores afetivos aos estímulos. Da mesma forma que o hipocampo, ela também passa a enviá-los aos centros talâmicos.

Nesses processos, o hipocampo e a amígdala interagem, reverberando importantes informações e usando o córtex límbico como uma memória de trabalho (Taylor et al., 2000). Recebendo os estímulos destas regiões, o tálamo e hipotálamo, por sua vez, retransmitem-nos ao córtex pré-frontal (Mishkin et al., 1982), no lobo frontal do cérebro.

A participação do córtex pré-frontal durante a memorização decorre de duas funções: ajudar a amígdala a determinar o valor afetivo da informação; e servir como uma memória de trabalho capaz de realizar processamento cognitivo quando requisitada.

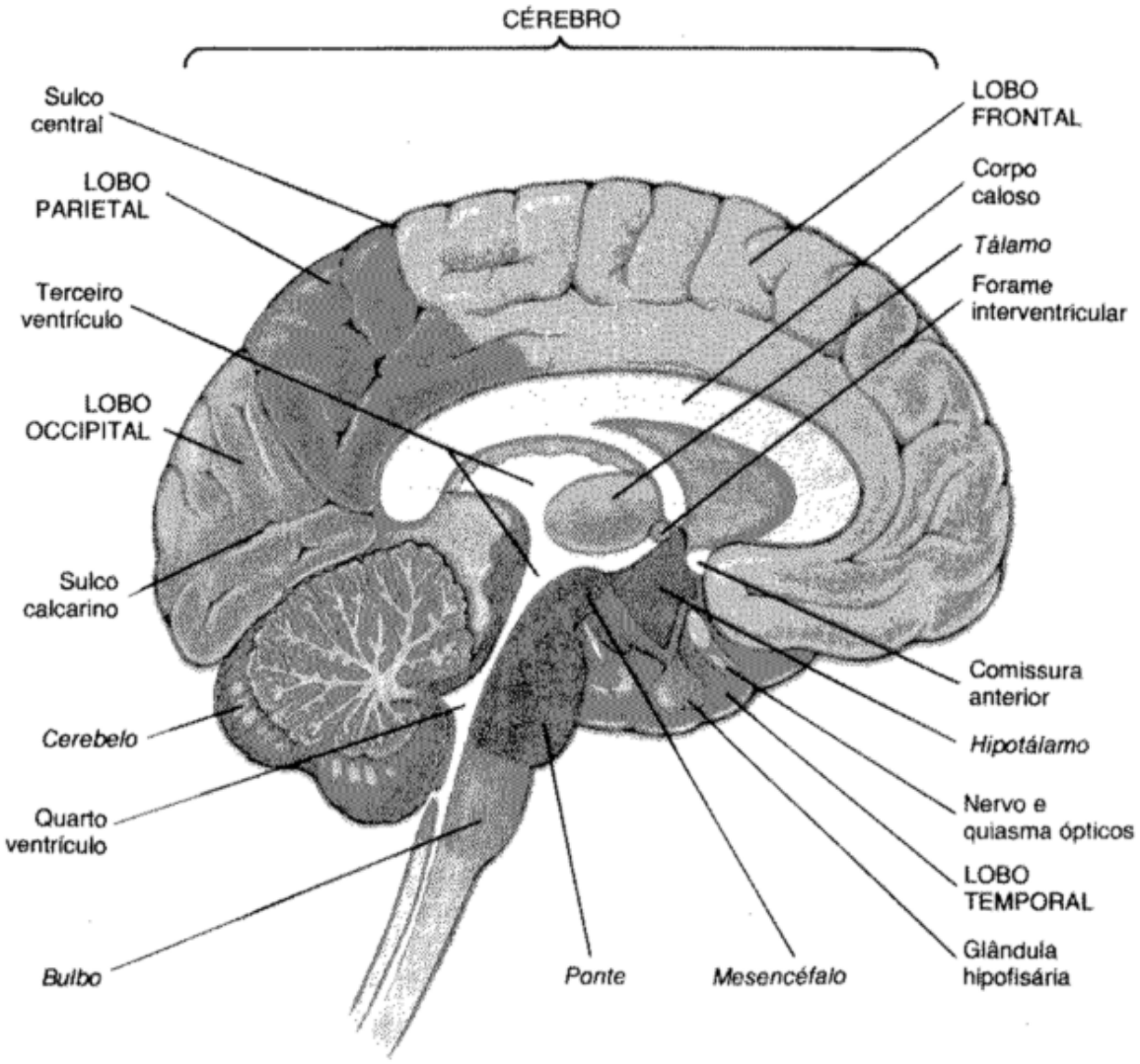

Figura 2.1: Anatomia da metade esquerda do encéfalo (Guyton, 1993).

Por fim, as estruturas diencefálicas (o tálamo e o hipotálamo) e o córtex pré-frontal enviam sinais ao prosencéfalo basal, parte do sistema límbico responsável por liberar o neurotransmissor acetilcolina (linhas pontilhadas na Figura 2.2). Num cenário plausível, a liberação desta substância incidiria de volta sobre as regiões sensoriais do córtex e do sistema límbico, iniciando uma série de eventos que modificariam as sinapses, fortalecendo conexões neurais e transformando a percepção sensorial num rastro real de memória (Mishkin et al., 1982).

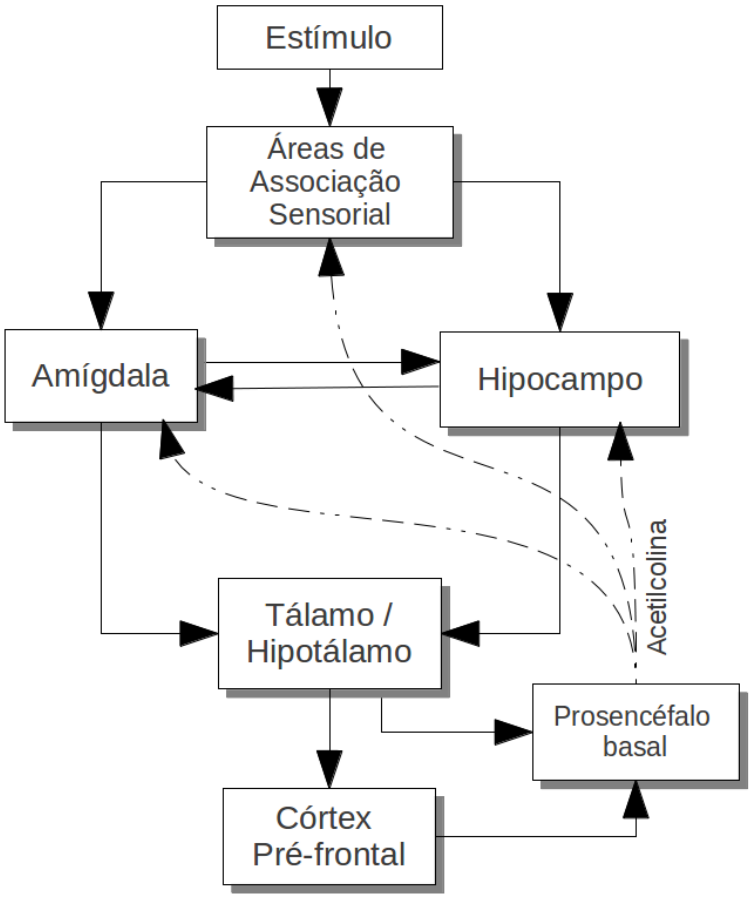

Figura 2.2: Fluxo de informação entre diferentes regiões do encéfalo humano durante o processo de memorização.

2.2.2 Processo de Reconhecimento

Assim como na memorização, os primeiros passos do reconhecimento consistem no processamento dos estímulos nas áreas sensoriais do córtex cerebral e na recuperação de seu significado nas regiões de associação. Também de forma análoga, Diana et al. (2007) sugere que a extração de informação contextual é realizada pelo córtex parahipocampal, enquanto o córtex perirrinal identifica o conteúdo específico do item apresentado. Essas informações são então enviadas a uma memória de trabalho no sistema límbico, provavelmente localizada no córtex límbico ou no lóbulo quadrado (Taylor et al., 2000).

Enquanto o hipocampo associa o contexto espacial e temporal à informação específica do estímulo numa única representação (Schacter, 1996; Fletcher et al., 1997; Aggleton e Brown, 1999; Diana et al., 2007), a amígdala avalia o estímulo de forma afetiva (Cahill et al., 1996; Henson et al., 1999; Wilson e Rolls, 1993).

Após definidos os atributos contextuais e afetivos da informação, o córtex pré-frontal passa a coordenar a busca de uma experiência equivalente na memória do indivíduo. Suas diferentes regiões assumem papéis distintos nesta etapa do reconhecimento. O córtex pré-frontal medial é responsável por recuperar informações armazenadas na memória de longo prazo dos lobos cerebrais parietal, occipital e temporal (Fletcher et al., 1997). Henson et al. (1999) argumenta que sua região direita, por sua vez, monitora a estruturação de uma "chave de busca" (um conjunto de sinais sensoriais ou de atributos, como visuais, fonéticos, semânticos, contextuais e afetivos), avaliando a resposta obtida e determinando se a chave formada é apropriada. Por fim, a porção esquerda do córtex pré-frontal serviria como a memória de trabalho deste processamento (Taylor et al., 2000).

Além do córtex pré-frontal, o sistema límbico também auxilia o processo de busca, mantendo a chave de busca sendo usada em sua memória. Isto ocorre até que a experiência seja recuperada com sucesso ou o córtex pré-frontal considere-a absenta da memória (Burgess e Shallice, 1996; Shallice e Burgess, 1996; Schacter et al., 1998; Henson et al., 1999). O julgamento final do indivíduo é então manifestado em seu comportamento.

2.2.3 Premissas Incorporadas ao Modelo Proposto

De acordo com os temas cobertos nesta Seção 2.2, espera-se que o modelo neural apresentado neste trabalho cubra as seguintes premissas neurofisiológicas:

a) Durante a associação sensorial, o córtex parahipocampal monta uma representação do contexto espacial e temporal em que a informação foi recebida, enquanto o córtex perirrinal extrai a representação dos aspectos específicos do estímulo;

b) Durante a memorização, o hipocampo é responsável por atribuir relevância em termos de novidade às informações que deverão ser armazenadas na memória, afetando a força com que isso será feito;

c) Durante a memorização, circuitos reverberantes consolidam as memórias de longo prazo no córtex cerebral;

d) Durante a memorização e o reconhecimento, o hipocampo associa o contexto temporal e espacial ao conteúdo específico do estímulo, originando uma representação neural única;

e) Os estímulos durante a memorização iniciam-se nas áreas de associação sensorial do córtex, seguindo para o sistema límbico, para o diencéfalo e retornando aos lobos occipital, parietal e temporal do córtex, onde as sinapses serão fortalecidas e uma memória de longo prazo será formada;

f) Os estímulos durante o reconhecimento iniciam-se nas áreas de associação sensorial do córtex, seguindo para o sistema límbico, para o diencéfalo e então para o córtex pré-frontal, onde a decisão final acerca do reconhecimento é tomada.

3

Proposição do Modelo Neural

O sistema modular de redes neurais proposto por Pacheco (2004) e adaptado neste trabalho é descrito em detalhes neste capítulo. Na Seção 3.1, apresentamos a organização geral do modelo. Na Seção 3.2, abordamos as formas de representação adotadas para as informações superficiais (literais) e semânticas das palavras, isto é, os dados de entrada do sistema. Na Seção 3.3, descrevemos o algoritmo do primeiro módulo do sistema, o Módulo de Contexto. Nas Seções 3.4 e 3.5, introduzimos o funcionamento e as aplicações dos módulos seguintes, o Módulo de Essência e o Módulo de Literalidade. Por fim, o último componente do sistema, o Módulo de Decisão, é abordado na Seção 3.6.

3.1 Visão Geral

O modelo proposto visa reproduzir as tarefas principais executadas pelas estruturas cerebrais da Figura 2.2, estudadas na Seção 2.2, durante os processos de memorização e reconhecimento. Nesse sentido, são aplicadas abstrações das regiões encefálicas na forma de módulos de redes neurais (*artificial neural networks*) com aprendizagem não-supervisionada (*unsupervised learning*).

A Figura 3.1 expõe os módulos funcionais do sistema: o de associação sensorial, contexto, essência, literalidade e decisão. A relação entre os módulos de Literalidade e Essência está de acordo com os postulados da Teoria do Rastro Difuso, no que as informações *gist* e *verbatim* são processadas paralelamente.

O **módulo de associação sensorial** transmite o estímulo de entrada ao módulo de contexto como um código contendo representações das informações sensoriais e semânticas do estímulo. Como este trabalho preocupa-se em reproduzir experimentos no paradigma DRM, em que os itens estudados consistem de palavras, as informações

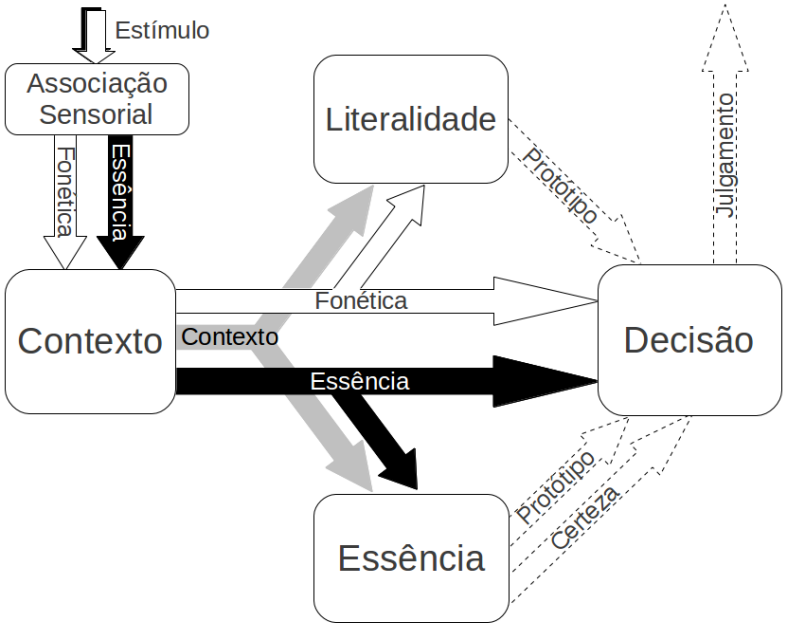

Figura 3.1: Visão geral dos módulos neurais do sistema proposto por Pacheco (2004).

literal e de essência equivalerão a codificações dos fonemas e do significado da palavra apresentada, respectivamente. Como foi visto no Capítulo 2, esse processamento ocorre nas áreas de associação do cérebro (lobos parietal, occipital e temporal).

O **módulo de contexto** mantém um histórico dos estímulos recebidos e associa cada um deles às circunstâncias (semânticas) em que foram apresentados, criando pares significado-contexto. Do ponto de vista neurofisiológico, a região análoga a esse módulo seria o hipocampo, associando as representações fornecidas pelo córtex perirrinal e o córtex parahipocampal. Além de propagar as representações literal e de essência, o módulo de contexto também alimenta o resto do sistema com os dados relativos ao contexto atual e ao contexto associado do estímulo de entrada.

Os módulos de essência e literalidade trabalham simultaneamente sobre os sinais recebidos do módulo de contexto. O **módulo de essência**, fornecido com a semântica e o contexto associado ao estímulo, gera protótipos de significado, adaptando-os de acordo com o nível de novidade do conteúdo recebido. Além disso, durante o reconhecimento, esse módulo estima o grau de certeza em sua própria resposta. Podemos, assim, verificar duas analogias a conceitos neurofisiológicos: a atribuição de relevância em termos de novidade à informação, pelo hipocampo, e a idéia de existência de um grau contínuo na manifestação de familiaridade.

O **módulo de literalidade**, por outro lado, recebe as informações contextuais e fonéticas, armazenando protótipos de palavras sonoramente similares de forma mediada

pelo contexto. No cérebro, a memória de longo prazo de informações literais localiza-se especialmente no lobo temporal (Taylor et al., 2000).

Por fim, o **módulo de decisão** é alimentado com as representações fonética e semântica do estímulo de entrada e as saídas dos módulos de essência e literalidade, que consistem dos protótipos de essência e literalidade recuperados da memória do sistema, respectivamente (com a adição do grau de certeza do módulo de essência em sua resposta). De forma análoga ao córtex pré-frontal (Seção 2.2.3), a função do módulo de decisão é definir o tipo de julgamento que ocorrerá, resultando na aceitação ou rejeição da palavra apresentada. De acordo com a Teoria do Rastro Difuso, cinco são os possíveis tipos de decisão, excluindo-se a tendência (*bias*) de cada indivíduo (Seção 2.1.5): identidade, recordação errônea, não-identidade, recordação fantasma e similaridade.

3.2 Módulo de Associação Sensorial

O Módulo de Associação Sensorial é responsável pela construção da representação neural dos estímulos. De forma a evitar-se limitações no sistema proposto, a representação dos dados de entrada deve, inerentemente, possuir as seguintes características:

- expansibilidade, sendo possível codificar qualquer palavra;
- similaridade semântica, sendo gerados códigos parecidos para palavras próximas em termos de significado;
- similaridade fonética, sendo gerados códigos parecidos para palavras próximas em termos de fonética;
- e simplicidade, através de uma representação simples e compacta que mantenha as características principais dos dados sem comprometer o processamento computacional ou sua interpretação.

3.2.1 Representação Literal

A representação literal proposta por Pacheco (2004) é baseada na representação de Hinton e Shallice (1991), que considera a grafia das palavras. No entanto, na maioria dos experimentos com falsas memórias apresenta-se sinais de áudio, justificando a adaptação dessa representação para congregar as propriedades fonéticas das palavras.

Cada fonema pode ser caracterizado por traços distintivos, relativos às propriedades articulatórias necessárias à sua pronúncia. Jakobson et al. (1952) definiu 12 desses traços: (1) vocálico/ não-vocálico, (2) consonantal/ não-consonantal, (3) interrompido/ contínuo, (4) brusco/ fluente, (5) estridente/ doce, (6) sonoro/ não-sonoro, (7) compacto/ difuso, (8) grave/ agudo, (9) rebaixado/ sustentado, (10) incisivo/ raso, (11) tenso/ frouxo, (12) nasal/ oral.

Apesar de ser possível representar 2^{12} fonemas distintos através de todos os traços, a maioria das línguas contém apenas uma fração dessa quantidade, como o Inglês, que faz uso de apenas 9 traços (Cherry, 1974). Nesse sentido, o código binário adotado apresenta 9 bits por fonema, mapeamento demonstrado na Tabela 3.1.

Tabela 3.1: Fonemas da língua inglesa (Proctor, 1995) e suas representações.

Símbolo fonético	Codificação binária	Símbolo fonético	Codificação binária
o ɐ ɔ	101110000	m	010101000
a	101100000	f	010100110
e	101000000	p	010100100
u ʊ w	100110000	v	010100010
ə	100100000	b	010100000
i I	100000000	n	010001000
l	110000000	s	010001000
ŋ	011001000	θ	010000110
ʃ	011000110	t	010000100
ʌʃ	011000101	z	010000011
k	011000100	ð	010000010
ʒ	011000010	d	010000000
ǯ dʒ	011000001	h r	000000100
g	011000000	#	000000000

Dado a codificação binária de cada fonema, Pacheco (2004) realizou testes com quatro tipos de estratégias distintas para representar cadeias de fonemas. A abordagem que apresentou os melhores resultados, gerando códigos próximos para palavras com estruturas fonéticas semelhantes, foi similar à de Hinton e Shallice (1991). As regras de codificação literal definidas foram as seguintes:

a) cada palavra será representada como uma lista de fonemas alternando entre consoantes e vogais, começando por uma consoante;

b) caso o primeiro fonema de uma palavra seja uma vogal, os primeiros 9 bits da representação desta palavra serão preenchidos com bits 0 (símbolo '#' na Tabela 3.1);

c) caso dois fonemas consecutivos sejam do mesmo tipo (vogal ou consoante), ao invés de serem separados por um preenchimento com '#', serão aglutinados numa só posição, através de uma operação OR binária entre suas cadeias de 9 bits;

d) o tamanho da representação será fixo para todas as palavras, correspondendo a n cadeias de 9 bits, onde n corresponde ao maior número de fonemas presentes numa palavra da base de dados usada;

e) caso o número de fonemas numa palavra seja menor que o número máximo de fonemas definido, a parte menos significativa de sua representação será preenchida com símbolos '#'.

De acordo com as regras utilizadas, a palavra *walk*, por exemplo, seria decomposta em *walk* = /uok/ = /#uok/ = /#(u ∧ o)k/ = "000000000 101110000 011000100". Supondo que o maior comprimento de uma palavra da base utilizada fosse de 7 fonemas, a parte menos significativa da representação de *walk* seria preenchida com quatro cadeias nulas '#'.

Modificações na Representação

Neste trabalho, adotou-se a mesma representação fonética utilizada por Pacheco (2004). É importante salientar, no entanto, que a base de palavras aqui utilizadas foi expandida para incluir as 24 listas do paradigma DRM, enquanto Pacheco limitou-se a realizar testes com apenas 12 das listas. Como resultado, tivemos um crescimento do tamanho da cadeia de bits utilizada para a informação literal, que passou a conter 63 bits (ou características).

Na reconstrução da base de palavras, utilizou-se o mapeamento fonético indicado pelo dicionário da Carnegie Mellon University (CMU, 2011), disponível em domínio público e contendo a pronúncia norte-americana de mais de 100.000 palavras.

3.2.2 Representação de Essência

Pacheco (2004) propôs 129 propriedades semânticas para a representação de essência de 12 das listas DRM, cada qual podendo variar continuamente num intervalo entre 0 e 1. Apesar da codificação proposta ter capturado intuitivamente as diferenças e similaridades de significado entre os itens das listas, e garantido que cada palavra possuísse entre 16% e 30% do total de propriedades, como sugerido por Hinton e Shallice (1991), a abordagem não é segura o suficiente para garantir a ausência de uma tendência (*bias*) na representação, pelo menos não enquanto este processo for manual.

Espaço de Associação de Palavras (WAS)

Neste trabalho, decidimos adotar a codificação semântica proposta por Steyvers et al. (2004), que aplicou e comparou a performance de diferentes técnicas de dimensionamento na criação de um hiperespaço para representação do significado de palavras, dentre elas a de *singular value decomposition* (SVD). O espaço multidimensional originado foi chamado de *Word Association Space* (WAS).

A definição do hiperespaço WAS provém da base de normas associativas construída por Nelson e Schreiber (1998), que necessitou de aproximadamente uma década de experimentos com humanos. As normas associativas são definidas entre duas palavras distintas, a **sugestão** (*cue*) e a **resposta** (*response*), e consistem da probabilidade de um indivíduo mencionar a palavra de resposta como a primeira associação à palavra sugerida, após lê-la ou ouvi-la. Contendo valores de norma para mais de 5000 palavras, a base da University of South Florida encontra-se disponível publicamente.

Pode-se notar, então, que uma matriz de normas associativas, onde cada linha representa uma sugestão e cada coluna representa uma resposta, não necessariamente será simétrica. Isto porque muitas pessoas podem, por exemplo, pensar em *cat* como a associação mais forte de *pet*, e, no entanto, considerar *dog* a associação primária de *cat*. Ou seja, as normas de associação devem ser tratadas como propriedades assimétricas por natureza.

Através da matriz de normas associativas, pode-se aplicar uma técnica de dimensionamento como a de SVD para geração de um espaço multidimensional de representação semântica, onde estima-se a posição relativa das palavras em termos de significado (Figura 3.2). Apesar da decomposição matricial SVD poder ser aplicada a matrizes assimétricas, os resultados tornam-se mais interpretáveis no caso de uma matriz simétrica (Steyvers et al., 2004), o que é desejável, considerando-se que utilizaremos a representação resultante para modelar um processo cognitivo.

Nesse sentido, antes da técnica de dimensionamento, optamos por transformar a matriz de associação assimétrica numa matriz simétrica. A abordagem adotada para a operação foi a segunda proposta no trabalho de Steyvers et al. (2004), em que a matriz simétrica é construída elicitando relações indiretas entre as palavras. A razão é que a operação de SVD aplicada sobre esta matriz levou a melhores resultados de correlação que quando aplicada sobre a matriz contendo apenas normas diretas, melhorando a qualidade na representação semântica dos dados.

As equações 3.1 e 3.2 indicam o pré-processamento realizado sobre a matriz de normas associativas, onde o valor na posição (i, j) é indicado por S_{ij}. Enquanto a operação 3.1 incorpora apenas simetria à matriz, a 3.2 torna-a mais robusta, adicionando valores

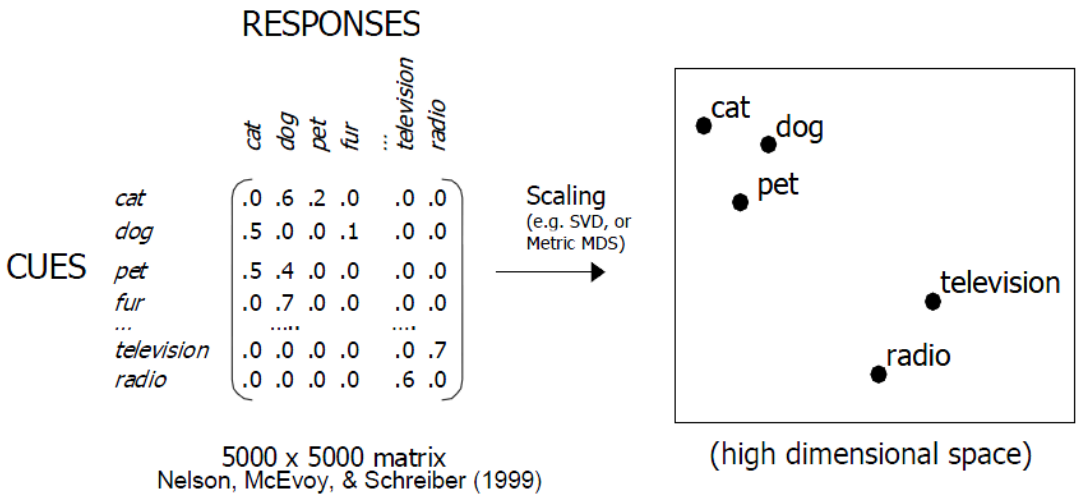

Figura 3.2: Ilustração da criação de um Espaço de Associação de Palavras (WAS) a partir de uma matriz contendo normas associativas (Steyvers et al., 2004).

de associação indireta (afinal, a similaridade de palavras indiretamente associadas não deve ser subestimada).

$$S_{ij}^{(1)} = A_{ij} + A_{ji} \tag{3.1}$$

$$S_{ij}^{(2)} = S_{ij}^{(1)} + \sum_{k} S_{ik}^{(1)} S_{kj}^{(1)} \tag{3.2}$$

Finalmente, a representação semântica é obtida através da decomposição por SVD da matriz de associação simétrica, o que nos fornece vetores singulares para cada linha e coluna da matriz original, ou seja, para cada palavra. A partir daí, o número de dimensões do espaço multidimensional WAS pode ser ajustado de acordo com a aplicação. Deve-se levar em conta que um pequeno número de dimensões será incapaz de capturar detalhes suficientes da estrutura associativa original, enquanto que um número muito grande tenderá a descomprimir a representação, mascarando relações entre os dados.

Neste trabalho, optamos por manter o número de características semânticas igual a 129, como proposto por Pacheco (2004). No entanto, conferimos embasamento experimental, através de dados reais de associação, à construção desse espaço multidimensional semântico, evitando o aparecimento de *bias* nas codificações.

Modificações na Representação

A principal diferença da nova representação semântica em relação à proposta por Pacheco (2004), capaz de mudar a dinâmica de todo o sistema modular proposto, é o fato de cada característica corresponder agora a um valor real entre -1 e 1. Isto é, enquanto a antiga representação baseava-se em graus de pertinência unilaterais a diferentes características, a nova permite que significados opostos localizem-se numa mesma escala bilateral, ampliando o escopo da representação e tornando-a mais intuitiva.

Como será descrito nas próximas seções, em razão da mudança na codificação de essência das palavras, foi necessário reconsiderar alguns aspectos dos módulos subsequentes, como os seus parâmetros, para tornar o sistema capaz de reproduzir os experimentos de falso reconhecimento do Capítulo 4.

3.3 Módulo de Contexto

O Módulo de Contexto possui a função de atribuir um contexto temporal aos dados de entrada do sistema, através da manutenção de um histórico dos traços semânticos que enfatiza dados recentes em detrimento de dados antigos. Como várias conexões recorrentes são observadas nas regiões corticais da memória, uma rede neural recorrente é uma escolha adequada à modelagem deste módulo.

Como pode ser visto na Figura 3.3, as entradas do módulo consistem das representações literal e de essência da palavra apresentada, de acordo com as codificações da Seção 3.2. As saídas são as mesmas entradas, inalteradas, e três variáveis produzidas pelo módulo: o conceito semântico recuperado do protótipo vencedor; o contexto associado a esse protótipo; e o presente contexto semântico. Esses três valores, juntos, formam o conjunto de dados contextuais do sistema, representado pela seta cinza na Figura 3.1.

3.3.1 Módulo de Contexto: Arquitetura, Representação e Regras de Propagação

O algoritmo do módulo de contexto possui sua inspiração no modelo ART2 (*Adaptive Resonance Theory 2*) (Carpenter e Grossberg, 1987), uma vez que se considera uma aprendizagem não-supervisionada e incremental para este módulo. A estrutura ART2 é capaz de agrupar padrões de valores reais, associar estímulos de diferentes naturezas e ajustar o grau de similaridade dos padrões agrupados, lida com plasticidade e estabilidade, e é plausível do ponto de vista neurológico.

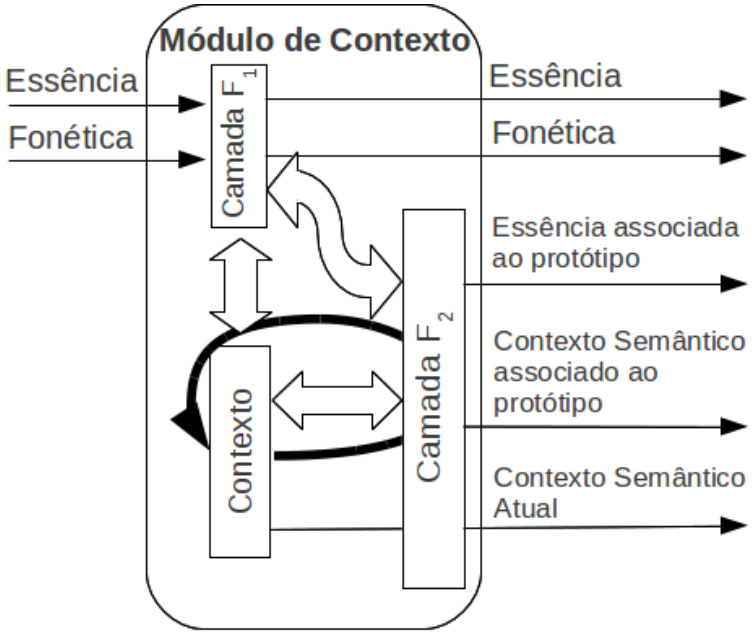

Figura 3.3: Arquitetura do Módulo de Contexto.

As camadas de entrada e saída do módulo são denominadas F_1 e F_2, respectivamente. A camada F_1 contém as unidades $p_i, q_i, r_i, s_i, u_i, v_i, w_i, x_i$, que formam os vetores **P**, **Q**, **R**, **S**, **U**, **V**, **W**, **X**. Estas unidades estão conectadas entre si, servindo para processar o padrão de entrada, redimensionando-o e aplicando supressão de ruído. Os nodos **W** recebem o padrão de entrada, enquanto o vetor **P** é responsável pela interface com a camada F_2. Já as unidades de **R** verificam se a correspondência entre a entrada e os pesos retornados da camada de saída através de **P** ultrapassa ou não o parâmetro de vigilância ρ.

A camada F_2 é formada por unidades $\mathbf{Y}_j$, onde j varia no intervalo $[0, n]$ e n corresponde ao número de protótipos da camada. Cada um destes nodos da camada de saída possui uma matriz de pesos *bottom-up* B e uma matriz *top-down* T, contendo, respectivamente, valores $b_{i,j}$ e $t_{i,j}$.

A camada adicional, que diferencia o módulo da rede ART2 original, é a de contexto, que armazena um histórico da essência dos estímulos recebidos. Essa nova camada implica na existência de unidades análogas às da camada F_1: as variáveis uc_i, que recebem o valor de u_i e formam o atual contexto através de conexões recorrentes de peso *back*, e pc_i, que realizam a interface da camada de contexto com a de saída F_2. Além disso, cada nodo na camada F_2, $\mathbf{Y}_j$, passa a ter um par correspondente que memoriza, por sua vez, o contexto semântico em que os estímulos ocorreram. Essas novas unidades de saída $\mathbf{YC}_j$ também possuem matrizes de peso *bottom-up*, BC, e

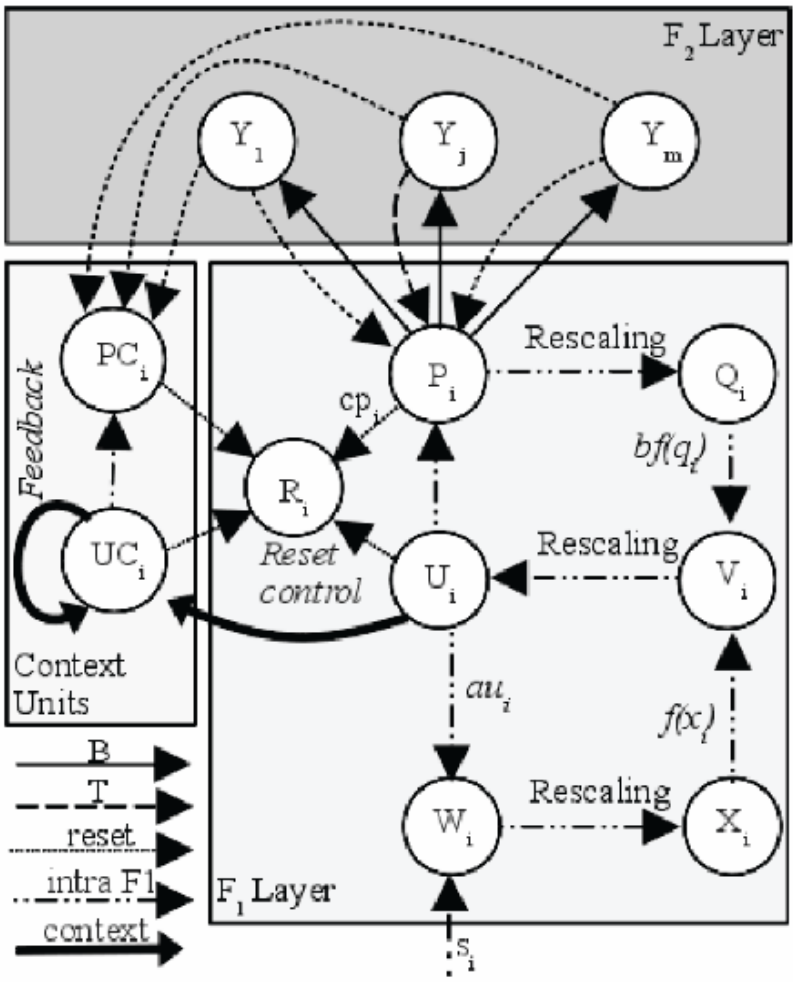

Figura 3.4: Unidades do Módulo de Contexto, baseadas na teoria ART2.

top-down, TC, que contém os valores $bc_{i,j}$ e $tc_{i,j}$, respectivamente. Através desta nova camada, pretende-se que o contexto formado e armazenado aproxime o significado geral de uma lista de palavras, supostamente semelhante ao distraidor crítico da mesma.

A Figura 3.4 permite a visualização da arquitetura do módulo de contexto, as unidades pertencentes a cada camada e suas relações.

O algoritmo de treinamento do módulo de contexto é descrito no Algoritmo 1. Basicamente, durante a fase de treinamento (ou memorização), a informação de essência do estímulo é comparada com a informação armazenada na camada F_2 escolhida para representá-la. Caso a similaridade entre ambas seja superior ao parâmetro de vigilância ρ, então o nodo equivalente da camada de saída é adaptado para aproximar o padrão de entrada, da mesma forma que o nodo de contexto associado será adaptado para

Algorithm 1 Treinamento do Módulo de Contexto

Inicialize a camada F_2 contendo $n = 200$ unidades, com pesos aleatórios $b_{i,j} \in [-8 \times 10^{-4}, 8 \times 10^{-4}]$ e pesos $t_{i,j} = 0$
Inicialize os parâmetros $a = 10$, $b = 10$, $c = 0.1$, $d = 0.9$, $e = 10^{-5}$, $\alpha = 0.7$, $\theta = 10^{-3}$, $\rho = 0.989$, $numEpochs = 1$, $numIterations = 1$, $back = 0.9$, $contextWeight = 0$, $d_{context} = 0.9$, $\alpha_{context} = 0.7$
repeat
  for cada vetor de entrada s **do**
    Execute a inicialização das unidades de F_1 (Algoritmo 2)
    Faça $reset = true$
    while reset **do**
      Encontre a unidade de F_2 com maior saída: $Y_J = max[Y_j], 1 \leq j \leq q$
      if $Y_J = -1$ **then**
        $J =$ o índice de uma unidade inativa
        $reset = false$
      end if
      if reset **then**
        $ui = vi/(e + ||v||)$, $p_i = u_i + dt_{J,i}$, $pc_i = tc_{J,i}$
        $r_i = (u_i + cp_i + contextWeight)/(e + ||u|| + c||p|| + contextWeight||pc||)$
        if $||r|| < (\rho - e)$ **then**
          $reset = true, Y_J = -1$
        else
          $reset = false, w_i = s_i + au, x_i = w_i/(e + ||w||)$
          $q_i = p_i/(e + ||p||), v_i = f(x_i) + bf(q_i)$
        end if
      else
        repeat
          Execute a adaptação da unidade vencedora (Algoritmo 3)
        until $numIterations$
      end if
    end while
  end for
until $numEpochs$

aproximar o contexto naquele instante. Por outro lado, caso a similaridade seja inferior à vigilância, o padrão de entrada é armazenado num novo nodo, criado para que nenhuma informação prévia seja modificada inapropriadamente.

A função de supressão de ruído $f(x)$, utilizada na camada F_1 do módulo, é definida da seguinte forma:

$$f(x) = \begin{cases} x & \text{if } x \geq \theta \\ 0 & \text{if } x < \theta \end{cases}$$

O reconhecimento de padrões é executado da mesma forma que o treinamento neste módulo, com algumas exceções. Primeiramente, não ocorre armazenamento

Algorithm 2 Treinamento do Módulo de Contexto (Inicialização de F_1)

1: Inicialize a ativação das unidades de F_1: $u_i = 0$, $w_i = s_i$, $p_i = 0$, $q_i = 0$,
2: $x_i = s_i/(e + ||s||)$, $v_i = f(x_i)$
3: Atualize a ativação das unidades de F_1: $u_i = v_i/(e + ||v||)$, $w_i = s_i + au_i$, $p_i = ui$,
4: $x_i = w_i/(e + ||w||)$, $q_i = p_i/(e + ||p||)$, $v_i = f(x_i) + bf(q_i)$
5: Propague os valores para as unidades UC: $uc_i = (back)(uc_i) + (1 - back)f(u_i)$
6: Normalize os valores de UC: $uc_i = uc_i/(e + ||uc||)$
7: Propague esses valores para as unidades PC: $pc_i = uc_i$
8: Calcule as ativações das unidades de F_2: $Y_j = (1 - contextWeight)\sum_i b_{i,j}p_i + contextWeight\sum_i(bc_{i,j}pc_i)$
9: {Unidades inativas de F_2 devem possuir ativação igual a -1}

Algorithm 3 Treinamento do Módulo de Contexto (Adaptação de F_2)

1: Atualize a unidade vencedora Y_J:
2: $t_{J,i} = \alpha du_i + [1 + \alpha d(d-1)]t_{J,i}$
3: $b_{i,J} = \alpha du_i + [1 + \alpha d(d-1)]b_{i,J}$
4: $tc_{J,i} = \alpha_{context}d_{context}uc_i + [1 + \alpha_{context}d_{context}(d_{context} - 1)]tc_{J,i}$
5: $bc_{i,J} = \alpha_{context}d_{context}uc_i + [1 + \alpha_{context}d_{context}(d_{context} - 1)]bc_{i,J}$
6: Normalize os vetores de pesos:
7: $t_{J,i} = t_{J,i}/||t_J||$
8: $b_{i,J} = b_{i,J}/||b_J||$
9: $tc_{J,i} = tc_{J,i}/||tc_J||$
10: $bc_{i,J} = bc_{i,J}/||bc_J||$
11: Atualize as unidades de F_1:
12: $u_i = v_i/(e + ||v||)$, $w_i = s_i + au_i$, $p_i = ui + dt_{J,i}$,
13: $x_i = w_i/(e + ||w||)$, $q_i = p_i/(e + ||p||)$, $v_i = f(x_i) + bf(q_i)$

na camada F_2. Além disso, realiza-se uma adaptação no parâmetro de vigilância ρ a cada padrão apresentado. Começando com o valor máximo de 1 (que permite apenas equivalências entre a entrada e o protótipo armazenado), ele é reduzido suavemente, a uma taxa de 5%, a cada tentativa falha de busca por um nodo contendo um protótipo suficientemente similar ao padrão de entrada.

Quando uma boa representação para o estímulo é encontrada, as informações semânticas e contextuais armazenadas no nodo equivalente, e seu par, são propagadas para a saída do módulo de contexto. Esses dados são formados por: a representação de essência $\mathbf{x}_{gist}$ do estímulo, a sua representação literal $\mathbf{x}_{verbatim}$, o protótipo semântico recuperado $\mathbf{w}^c_{gist}$, o contexto associado a este protótipo $\mathbf{w}^c_{context}$ e o contexto atual $\mathbf{x}_{context}$.

3.4 Módulo de Essência

O Módulo de Essência (ou módulo *gist*) recebe a representação semântica do estímulo e as três saídas produzidas pelo módulo de contexto, para, desta forma, construir protótipos dos padrões de entrada. Propaga, então, para o resto do sistema, a essência associada ao protótipo vencedor, numa competição que considera também a parcela contextual dos dados. Além disso, libera o grau de certeza em sua resposta, correspondente à probabilidade de que o padrão de entrada tenha sido corretamente associado ao protótipo escolhido.

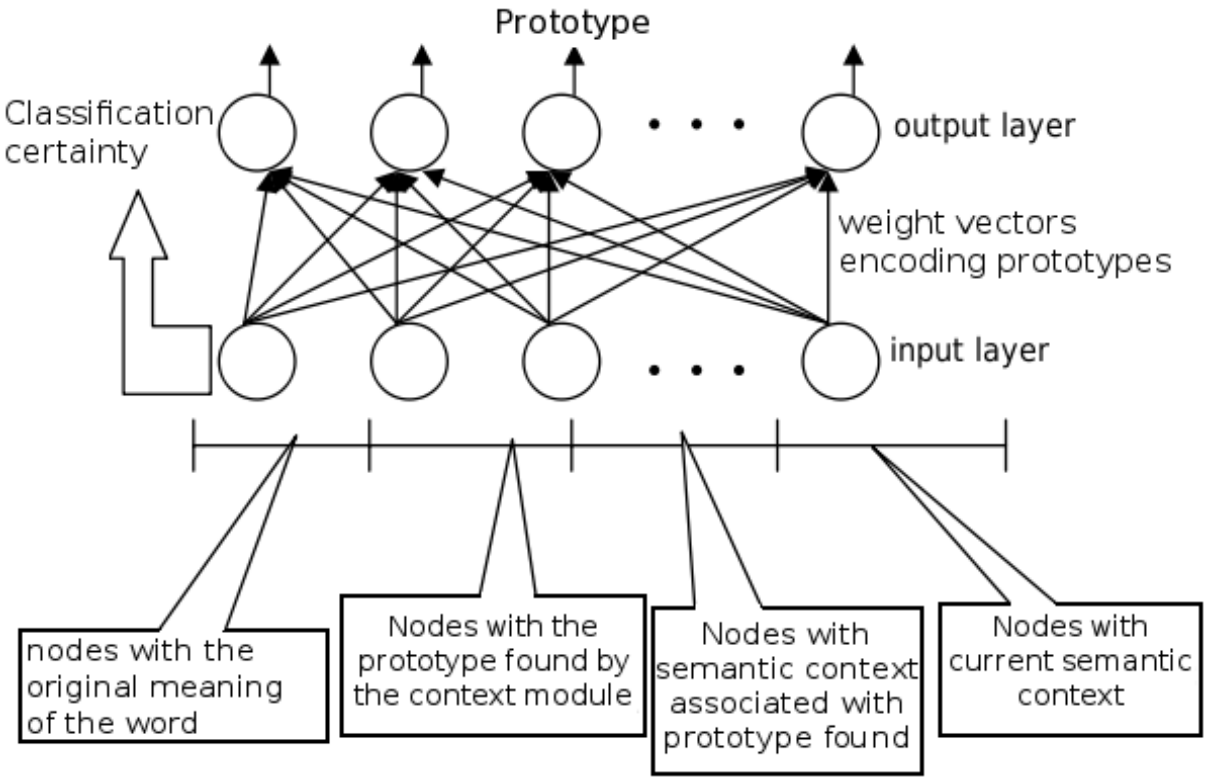

Figura 3.5: Arquitetura do Módulo de Essência.

Durante a fase de treinamento, o módulo *gist* é responsável por gerar protótipos do significado dos alvos estudados, sem levar em conta detalhes literais (*verbatim*). Dessa forma, uma alta similaridade entre itens de uma lista prediz seu armazenamento num mesmo protótipo, que codifica as características semânticas mais importantes compartilhadas por ambas as palavras.

Também espera-se que palavras de uma mesma lista formem contextos semelhantes, e palavras com contextos distantes pertençam a listas distintas. Essa é a razão pela qual o módulo de essência considera tanto a representação *gist* do estímulo quanto os dados

de contexto semântico associados ao mesmo. Além disso, o módulo incorpora a idéia de discriminação de informação nova, criando um novo protótipo, ao invés de adaptar o protótipo vencedor, caso o contraste entre a entrada e a informação previamente armazenada seja significativo.

O algoritmo que inspira este módulo é o LVQ (*Learning Vector Quantization*) (Kohonen, 1989), introduzido como um método supervisionado para ajuste fino de mapa auto-organizável (SOM). Apesar de o módulo de essência ser naturalmente não-supervisionado, um detalhe específico durante o seu treinamento é inspirado pelo paradigma supervisionado de aprendizagem, como veremos adiante. Adicionalmente, a avaliação do grau de certeza na tarefa de reconhecimento segue a abordagem proposta pelo SNSI (*Stochastic Neural Sequence Identifier*) (Araújo e Henriques, 2002) para obtenção de probabilidades de associação.

3.4.1 Módulo de Essência: Arquitetura, Representação e Regras de Propagação

Na Figura 3.5, pode-se observar a estrutura do módulo, baseada na rede SOM. Todos os valores de entrada estão conectados a todos os valores de saída; e estas conexões possuem pesos que representam os protótipos de cada categoria.

Os nodos de entrada formam o padrão $\mathbf{x}_{in}^{g}$, equivalente à concatenação da codificação de essência $\mathbf{x}_{gist}$ e das três saídas do Módulo de Contexto (o protótipo semântico recuperado $\mathbf{w}_{gist}^{c}$, o contexto associado a esse protótipo $\mathbf{w}_{context}^{c}$ e o contexto atual $\mathbf{x}_{context}$). Todos os valores de entrada pertencem ao intervalo real [-1,1], decorrente da representação adotada para os dados (Seção 3.2.2). A matriz de pesos da rede $\mathbf{W}^{g}$, portanto, também deve apresentar valores neste intervalo.

O número de nodos na camada de entrada é N_{in}^{g}, onde $N_{in}^{g} = 4 \times 129$, uma vez que cada um dos vetores de entrada possui 129 características. Já a camada de saída contém um número arbitrário de nodos N_{out}^{g}, cujos pesos devem ser inicializados no início da execução do módulo. Neste trabalho, utilizou-se um valor $N_{out}^{g} = 200$, enquanto a matriz de pesos $\mathbf{W}^{g}$, de dimensões $N_{in}^{g} \times N_{out}^{g}$, foi aleatoriamente inicializada com valores contidos no intervalo [-0.005,0.005].

O treinamento do módulo de essência ocorre em duas etapas: construção do mapa auto-organizável e ajuste da probabilidade de associação entrada-protótipo. A ativação de cada nodo considera a distância Euclidiana entre o padrão de entrada $\mathbf{x}_{in}^{g}$ e o vetor $\mathbf{w}_{i}^{g}$ que representa o i-ésimo protótipo, da seguinte forma:

Algorithm 4 Treinamento do Módulo de Essência

Inicialize a matriz de pesos $\mathbf{W}^g$ com valores aleatórios entre [-0.005,0.005]
Inicialize a taxa de aprendizagem, $\eta = 0.3$, e o limiar de resposta, $r_{min} = 0$
for cada padrão do conjunto de treinamento **do**
  Normalize o vetor de codificação semântica recebido pelo módulo $\mathbf{x}_{gist}$
  Defina o padrão $\mathbf{x}^g_{in}(t)$ como uma concatenação dos vetores de entrada $\mathbf{x}_{gist}$, $\mathbf{w}^c_{gist}$, $\mathbf{w}^c_{context}$ e $\mathbf{x}_{context}$ (Figura 3.5)
  if houve um intervalo longo sem apresentação de padrão **then**
    aumente r_{min} (Equação 3.5)
  end if
  Calcule a resposta (ativação) de cada nodo na camada de saída (Equação 3.3)
  Defina como vencedor o nodo ativo mais próximo a $\mathbf{x}^g_{in}(t)$
  if a ativação do nodo vencedor é menor que r_{min} **or** todos os nodos estão inativos **then**
    Defina como vencedor o nodo inativo mais próximo
    Atualize a resposta mínima r_{min} (Equação 3.4)
    Ajuste a taxa de aprendizagem para que o padrão de entrada seja copiado: $\eta = 1$
    Ajuste o protótipo do nodo vencedor (Equação 3.6)
    Retorne a taxa de aprendizagem para seu valor inicial: $\eta = 0.3$;
  else
    Atualize a resposta mínima r_{min} (Equação 3.4)
    Ajuste o protótipo do nodo vencedor (Equação 3.6)
    Normalize os vetores componentes do nodo vencedor separadamente
    Reduza a taxa de aprendizagem: $\eta = \eta * 0.95$
  end if
end for

$$R_i = 1/(\|\mathbf{x}^g_{in} - \mathbf{w}^g_i\| + e) \tag{3.3}$$

onde, $e = 10^{-9}$, é usado para se evitar divisão por zero.

Supondo que a maior ativação dentre os nodos ativos da rede (que já tiveram pelo menos um padrão de entrada associado a si) tenha sido, R^{max}_i, é verificado se este valor é superior a um limiar dinâmico r_{min}. Caso o seja, o protótipo do nodo vencedor é adaptado; caso não, entende-se que não há nodo ativo suficientemente próximo do padrão de entrada, o que leva à adaptação do nodo inativo mais semelhante ao mesmo e à vitoria de sua primeira competição. Essa regra visa preservar os protótipos já treinados, de forma que as características armazenadas de longo termo não sejam perdidas.

Após a apresentação de cada padrão de entrada, o limiar dinâmico é atualizado de acordo com as equações a seguir:

Algorithm 5 Algoritmo de Ajuste da Probabilidade de Gibbs (Ajuste de β)

Construa o conjunto de teste com padrões aleatoriamente selecionados do conjunto de treinamento
Descarte os nodos inativos (que nunca venceram no treinamento)
Faça $l_{doubt} = 0.6$, $l_{context} = 0.25$
for cada possível valor de β **do**
  $C_{total} = 0$
  for cada padrão do conjunto de teste **do**
    Calcule a resposta de cada nodo da camada de saída (Equação 3.3)
    Determine o nodo vencedore $\mathbf{w}_s$ (Equation 3.8)
    Calcule a probabilidade de Gibbs ($P(c_i/\mathbf{x})$) (Equação 3.7)
    if $P(c_i/\mathbf{x}) > l_{doubt}$ (Equação 3.9) **then**
      Calcule a distância contextual entre o protótipo e a entrada: $d^c = \|\mathbf{w}_s^c - \mathbf{x}^c\|$
      if $d^c \leq l_{context}$ **then**
        $C = 0$
      else
        $C = 1$
      end if
    else
      $C = 1 - P(c_i/\mathbf{x})$
    end if
    $C_{total} = C_{total} + C$
    if C_{total} é o menor custo encontrado **then**
      Memorize o valor de β associado a este custo
    end if
  end for
  Iguale β ao valor que gerou o menor custo total C_{total}
end for

$$r_{min}(t) = \begin{cases} \alpha_r^g \lambda R_i^{max}(t) + (1 - \alpha_r^g)\lambda r_{min}(t-1) \\ \qquad\qquad \text{if } R_i^{max}(t) \geq r_{min}(t-1) \\ R_i^{max}(t) \\ \qquad\qquad \text{otherwise} \end{cases} \tag{3.4}$$

onde $R_i^{max}(t)$ é a resposta (ativação) do nodo vencedor na iteração t, obtido da maximização da Equação 3.3, α_r^g é a taxa de aprendizagem e $\lambda \in [0,1]$ controla a influência das últimas respostas. Esses parâmetros foram ajustados experimentalmente, assumindo os valores $\alpha_r^g = 0.07$ e $\lambda = 0.9$.

Na inicialização do módulo de essência, r_{min} é atribuído um valor nulo, uma vez que ainda não se possui informação acerca da apresentação de um tema específico. Dessa forma, e também por todos os nodos estarem inativos, o protótipo mais semelhante à

Algorithm 6 Algoritmo de Reconhecimento do Módulo de Essência

1: **for** cada padrão $\mathbf{x}(t)$ de entrada **do**
2: Calcule a resposta de cada unidade de saída (Equação 3.3)
3: Encontre o nodo vencedor $\mathbf{w}_s$ (Equação 3.8)
4: Calcule a probabilidade de Gibbs $P(c_i/\mathbf{x})$ (Equação 3.7)
5: Retorne a probabilidade calculada como o grau de certeza do módulo ($P(c_i/\mathbf{x})$)
6: Retorne o vetor de essência armazenado em $\mathbf{w}_s$, isto é, a parte do protótipo que armazenou a representação de essência x_{gist} (uma das entradas do módulo)
7: **end for**

entrada será o vencedor.

Para simular o efeito de pausa entre a apresentação de listas de palavras (que, nos experimentos do paradigma DRM, representa um intervalo de 10 segundos), adota-se a Equação 3.5. Através dela, o limiar r_{min} sofre um aumento proporcional ao seu valor anterior, dificultando a associação de padrões de entrada com protótipos já formados.

$$r_{min}(t) = (1 + \alpha_r^g) r_{min}(t-1) \tag{3.5}$$

Após determinado o nodo vencedor, o mesmo é atualizado de acordo com a Equação 3.6:

$$\mathbf{w}_i^g(t) = (1-\eta)\mathbf{w}_i^g(t-1) + \eta \mathbf{x}_{in}^g(t) \tag{3.6}$$

onde $\mathbf{x}_{in}^g(t)$ representa o padrão de entrada no instante t e η corresponde à taxa de aprendizagem da rede, entre [0,1]. Essa taxa é inicializada em 0.3 e decai 5% a cada padrão apresentado. No entanto, é retornada ao valor inicial quando um novo protótipo é ativado. O objetivo desta dinâmica é reproduzir o efeito de discriminação do teor de novidade do estímulo, que, no caso de uma informação inédita, aumenta a força com que ela é memorizada e, na ocorrência de informações repetitivas, induz à habituação.

O Algoritmo 4 apresenta em detalhes o passo-a-passo do processo de treinamento deste módulo.

A segunda fase do treinamento do módulo de essência, como proposto por Pacheco (2004), consiste do ajuste dos valores de probabilidade de classificação calculados. Neste trabalho, no entanto, modificamos a forma como isso é feito, utilizando todas as três saídas produzidas pelo módulo de contexto, ao invés de apenas uma delas, durante o cálculo da função de custo.

O parâmetro relativo à definição dessas probabilidades, β, deve ser ajustado

A probabilidade de a associação entre um padrão de entrada $\mathbf{x}$ e o protótipo $\mathbf{w}_i$,

representando a classe c_i, estar correta pode ser calculada de acordo com a **distribuição de Gibbs** (*Gibbs distribution*) (Haykin, 1999):

$$P(c_i/\mathbf{x}) = \frac{\exp(-\beta d_i)}{\sum_j exp(-\beta d_j)} \tag{3.7}$$

onde β indica o inverso da temperatura e corresponde ao parâmetro que será ajustado durante esta fase do treinamento do módulo de essência. Temos $d_i = \|\mathbf{x} - \mathbf{w}_i\|$, $d_j = \|\mathbf{x} - \mathbf{w}_j\|$, $j = 1,2,..,nc$ e nc equivalente ao número de protótipos formados no módulo.

A distribuição de probabilidade de Gibbs tende à uniformidade quanto mais próximo o parâmetro β for de 0. De forma oposta, à medida que este parâmetro cresce, a probabilidade de associação correta entre a entrada e o protótipo mais próximo dispara, tendendo a 100%, tornando mais difícil discriminar os outros protótipos entre si.

Nesse sentido, deve-se escolher um valor apropriado para β, o que é feito através de uma função de custo $C(c_s,\mathbf{x})$. Seja c_s a classe do protótipo mais próximo à entrada $\mathbf{x}$, $\mathbf{w}_s$ (encontrado através da Equação 3.8), e n o número de protótipos formados, a função de custo pode ser calculada de acordo com a Equação 3.9:

$$\mathbf{w}_s, \text{ such that } \|\mathbf{x} - \mathbf{w}_s\| < \|\mathbf{x} - \mathbf{w}_i\|, \text{ for } \forall i \neq s, i = 1,...,n \tag{3.8}$$

$$C(c_s,\mathbf{x}) = \begin{cases} 0 & \text{if } \|\mathbf{w}_s^c - \mathbf{x}^c\| \leq l_c \\ & \text{and } P(\mathbf{w}_s/\mathbf{x}) > l_{doubt} \text{ (classe correta)} \\ 1 & \text{if } \|\mathbf{w}_s^c - \mathbf{x}^c\| > l_c \\ & \text{and } P(\mathbf{w}_s/\mathbf{x}) > l_{doubt} \text{ (classe incorreta)} \\ 1 - P(\mathbf{w}_s/\mathbf{x}) & \text{if } P(\mathbf{w}_s/\mathbf{x}) \leq l_{doubt} \text{ (não-classificado)} \end{cases} \tag{3.9}$$

onde $\mathbf{x}^c$ representa a concatenação das três saídas produzidas pelo módulo de contexto, $\mathbf{w}^c_{gist}$, $\mathbf{w}^c_{context}$ e $\mathbf{x}_{context}$, isto é, $\mathbf{x}^c$ representa a informação contextual do vetor de entrada $\mathbf{x} \equiv \mathbf{x}^g_{in}$. Dois parâmetros são definidos para o cálculo do custo de β: l_c representa a distância mínima entre a informação contextual de entrada e a do protótipo para que a classificação seja considerada correta, e l_{doubt} consiste do limiar de dúvida, abaixo do qual considera-se que o classificador admitiu indecisão e reduz-se a penalidade associada ao erro.

Através da Equação 3.9, a inspiração do módulo na aprendizagem supervisionada do algoritmo LVQ fica clara, em razão da utilização de informação contextual no julgamento da resposta do classificador treinado.

O Algoritmo 5 consiste de calcular o valor ótimo de β dentro de um conjunto de possibilidades, buscando-se minimizar a soma do custo de diferentes classificações. Experimentos realizados mostraram que, em geral, para os limiares de contexto e decisão escolhidos, este parâmetro foi ajustado para um valor localizado entre $[1.0, 2.0]$.

A operação do módulo de essência, após definidos os protótipos, é descrita no Algoritmo 6. As saídas propagadas ao resto do sistema consistem do vetor de essência associado ao protótipo vencedor, $\mathbf{w}^g_{gist}$, e ao grau de certeza na resposta de classificação, DC^g.

3.5 Módulo de Literalidade

O Módulo de Literalidade possui a função de formar protótipos da informação literal (*verbatim*) do estímulo. Suas entradas são três: a representação literal $x_{verbatim}$ de entrada do sistema; o contexto associado ao protótipo vencedor do módulo de contexto, $w^c_{context}$; e o contexto atual calculado pelo módulo de contexto, $x_{context}$.

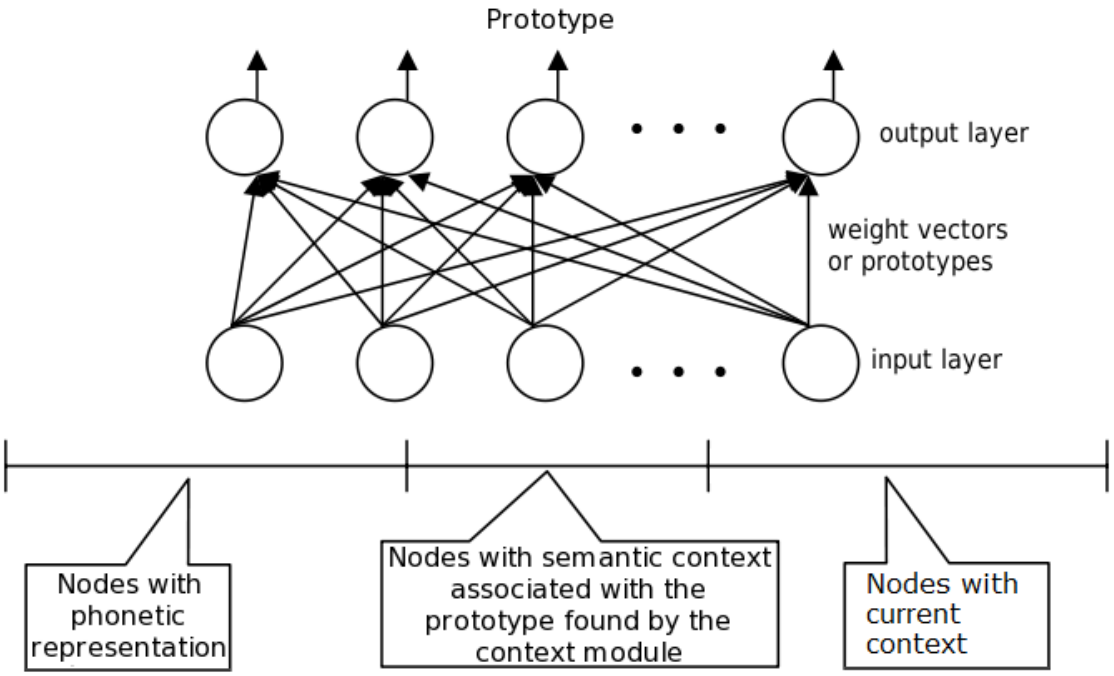

Figura 3.6: Arquitetura do Módulo de Literalidade.

De acordo com a Teoria do Rastro Difuso, o julgamento de identidade ocorre quando a familiaridade com o item apresentado induz à extração de traços literais do material estudado, que coincide com o estímulo recebido, resultando no reconhecimento. O processo de não-identidade também pode ocorrer, decorrendo da extração de traços

Algorithm 7 Treinamento do Módulo de Literalidade

Inicialize um conjunto de N_{max} centros inativos
for cada padrão de entrada **do**
  if não houver centros ativos **then**
    Determine os $N_{candidates}$ centros **inativos** mais próximos ao padrão de entrada considerando a entrada literal e as duas entradas contextuais (Equação 3.11)
    Dentre esses, encontre o candidato mais próximo considerando apenas a informação literal (Equação 3.10)
    Desloque o centro vencedor para a posição do padrão de entrada
  else
    Aplique o mecanismo de decaimento da memória (Equação 3.13)
    Determine os $N_{candidates}$ centros **ativos** mais próximos ao padrão de entrada considerando a entrada literal e as duas entradas contextuais (Equação 3.11)
    Dentre esses, encontre o candidato mais próximo considerando apenas a informação literal (Equação 3.10)
    if a distância **literal** do centro mais próximo for suficiente (abaixo do limiar σ) **then**
      Aproxime o centro vencedor do padrão de entrada (Equação 3.12)
    else
      Determine os $N_{candidates}$ centros **inativos** mais próximos ao padrão de entrada considerando a entrada literal e as duas entradas contextuais (Equação 3.11)
      Dentre esses, encontre o candidato mais próximo considerando apenas a informação literal (Equação 3.10)
      Desloque o centro vencedor para a posição do padrão de entrada
    end if
  end if
end for

literais associados aos alvos estudados, que levam à constatação de que o estímulo é novo e não foi experienciado previamente. Percebe-se que esses julgamentos são passíveis de modelagem através do módulo de literalidade, uma vez que os sinais de entrada do módulo contém dados literais e de contexto semântico.

3.5.1 Módulo de Literalidade: Arquitetura, Representação e Regras de Propagação

A arquitetura do módulo de literalidade baseia-se no Hipermapa (*Hypermap*) (Kohonen et al., 1991). Nesta rede neural, utiliza-se duas medidas de distância entre o valor de entrada e os protótipos armazenados: uma é aplicada inicialmente para a escolha de um subconjunto de nodos candidatos a vencer a competição, formado pelos $N_{candidates}$ centros mais próximos de acordo com esta distância; a segunda medida é usada para selecionar o protótipo vencedor, dentre os candidatos obtidos.

Como o módulo de literalidade recebe uma vetor de representação literal e dois

Algorithm 8 Reconhecimento do Módulo de Literalidade

1: **for** cada padrão de entrada **do**
2: Aplique o mecanismo de decaimento de memória (Equação 3.13)
3: Determine os $N_{candidates}$ centros ativos mais próximos ao padrão de entrada considerando a entrada literal e as duas entradas contextuais (Equação 3.11)
4: Dentre esses, encontre o candidato mais próximo considerando apenas a informação literal (Equação 3.10)
5: **end for**

vetores de conteúdo semântico, usaremos a distância total entre a entrada e os centros da rede para selecionar os candidatos a vencer a competição. Em seguida, aplicaremos apenas a informação literal para discriminar os nodos candidatos, conforme a idéia da FTT de traços *gist* influenciando na extração de memória *verbatim* relativa à material estudado.

A distância entre o vetor literal de entrada e a informação literal de um centro c_i^v será dada por:

$$d^v = \|\mathbf{x}^v - \mathbf{c}_i^v\| / \|\mathbf{x}^v\| \tag{3.10}$$

Por sua vez, a distância entre os dois vetores contextuais recebidos do módulo de contexto e a informação correspondente prototipada num dos centros é:

$$\mathbf{d}^c = \|\mathbf{x}^c - \mathbf{c}_i^c\| \tag{3.11}$$

O processo de seleção de candidatos ocorre tanto durante o treinamento dos protótipos, na fase de estudo, quanto durante a fase de reconhecimento, na qual não há adaptação do nodo vencedor. Os procedimentos para ambas as situações estão descritos nos Algoritmos 7 e 8. O valor adotado para $N_{candidates}$ foi 7.

Durante o treinamento, a adaptação do centro vencedor foi definida com uma taxa de aprendizagem $\alpha^v = 0.8$ (Equação 3.12).

$$\mathbf{c}_i^v = \mathbf{c}_i^v(1 - \alpha^v) + \mathbf{x}^v\alpha^v \tag{3.12}$$

Uma peculiaridade da fase de treinamento é que, como o hipermapa baseia-se em centros que aproximam-se, ao longo do tempo, das posições dos padrões apresentados, deve ser definida uma distância literal mínima σ, para discriminar protótipos que, apesar de vencedores, não são suficientemente próximos da entrada. Isso induz à ativação de protótipos que encontravam-se inativos, melhorando a representatividade do material estudado pelos centros do hipermapa. O valor atribuído para o parâmetro σ foi 0.6.

Pacheco (2004), ao propor o algoritmo deste módulo, introduziu um dos aspectos que diferenciam a memória *verbatim* da memória *gist*: a maior taxa de decaimento. Isso foi feito através da adição de um ruído aleatório a cada um dos centros, toda vez que um padrão fosse apresentado durante o treinamento e durante o reconhecimento. O ruído, no entanto, não foi intuitivamente definido, pois apenas 10% de suas posições poderiam conter valores diferentes de nulo.

Neste trabalho, propomos a aplicação de um ruído aleatório de distribuição gaussiana, com média 0 e desvio padrão igual a 2 (valor este que foi definido experimentalmente). Os centros formados, dessa forma, sofrerão interferência a curto prazo, possibilitando que novos padrões sejam associados aos mesmos e tornando a memorização no módulo de literalidade mais instável, de um modo geral, que o armazenamento no módulo de essência.

A adição do ruído é feita através da seguinte equação:

$$\mathbf{c}_i^v(t) = \epsilon \mathbf{c}_i^v(t-1) + [1-\epsilon].[\mathbf{c}_i^v(t-1)[1-\psi] + \mathbf{nv}\psi] \tag{3.13}$$

onde $\epsilon = 0.3$ e $\psi = 0.002$ representam, respectivamente, a proporção de valor residual dos centros que não é esquecida e a velocidade com que o centro de cada agrupamento é esquecido.

Apesar de os centros do módulo de literalidade armazenarem tanto informação literal quanto contextual, apenas o conteúdo *verbatim* é transmitido adiante para o resto do nosso sistema. Ou seja, a saída do sistema consiste do protótipo $w^v_{verbatim}$, onde o índice v indica que o protótipo foi produzido pelo módulo de literalidade (*verbatim*).

A estrutura do módulo de literalidade pode ser visualizada na Figura 3.6.

3.6 Módulo de Decisão

Como último componente do sistema, o Módulo de Decisão é responsável pelo julgamento do estímulo de entrada. Do módulo de contexto, recebe as representações de essência e fonética x_{gist} e $x_{verbatim}$. Do módulo de essência, adquire o valor semântico armazenado no protótipo vencedor w^g_{gist} e o grau de certeza DC^g. Por fim, o módulo de literalidade lhe fornece a informação literal do centro vencedor $w^v_{verbatim}$.

A arquitetura deste módulo é a mais simples e direta de todo o sistema: as entradas citadas alimentam uma camada com três unidades de função de base radial (RBF), cujos centros e raios são determinados dinamicamente. Cada unidade é relativa a um de três processos de julgamento da FTT: identidade, não-identidade e similaridade.

As unidades competem entre si, de forma que a decisão de reconhecimento será baseada naquela que apresentar a maior ativação. Como visto no Capítulo 2, tanto a identidade quanto a similaridade resultarão no reconhecimento do padrão; a não-identidade, por outro lado, levará à sua rejeição. A Tabela 3.2 mostra como os diferentes sinais de entrada são usados na definição da RBF dos nodos.

Tabela 3.2: Funções desempenhadas pelas entradas do Módulo de Decisão.

Variáveis\Nodos:	Identidade	Não-Identidade	Similaridade
Centro	$x_{verbatim}$	$x_{verbatim}$	x_{gist}
Raio	$\|x_{gist} - w^g_{gist}\|$	$\|x_{gist} - w^g_{gist}\|$	$\|x_{verbatim} - w^v_{verbatim}\|$
Estímulo de entrada	$w^v_{verbatim}$	$w^v_{verbatim}$	w^g_{gist}
Ajuste			DC^g

As codificações semântica e fonética de entrada do sistema representam os centros das funções RBF dos nodos de similaridade e identidade, respectivamente, pois correspondem aos sinais que devem ser usados como parâmetros ao correto reconhecimento.

Para que haja influência mútua entre os processamentos *gist* e *verbatim*, faz-se com que o raio da função de similaridade seja modulado pela semelhança literal entre a entrada do sistema e o protótipo extraído do módulo de literalidade, e vice-versa. Consequentemente, a extração de traços *verbatim* da memória semelhantes ao padrão de entrada aumentará as chances de um reconhecimento por similaridade, e traços *gist* extraídos próximos à representação de essência do estímulo facilitarão o julgamento de identidade. De forma análoga, a extração de traços memorizados distintos em qualquer dos dois casos resultará num maior grau de rigidez do julgamento paralelo.

A ativação do nodo de identidade é dada pela Equação 3.14:

$$z^v = \frac{\mu}{(\sigma^g)^2} \exp\left(-\frac{\|x_{verbatim} - w^v_{verbatim}\|}{(\sigma^g \kappa)^2}\right) \tag{3.14}$$

onde $\sigma^g = \|x_{gist} - w^g_{gist}\|$ é o raio, $\kappa = 0.45$ é o parâmetro de ajuste do raio, e $\mu = 18.806$ ajusta o tamanho da área abaixo da função de base radial, aumentando a magnitude em torno do centro. Estes dois últimos parâmetros foram definidos empiricamente, para aproximar o comportamento humano no paradigma DRM.

A saída da função de similaridade, de forma semelhante, é dada pela Equação 3.15:

$$z^g = DC^g \frac{\mu}{(\sigma^g)^2} \exp\left(-\frac{\|x_{gist} - w^g_{gist}\|}{(\sigma^v \kappa)^2}\right) \tag{3.15}$$

onde o grau de certeza DC^g é usado para aumentar a magnitude da RBF em torno do centro, $\sigma^v = \|x_{verbatim} - w^v_{verbatim}\|$ representa o raio, $\kappa = 1.1$, e $\mu = 18.806$.

Por último, a ativação da RBF relativa à não-identidade é dada por uma constante a menos da resposta de identidade, como na equação a seguir:

$$z^{nv} = m - z^v \tag{3.16}$$

onde $m = 9$ indica a ativação máxima possível para este nodo.

Quando a magnitude da RBF de identidade é muito pequena, torna-se muito provável um julgamento de rejeição por recordação, equivalente à não-identidade. Quando essa magnitude é moderada ou insuficientemente alta, o nodo também é ativado e rejeita-se o padrão, apesar de o processo de não-identidade não caracterizar bem este evento.

A razão é que, de acordo com a FTT, a não-identidade não precisa ocorrer de fato para que o padrão seja rejeitado, sendo necessário apenas que não se manifestem nenhuma das fenomenologias de familiaridade ou recordação (desconsiderando-se qualquer tipo de *bias* de decisão). Apesar de a rejeição no sistema proposto ser sempre causada pela ativação do nodo de não-identidade, a função deste nodo (Equação 3.16) é definida de forma a agregar tanto a rejeição por recordação de alvo (não-identidade) quanto a rejeição por falha de recuperação de quaisquer traços de memória, *gist* ou *verbatim*. Desta forma, seria mais preciso denominar esta unidade por unidade de "não-identidade/ indecisão". Restringimo-nos a chamá-la de "não-identidade" por motivos práticos.

4

Validação do Modelo Proposto

Neste capítulo, serão mostrados alguns dos experimentos que foram realizados a fim de validar os módulos componentes do sistema proposto para modelagem do fenômeno de falsas memórias. A Seção 4.1 analisa a performance do módulo de contexto. Na Seção 4.2, detalharemos os experimentos com o módulo de essência. Em seguida, abordaremos o desempenho do módulo de literalidade, na Seção 4.3. Por fim, na Seção 4.4, reproduziremos os dois experimentos do paradigma DRM (Brainerd e Reyna, 1998b), e analisaremos a eficácia com que o novo sistema modela os processos cognitivos de memorização e reconhecimento.

4.1 Validação do Módulo de Contexto

As simulações do módulo de contexto visam à análise da capacidade deste módulo de agrupar os padrões de entrada com base no significado semântico e de produzir e associar informação contextual relevante aos mesmos.

Primeiramente, na Seção 4.1.1, realizamos um experimento para verificar se o contexto semântico formado pelo módulo representa bem o significado geral de uma lista de palavras sendo apresentada. Já na Seção 4.1.2, analisamos a operação regular do módulo durante a apresentação de padrões treinados e não-treinados.

4.1.1 Avaliação da Formação de Contexto

Nesta simulação, as 24 listas de palavras do paradigma DRM (Roediger e McDermott, 1995) foram apresentadas ao módulo de contexto. O objetivo foi o de calcular a distância entre o contexto atual e o padrão sendo apresentado num determinado instante, para verificar se o contexto semântico tende a se aproximar do significado geral de uma lista com o tempo.

A cada entrada, o vetor *gist* foi comparado com o contexto existente antes de sua apresentação, através do cálculo do cosseno do ângulo entre os vetores (Equação 4.1).

$$cos(\mathbf{x},\mathbf{y}) = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2}\sqrt{\sum_i y_i^2}} \tag{4.1}$$

Os resultados (Figura 4.1) mostram que a distância entre o vetor de entrada e o contexto tende a decrescer ao longo do estudo de uma lista (e o cosseno tende a crescer), aumentando bruscamente quando uma nova lista começa a ser apresentada (o cosseno diminui nesse caso). Isso significa que o vetor contextual do módulo está gerando o comportamento desejado, aproximando-se do significado comum de uma sequência específica de palavras.

No entanto, algumas oscilações podem ser percebidas, como palavras que aparentemente são intrusas nas listas *chair* e *music* e *sleep*, por exemplo. Além disso, listas como *girl* e *man* aparentam ser bastante heterogêneas em termos de essência. A razão é que algumas das palavras das listas DRM são associadas a outros conceitos que não os da lista quando apresentadas fora de contexto em testes de associação livre, como o que foi realizado para gerar a base de normas associativas por trás da representação de essência adotada (Nelson e Schreiber, 1998). Dessa forma, acabam existindo exceções dentro das listas, como a palavra *molasses* (melaço), por exemplo, que no paradigma DRM pertence à lista *slow*, mas que é codificada com o sentido de *sweet* na nossa base de palavras.

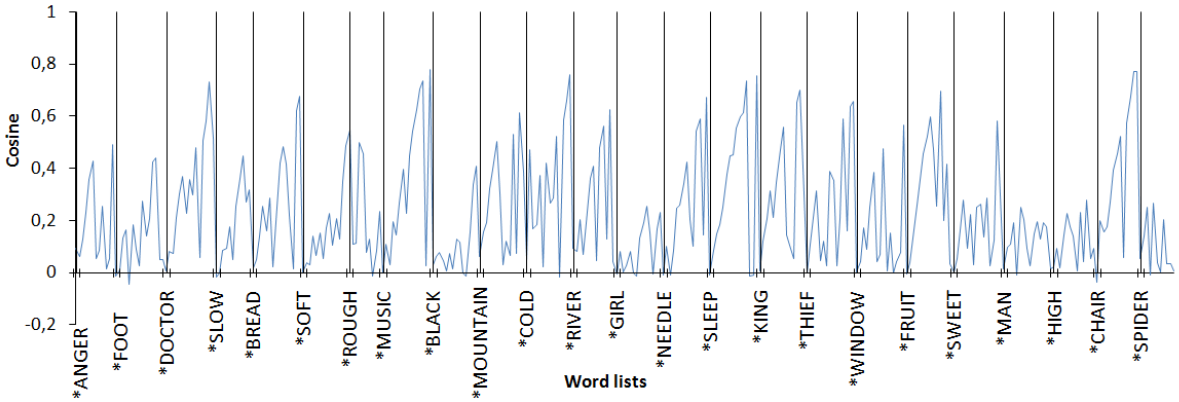

Figura 4.1: Cosseno entre os padrões apresentados e o contexto formado naquele instante: o eixo x mostra os temas de cada lista (distraidores críticos) na posição da primeira palavra da lista.

4.1.2 Operação Regular

A capacidade de generalização do módulo de contexto foi avaliada através do treinamento com quatro listas (*foot*, *fruit*, *black* and *chair*) e do reconhecimento destas e da lista *bread*, não-estudada. Nessa simulação, o parâmetro $\rho = 0.989$ permite que agrupamentos entre palavras semanticamente muito semelhantes sejam formados.

A Tabela 4.1 lista as palavras associadas aos agrupamentos produzidos no treinamento. O reconhecimento é descrito nas Tabelas 4.2 e 4.3, relativas respectivamente à palavras treinadas e não-treinadas.

Tabela 4.1: Agrupamentos formados na camada F_2 após treinamento do módulo com as palavras das listas *Foot*, *Fruit*, *Black* e *Chair*. As palavras que foram associadas ao mesmo protótipo estão entre parênteses.

1(shoe, toe, ankle, inch, sock), 2(hand, arm), 3(kick, soccer, basket), 4(sandals, boot), 5(yard), 6(walk), 7(smell), 8(mouth), 9(apple, orange, kiwi, citrus, pear, banana, berry, cherry, juice), 10(vegetable, salad), 11(bowl), 12(cocktail), 13(white, coal, brown, gray), 14(dark), 15(cat), 16(night), 17(funeral, death), 18(color, blue), 19(grief), 20(ink), 21(bottom), 22(table, seat, sit, couch, desk, recliner, sofa, cushion, stool, sitting, bench), 23(legs), 24(wood)

Tabela 4.2: Reconhecimento de algumas listas treinadas pelo Módulo de Contexto (ρ = 0.989). Distraidores críticos encontram-se em itálico.

Palavra	Agrupamento	Cosseno	Palavra	Agrupamento	Cosseno
Foot	1	0.9851	ankle	1	0.8514
hand	2	0.9480	boot	4	0.9550
kick	3	0.9363	sock	1	0.9812
soccer	3	0.9054	mouth	8	1.0000
Fruit	9	0.9879	pear	9	0.9824
banana	9	0.9664	berry	9	0.9956
apple	9	0.9909	basket	3	0.8778
orange	9	0.9890	salad	10	0.9345
citrus	9	0.9633	cocktail	12	1.0000
Black	13	0.9934	color	18	0.9806
white	13	0.9820	blue	18	0.9565
cat	15	1.0000	ink	20	1.0000
night	16	1.0000	coal	13	0.9150
Chair	22	0.9841	wood	24	1.0000
table	22	0.9797	desk	22	0.9512
legs	23	1.0000	stool	22	0.9810
couch	22	0.9639	bench	22	0.9920

Os resultados de reconhecimento obtidos sugerem que a re-apresentação de palavras treinadas geralmente implica na vitória do protótipo em que foram agrupadas, uma vez que, na simulação, nenhuma das palavras das listas estudadas recuperou um outro protótipo. Além disso, a redução do parâmetro de vigilância ocorre principalmente durante o reconhecimento de itens não estudados, uma vez que não há bons representantes destes dentre os protótipos. Uma exceção foi a palavra *wine*, que, apesar de não treinada, foi associada com uma alta similaridade ao agrupamento de *cocktail*, o que indica que esta palavra talvez não corresponda como deveria ao tema *bread*.

É notável, também, que todos os distraidores críticos tenham sido associados com facilidade a protótipos de palavras de suas listas, o que é esperado.

Tabela 4.3: Reconhecimento de itens não-treinados pelo Módulo de Contexto. Distraidor crítico em itálico.

Palavra	Grupo	Cos	ρ	Palavra	Grupo	Cos	ρ
Bread	3	0.0232	0.9733	flour	13	0.2984	0.9792
butter	23	0.0225	0.9723	jelly	11	0.0799	0.9743
food	10	0.2589	0.9782	dough	23	0.0246	0.9733
eat	11	0.2183	0.9772	crust	10	0.2345	0.9782
sandwich	10	0.3302	0.9801	slice	10	0.3119	0.9801
rye	3	0.0306	0.9733	wine	12	0.8518	**0.9890**
jam	11	0.0718	0.9743	loaf	3	0.0228	0.9723
milk	11	0.0830	0.9743	toast	24	0.0324	0.9733

4.2 Validação do Módulo de Essência

O módulo de essência foi validado através de três experimentos: avaliação da formação regular de protótipos, análise do efeito causado pela apresentação das palavras das listas em ordem aleatória, e avaliação do grau de certeza calculado sobre a resposta.

4.2.1 Avaliação da Formação de Protótipo

A Tabela 4.4 apresenta o resultado da formação de protótipos decorrente do treinamento com 12 listas específicas. Essas listas foram escolhidas aleatoriamente dentre as 24 do paradigma DRM.

Pode-se observar que o módulo de essência tende a agrupar palavras de uma mesma lista, o que é desejável, uma vez que a maioria das listas apresentam um forte significado semântico em comum. Para facilitar esse agrupamento, o módulo recebe as

sáidas contextuais do módulo de contexto. Como se espera que, nos experimentos de falsas memórias, as palavras sejam estudadas por lista, o contexto em comum ajuda a aproximar itens instanciados sob um mesmo tema, mas distantes entre si quando considerada a representação de essência apenas.

Tabela 4.4: Protótipos formados numa execução do Módulo de Essência. Apresenta-se o número de palavras por lista que formam cada protótipo entre colchetes. Os protótipos com palavras de mais de uma lista encontram-se entre parênteses. 12 listas foram treinadas.

SLOW[1]	SLOW[6]
DOCTOR[14]	ANGER[9]
(FOOT[13], DOCTOR[1])	BREAD[15]
(FOOT[2], SOFT[1])	(ANGER[3], BLACK[1])
MUSIC[1]	(SLOW[4], ROUGH[1])
BLACK[11]	MUSIC[14]
(SOFT[2], BLACK[1])	(MOUNTAIN[2], COLD[1])
MOUNTAIN[10]	MOUNTAIN[1]
(SOFT[11], ROUGH[3], BLACK[1])	(MOUNTAIN[1], COLD[7])
COLD[4]	ROUGH[2]
COLD[1]	COLD[1]
RIVER[1]	RIVER[1]
RIVER[1]	RIVER[9]
(ROUGH[4], RIVER[1]).	

4.2.2 Avaliação de Contexto

Através da apresentação aleatória das palavras, perde-se a noção de contexto semântico das listas, sendo interessante analisar o efeitos dessa manipulação sobre a formação de protótipos de essência.

A Tabela 4.5 exibe os agrupamentos produzidos pelo módulo nessa situação. Percebe-se que bem menos protótipos foram formados em comparação com a operação regular da rede, o que significa que se perdeu informação discriminativa. A razão é simples: o vetor de contexto deixa de trazer informação relevante. Isso se deve ao fato de ele tornar-se equivalente a um valor médio, igualmente distante de todos os padrões, no espaço de representação semântica, mas ainda assim ser considerado pelo módulo de essência no cálculo da distância entre a entrada e os protótipos. Com uma informação a menos, todos os estímulos de entrada estarão mais capacitados a superarem facilmente o limiar de resposta da ativação do nodo vencedor, o que evita a criação de novos nodos e leva ao esquecimento de informações armazenadas de longa data.

Tabela 4.5: Formação de protótipos no Módulo de Essência após treinamento em ordem aleatória das palavras das listas. Apresenta-se o número de palavras por lista que formam cada protótipo entre colchetes. Os protótipos com palavras de mais de uma lista encontram-se entre parênteses. 12 listas foram treinadas.

(MOUNTAIN[11], RIVER[11], COLD[8], BLACK[1])
(COLD[6], MOUNTAIN[1])
(SLOW[11], DOCTOR[15], BREAD[13], ROUGH[1], FOOT[1])
(BLACK[12], MUSIC[15], SOFT[13], ROUGH[9], RIVER[1], ANGER[1], MOUNTAIN[2], BREAD[2])
(FOOT[14], ANGER[11], BLACK[1], RIVER[1], SOFT[1])

4.2.3 Avaliação dos Graus de Certeza

O último experimento de validação do módulo de essência apresentou um resultado intuitivo, como pode ser visualizado na Figura 4.2. Neste simulação, foram treinadas 12 listas selecionadas aleatoriamente e testadas no reconhecimento todas as 24 listas. Assim como esperado, o grau de certeza na resposta de associação do módulo, calculado através da probabilidade de Gibbs, assumiu valores bem maiores para as palavras treinadas (e seus distraidores) que para os padrões não-estudados.

Isso confirma a função do parâmetro de certeza, que, para distraidores não-relacionados, tornará mais difícil o reconhecimento por similaridade no módulo de decisão.

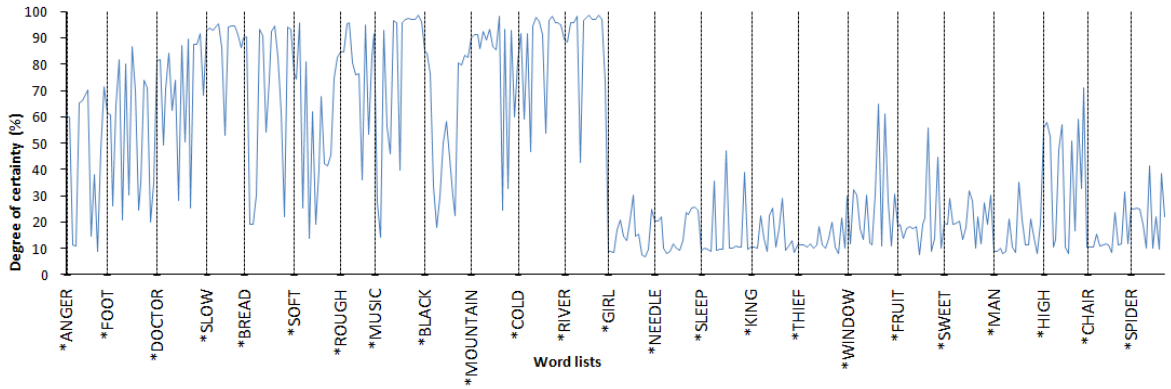

Figura 4.2: Gráfico do grau de certeza gerado pelo Módulo de Essência por palavra. As 12 primeiras listas foram treinadas, as 12 seguintes não. O grau de certeza foi significativamente mais alto para padrões treinados.

4.3 Validação do Módulo de Literalidade

As simulações com o módulo de literalidade foram realizadas para que três aspectos deste módulo pudessem ser compreendidos: a quantidade média de protótipos

formados durante o treinamento do módulo e a qualidade no armazenamento dos traços *verbatim* estudados; os efeitos causados pela variação no número de candidatos a vencedor; e as alterações causadas pela adição de ruído gaussiano aleatório durante o treinamento e reconhecimento.

4.3.1 Operação Regular

A operação regular do módulo, através do treinamento com 12 listas aleatórias e apresentadas de forma ordenada, resulta num número de agrupamentos médio em torno de 135 (Tabela 4.6), muitos dos quais são nodos-identidade, equivalentes aos padrões literais que os ativaram. A razão é que as listas são formadas por similaridade semântica, e não por semelhança fonética, tornando mais rara a compatibilidade literal entre palavras distintas. Se, por um lado, mais palavras serão reconhecidas com perfeição (de forma análoga à recordação de alvo), por outro, a interferência retroativa dificultará a extração de traços literais adequados em protótipos representativos de mais de um termo.

Tabela 4.6: Número de centros formados no Módulo de Literalidade. 12 listas foram treinadas.

Simulações:	1	2	3	4	5	6	7	Média	%
Total de centros	133	137	137	136	148	128	124	135	100.0%
1 palavra	110	117	116	117	135	104	100	114	84.4%
2 palavras	17	15	14	13	7	16	14	14	10.37%
3 palavras	6	3	5	5	4	4	8	5	3.7%
4 palavras	0	2	1	0	2	4	2	2	1.5%
5 palavras	0	0	1	1	0	0	0	1	0.7%
6 palavras	0	0	0	0	0	0	0	0	0.0%

Na Tabela 4.7, observa-se o resultado da aplicação do reconhecimento para 36 alvos aleatoriamente escolhidos, os 12 distraidores críticos relacionados, os 12 distraidores críticos não-relacionados e 12 palavras das listas não-estudadas. A média da distância literal para os alvos foi a única que se destacou como sendo claramente inferior às demais, o que demonstra o fato de a distância literal ser uma medida muito mais discriminativa de distraidores que a distância semântica.

4.3.2 Avaliação do Número de Candidatos a Vencedor

Neste experimento, o número de candidatos a vencedor selecionados através das informações fonéticas e contextuais foi variado entre 1 e 15. Os resultados demonstraram

Tabela 4.7: Distância literal média (7 execuções) entre o protótipo vencedor e a representação literal de entrada no Módulo de Literalidade para cada tipo de palavra.

Tipo de teste	Distância Literal
Alvos de listas treinadas	0.0444
Distraidores críticos de listas treinadas	0.6653
Distraidores críticos de listas não-treinadas	0.6068
Distraidores de listas não-treinadas	0.6616

que podem ser completamente diferentes caso consideremos um número igual ou diferente de candidatos nas fases de treinamento e de reconhecimento.

A Figura 4.3 exibe os efeitos causados pela variação em conjunto do número de candidatos durante a seleção do vencedor no treinamento e na classificação. Pode ser concluído que a redução na taxa de acerto do protótipo correto, e de sua escolha dentre os candidatos selecionados, decorre da formação de menos protótipos, o que leva ao esquecimento natural de algumas das informações armazenadas. Quanto menor o número de candidatos durante o treinamento, mais provável se torna a reprovação no limiar de distância literal e consequente ativação de um novo nodo. De certa forma, os dados de entrada acabam sendo "copiados" para a memória e seus protótipos sendo extraídos durante o reconhecimento, levando a uma maior taxa de acerto, como demonstra o gráfico.

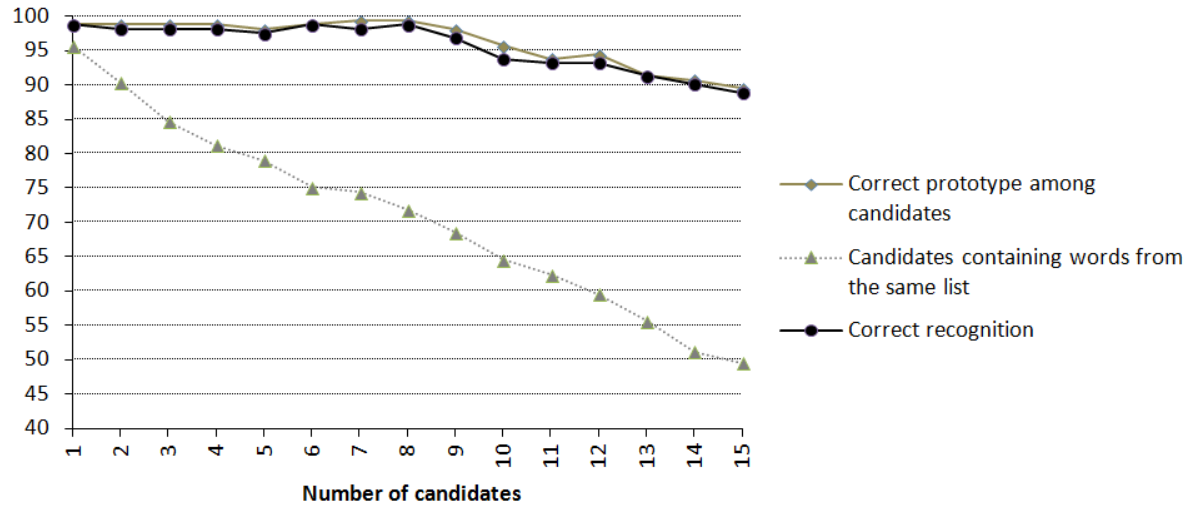

Figura 4.3: Impacto da variação do número de candidatos no Módulo de Literalidade. O número de candidatos durante o treinamento e o reconhecimento é o mesmo.

Já na Figura 4.4, temos que a manutenção de um número de candidatos fixo durante o treinamento resulta num crescimento da taxa de acerto na escolha do protótipo

vencedor. A razão é intuitiva e decorre do caso anterior: como o número de candidatos é fixo no treinamento, pode-se dizer que haverá pouca variação na aproximação das informações de treinamento pela rede, isto é, o número de agrupamentos formados será quase o mesmo entre duas iterações quaiquer do algoritmo. Logo, é fácil perceber que o aumento no número de candidatos considerado aumentará as chances de se escolher o centro correto.

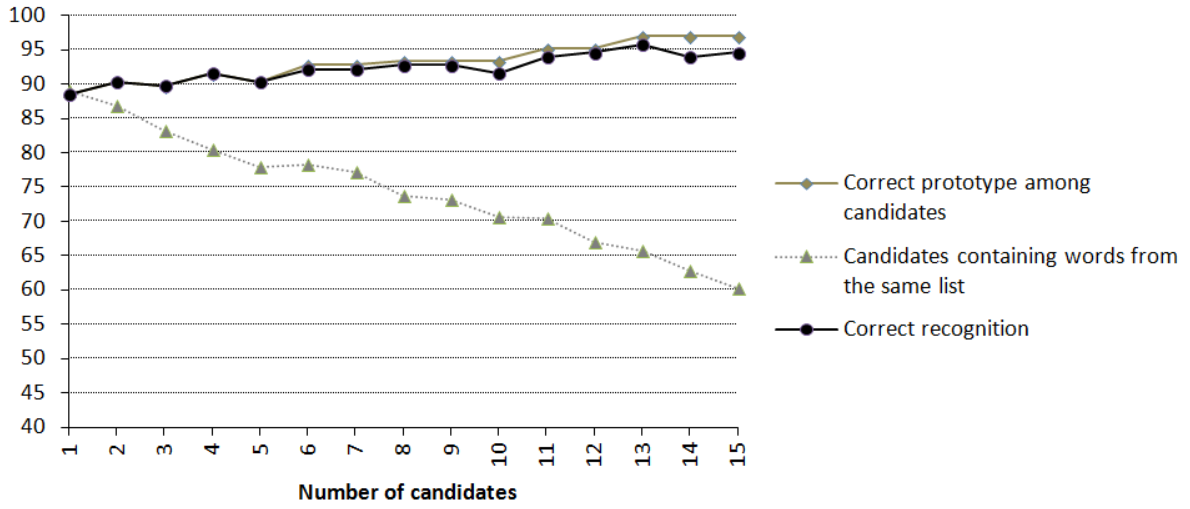

Figura 4.4: Impacto da variação do número de candidatos no Módulo de Literalidade. O número de candidatos durante o treinamento é 7, sendo variado apenas durante o reconhecimento.

4.3.3 Avaliação da Presença de Ruído

Os gráficos das Figuras 4.5 e 4.6 representam, respectivamente, as situações em que a operação do módulo de literalidade sofre decaimento por adição de ruído gaussiano ou não. Como o ruído é aplicado a cada apresentação de padrão durante o treinamento ou classificação, o gráfico da distância literal entre o protótipo do nodo vencedor e a entrada possui um formato de decaimento linear, deixando claro que centros formados a mais tempo tendem a apresentar um deslocamento maior em relação aos padrões que representa no espaço de dados fonéticos.

4.4 Reprodução dos Experimentos com Falsas Memórias

Nesta seção, os experimentos realizados por Brainerd e Reyna (1998b) baseados no paradigma DRM serão reproduzidos. Desta forma, obteremos parâmetros para a

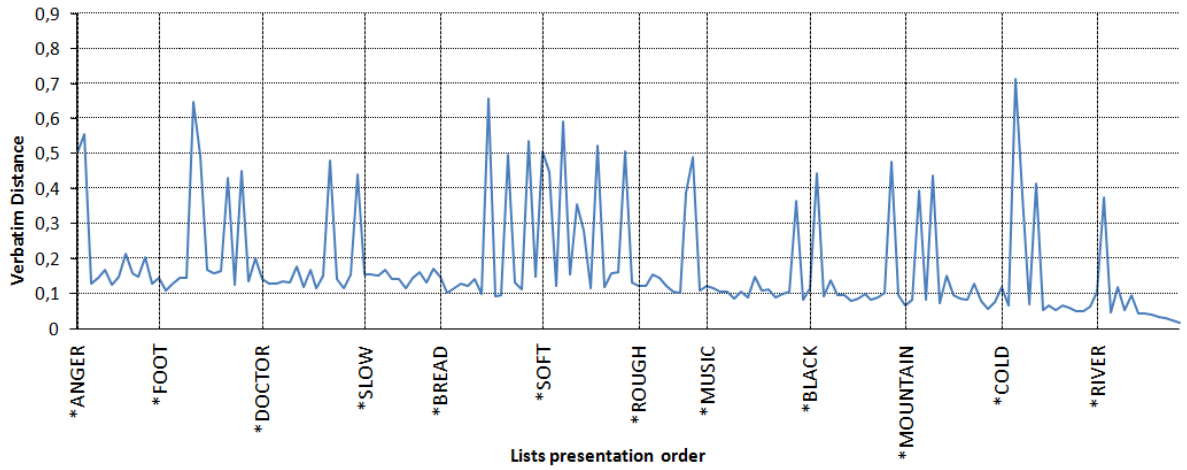

Figura 4.5: Evolução da distância verbatim entre a entrada e o centro vencedor no Módulo de Literalidade, aplicando-se ruído gaussiano aos centros ativos.

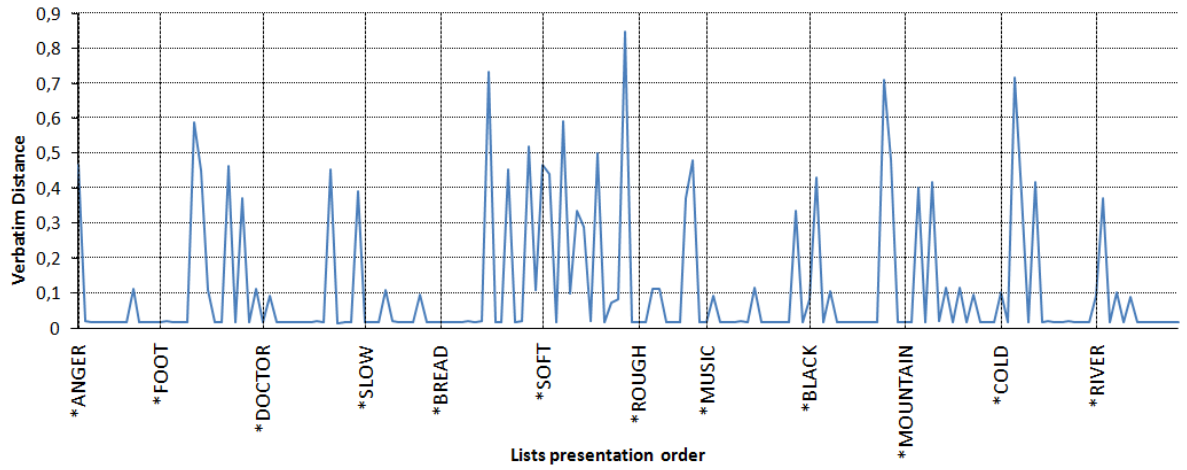

Figura 4.6: Evolução da distância verbatim entre a entrada e o centro vencedor no Módulo de Literalidade, sem ruído.

análise da eficácia do sistema implementado na modelagem de falsas memórias. O sistema será comparado com a sua primeira versão, proposta por Pacheco (2004).

4.4.1 Simulação 1 - Reconhecimento Regular

Como descrito no Capítulo 2, em Brainerd e Reyna (1998b), as listas DRM foram estudadas e reconhecidas por indivíduos sob duas condições de instrução distintas: reconhecer somente alvos estudados (Grupo 1) e reconhecer tanto alvos como distraidores relacionados (Grupo 2). A primeira simulação faz referência ao primeiro grupo.

Na Tabela 4.8, estão dispostas as taxas de reconhecimento de alvos (A), distraidores críticos relacionados (DCR), distraidores críticos não-relacionados (DCNR) e distraidores não-relacionados (DNR). Os resultados estão agrupados por forma de obtenção: através de experimentos com pessoas, através da reprodução desses experimentos com o sistema implementado por Pacheco (2004) (chamado de **NS**, acrônimo para *Neural System*) e através da reprodução dos mesmos com a versão modificada do NS, proposta neste trabalho, referenciada como **newNS**.

Tabela 4.8: Comparação do desempenho humano, do sistema proposto por Pacheco (2004) e do sistema adaptado neste trabalho no **primeiro** experimento do paradigma DRM. Para cada tipo de teste, é mostrada a taxa de reconhecimento e o desvio padrão. Os valores resultam de 30 execuções do sistema (treinamento + teste), nas quais se escolheu aleatoriamente 12 listas de estudo e 12 listas apenas de teste.

Tipos de teste	Reconhecimento %					STD %		
	Homem	NS	newNS	ΔNS	ΔnewNS	Homem	NS	newNS
A	61	63	62.78	2	1.78	8	6.8	8.07
DCR	63	59	65	4	2	13	8.1	13.68
DCNR	19	12	5.56	7	13.44	11	5.2	6.21
DNR	16	7	5.83	9	10.17	8	3.0	6.51

Pode-se concluir que as mudanças implementadas no sistema proposto melhoraram a eficácia na reprodução do comportamento do Grupo 1. As taxas de reconhecimento mais importantes, relativas aos alvos A e aos distraidores críticos relacionados DCR, foram melhor aproximadas pelo novo sistema, reduzindo o erro no acerto de A de 2% para 1.78% e o erro no falso reconhecimento de DCR de 4% para 2%. Os desvios padrões tornaram-se ainda mais próximos, sendo o relativo ao reconhecimento de alvos estudados praticamente idêntico ao gerado por humanos.

Além disso, as relações de magnitude entre os dados foram mantidas, com a taxa de reconhecimento de distraidores críticos relacionados sendo superior à de alvos, mesmo que por uma margem pequena. A antiga versão do sistema não chegou a alcançar este resultado, importante para a validação qualitativa do modelo.

As informações acerca dos distraidores não-relacionados DCNR e DNR não são tão relevantes quanto as mencionadas anteriormente devido ao fato de, na Teoria do Rastro Difuso, o reconhecimento de distraidores não-relacionados ser aproximado pelo valor de *bias*. Isto significa que a FTT atribui o reconhecimento desses tipos de testes a um processo completamente arbitrário, que resulta da falta de manifestação de quaisquer fenomenologias (i.e. recordação, familiaridade e recordação fantasma).

Nesse sentido, apesar do sistema proposto ter gerado um erro maior em relação

a essas duas taxas de falso reconhecimento, isso é um bom indício de que ele está funcionando de acordo com nossas expectativas. O newNS nunca apresentou a intenção de modelar a existência de uma tendência subjetiva (*bias*) à aceitação, logo, é esperado que as taxas relativas aos itens DCNR e DNR aproximem-se de zero.

Tabela 4.9: Probabilidade dos julgamentos por tipo de teste no **primeiro** experimento DRM. Considerando 30 execuções do sistema (treinamento + teste), com listas estudadas escolhidas aleatoriamente.

Tipo de teste	Identidade (%)	Similaridade (%)
A	52.59	10.19
DCR	3.61	61.39
DCNR	3.89	1.67
DNR	2.22	3.61

Na Tabela 4.9, vemos que a distribuição de probabilidade entre os possíveis julgamentos da rede apresentou o comportamento desejado, com os alvos sendo reconhecidos principalmente através de identidade (52.59%) e os distraidores críticos relacionados sendo reconhecidos majoritariamente por similaridade (61.39%). Isso diz respeito aos traços armazenados em cada um dos dois casos durante o estudo: enquanto os alvos podem induzir à recoleta de seus traços literais ou à extração de suas características semânticas, os distraidores relacionados devem recorrer aos traços *gist* dos alvos relacionados com mais frequência para serem reconhecidos.

4.4.2 Simulação 2 - Reconhecimento de Significado

Para reprodução dos experimentos do Grupo 2, a instrução de aceitação de quaisquer palavras que mantenham a essência das listas estudadas foi modelada através da atribuição de um valor unitário constante ao grau de certeza do módulo de essência, isto é, fazendo-se $DC^g = 1$. Os resultados da simulação podem ser vistos na Tabela 4.10.

Nesta simulação, o desempenho do sistema implementado newNS cresceu ainda mais em relação ao modelo original, chegando a reduzir o erro na reprodução da taxa de falso reconhecimento de distraidores críticos em um pouco mais de 7%. Além disso, o modelo revisto mostrou-se capaz de manter as relações de magnitude entre os dados, o que, assim como no experimento anterior, contribui para a plausibilidade do mesmo.

Em relação à distribuição dos tipos de julgamento modelados, podemos ver, na Tabela 4.11, que os padrões estudados costumam extrair traços literais de memória,

Tabela 4.10: Comparação do desempenho humano, do sistema proposto por Pacheco (2004) e do sistema adaptado neste trabalho no **segundo** experimento do paradigma DRM. Para cada tipo de teste, é mostrada a taxa de reconhecimento e o desvio padrão. Os valores resultam de 30 execuções do sistema (treinamento + teste), nas quais se escolheu aleatoriamente 12 listas de estudo e 12 listas apenas de teste.

Tipos de teste	Reconhecimento %					STD %		
	Homem	NS	newNS	ΔNS	ΔnewNS	Homem	NS	newNS
A	68	65	66.67	3	1.33	9	6.3	9.16
DCR	88	77	84.44	11	3.56	8	7.1	11.12
DCNR	28	11	13.06	17	14.94	15	5.6	9.79
DNR	25	7	10	18	15	14	3.0	8.16

enquanto os distraidores críticos levam à recuperação de traços *gist*, assim como esperado.

Tabela 4.11: Probabilidade dos julgamentos por tipo de teste no **segundo** experimento DRM. Considerando 30 execuções do sistema (treinamento + teste), com listas estudadas escolhidas aleatoriamente.

Tipo de teste	Identidade (%)	Similaridade (%)
A	54.07	12.59
DCR	2.22	82.22
DCNR	4.44	8.61
DNR	3.06	6.94

4.5 Considerações Finais

Como analisado na seção anterior, o modelo neural proposto neste trabalho adequou-se muito bem aos dados dos experimentos de falsas memórias reproduzidos. Houve melhora sobre todos os aspectos dos resultados, em relação ao modelo inicial proposto por Pacheco (2004). Até mesmo quando o modelo distanciou-se da taxa de reconhecimento dos distraidores não-relacionados apresentada pelas pessoas, seu embasamento na Teoria do Rastro Difuso foi suficiente para desmistificar este resultado.

A evolução do modelo pode ser compreendida em face de todas as revisões e mudanças realizadas no sistema modular originalmente proposto. Alguns exemplos de adaptações que o NS sofreu ao longo deste trabalho são: o uso de uma nova representação da essência do estímulo, com embasamento teórico e experimental; a aplicação das três saídas contextuais do módulo de contexto no cálculo da função de

custo do módulo de essência; a redefinição do algoritmo do módulo de literalidade, aplicando-se a seleção de candidatos também durante o treinamento; a adição de uma nova entrada a este módulo, o contexto atual; e a introdução de um ruído aleatório gaussiano para modelagem do decaimento de memória.

A descrição de todas estas modificações podem ser encontradas na apresentação do sistema proposto, ao longo do Capítulo 3.

5 Conclusão

Neste trabalho, levamos em consideração características psicológicas e neurofisiológicas relacionadas ao fenômeno de falsificação de memórias para ajustar e implementar uma arquitetura, originalmente proposta no trabalho de Pacheco (2004), capaz de reproduzir funções cognitivas relacionadas à memória e suas estruturas através de um conjunto de módulos neurais.

5.1 Objetivos Alcançados

Como visto no Capítulo 4, o modelo foi capaz de simular a ilusão de memória e os padrões observados em laboratório no reconhecimento de diferentes tipos de palavras por pessoas, em duas condições de instrução diferentes. Além disso, apresentou resultados significativamente melhores que os obtidos pela primeira versão do sistema, especialmente quando instruído para aceitar alvos e distraidores relacionados.

Dentre as premissas que deveriam ser incorporadas ao modelo, na Seção 2.1.6, de acordo com a Teoria do Rastro Difuso, percebe-se que o modelo proposto: a) armazena informações em paralelo nas memórias literal e de essência, através dos módulos de literalidade e de essência, respectivamente; b) recupera informações em paralelo dos módulos de literalidade e de essência; c) através do ruído aditivo gaussiano, resulta num decaimento mais rápido da memória literal; d) e costuma reconhecer os alvos por identidade e os distraidores relacionados por similaridade, como comprovado nas Seções 4.4.1 e 4.4.2.

Quanto aos possíveis julgamentos previstos pela FTT, o modelo proposto: e) aceita corretamente alvos cujos traços foram recuperados da memória literal (identidade); f) rejeita corretamente distraidores relacionados que recuperaram traços literais de um item contextual ou semanticamente relacionado (não-identidade); g) aceita corre-

tamente alvos cujos traços foram recuperados da memória de essência, assim como aceita incorretamente distraidores semelhantes a alvos cujos traços de essência foram extraídos (similaridade); h) e rejeita incorretamente alvos que recuperaram traços literais associados a um contexto semelhante, mas diferentes da sua codificação (recordação errônea), o que pode resultar da interferência retroativa presente na formação de agrupamentos com mais de uma palavra ou do decaimento devido ao ruído aditivo gaussiano.

Dentre as premissas neurofisiológicas que deveriam ser incorporadas ao modelo, citadas na Seção 2.2.3, percebe-se que o modelo proposto: a) monta uma representação contextual da informação recebida, de forma análoga ao córtex parahipocampal; b) atribui relevância em termos de novidade à informação, através do limiar dinâmico implementado no módulo de essência; c) armazena as representações literal e de essência na "memória" do sistema, os módulos de essência e de literalidade; d) e associa informações contextuais ao estímulo de entrada, no módulo de contexto. Todas as premissas neurofisiológicas, portanto, foram incorporadas no sistema.

Pode-se concluir que os conceitos psicológicos e neurofisiológicos que basearam a modelagem do sistema, tendo em vista a qualidade com que o mesmo reproduziu o fenômeno de falsas memórias, são robustos o suficiente para explicar os efeitos gerados nos experimentos do paradigma DRM aqui estudados.

No entanto, a abordagem de reconhecimento por recordação fantasma, extensão da Teoria do Rastro Difuso, ainda necessita de ser implementada e testada através do nosso modelo. Isso requer, futuramente, um estudo cuidadoso, uma vez que envolve a incorporação da aprendizagem de traços *verbatim* ilusórios, isto é, palavras que não foram estudadas devem, de alguma forma, estar presentes na memória literal.

5.2 Contribuições do Modelo Proposto

Pode-se dizer que, dentre as principais contribuições deste trabalho, estão inclusas:

- a análise do estado da arte na psicologia e neurofisiologia no que diz respeito aos processos de memorização e reconhecimento;
- a revisão e adequação do modelo neural proposto por Pacheco (2004) a estes conceitos;
- o uso de uma representação de essência plausível, baseada em dados de normas associativas coletados durante uma década de experimentos com indivíduos;

- a inclusão de influência do contexto semântico atual no armazenamento e recuperação de traços literais da memória *verbatim*;
- o uso de um ruído de distribuição gaussiana de média 0 para induzir a um decaimento da memória *verbatim* ao longo dos processos de memorização e reconhecimento;
- a modificação do algoritmo de treinamento do módulo de literalidade, adicionando-se uma seleção de candidatos que levará à escolha do protótipo vencedor baseada na distância literal, mas direcionada pela distância contextual;
- e a capacidade de aproximar de forma mais fidedigma os resultados de dois experimentos com humanos sob o paradigma DRM, executados sob condições de instrução distintas.

Por se adequar a certos padrões comportamentais observados em pessoas, o modelo proposto representa os primórdios de uma oportunidade estimulante: o uso de um sistema computacional para validação de hipóteses da psicologia e neurofisiologia acerca do funcionamento da memória humana. O aumento de sua robustez, através da reprodução de testes inspirados em outros paradigmas experimentais, possibilitará que, futuramente, este modelo venha a ser utilizado para simular e prever ações de memorização e reconhecimento, contribuindo ao estudo de doenças e lesões cerebrais.

5.3 Trabalhos Futuros

Este trabalho abre margens para uma série de oportunidades de estudo e projeto. Seria interessante, por exemplo, o desenvolvimento de um modelo matemático para cálculo direto da taxa com que cada um dos quatro processos de julgamento implementados pelo sistema ocorrem durante seu funcionamento (identidade, não-identidade, similaridade e rejeição por recordação). As taxas encontradas poderiam então ser validadas através do modelo estatístico *conjoint recognition* (Brainerd et al., 2001), ajudando na identificação e superação de possíveis limitações do sistema.

Como já foi mencionado na Seção 5.1, outro ponto que deve ser considerado é a inclusão da fenomenologia de recordação fantasma, provavelmente através da adição de um novo módulo à arquitetura do sistema.

Estudos como o de Stadler et al. (1999) demonstram uma alta variação no nível interno de coesão semântica entre diferentes listas do paradigma DRM. Nesse sentido, um trabalho futuro adicional consistiria de testar a performance do atual modelo

considerando-se o treinamento de listas específicas, sendo verificado se o comportamento gerado nestas situações permanece fidedigno ao manifestado por seres humanos.

Por fim, para que o modelo proposto possa evoluir e se tornar mais robusto, é importante também que ele seja aplicado a outros paradigmas de experimentos, além do DRM. Seria interessante, inclusive, que estes envolvessem objetos de estudo que não fossem apenas palavras, a exemplo da memorização de imagens ou sentenças, ampliando ainda mais o escopo do projeto.

Referências Bibliográficas

Aggleton, J. P., Brown, M. W., 1999. Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. Behavioral and Brain Sciences 22 (3), 425–44.

Araújo, A. F. R., Henriques, A. S., 2002. A stochastic neural model for fast identification of spatiotemporal sequences. Neural Processing Letters 16 (03), 277 – 292.

Atkinson, R. C., Juola, J. F., 1973. Factors influencing speed and accuracy of word recognition. Attention and performance IV, 583–612.

Atkinson, R. C., Juola, J. F., 1974. Search and decision processes in recognition memory. WH Freeman.

Brainerd, C., Poole, D., 1997. Long-term survival of children's false memories: A review. Learning and Individual Differences 9 (2), 125 – 151.

Brainerd, C., Reyna, V., Wright, R., Mojardin, A., 2003. Recollection rejection: False-memory editing in children and adults. Psychological Review 110 (4), 762–784.

Brainerd, C. J., Reyna, V. F., 1990. Gist is the grist: Fuzzy-trace theory and the new intuitionism. Developmental Review 10 (1), 3–47.

Brainerd, C. J., Reyna, V. F., 1998a. Fuzzy-trace theory and children's false memories. Journal of Experimental Child Psychology 71, 81–129.

Brainerd, C. J., Reyna, V. F., 1998b. When things that were never experienced are easier to"remember"than things that were. Psychological Science 9 (6), 484–489.

Brainerd, C. J., Reyna, V. F., Kneer, R., 1995. False-recognition reversal: When similarity is distinctive. Journal of Memory and Language 34 (2), 157 – 185.

Brainerd, C. J., Reyna, V. F., Mojardin, A. H., Jan 1999. Conjoint recognition. Psychol Rev 106 (1), 160–179.

Brainerd, C. J., Wright, R., May 2005. Forward association, backward association, and the false-memory illusion. J Exp Psychol Learn Mem Cogn 31 (3), 554–567.

Brainerd, C. J., Wright, R., Reyna, V. F., Mojardin, A. H., Mar 2001. Conjoint recognition and phantom recollection. J Exp Psychol Learn Mem Cogn 27 (2), 307–327.

Burgess, P. W., Shallice, T., 1996. Confabulation and the control of recollection. Memory 4 (4), 359–411.

Cahill, L., Haier, R. J., Fallon, J., Alkire, M. T., Tang, C., Keator, D., Wu, J., McGaugh, J. L., 1996. Amygdala activity at encoding correlated with long-term, free recall of emotional information. Proceedings of the National Academy of Sciences, U.S.A. 93 (15), 8016–8021.

Carpenter, G., Grossberg, S., 1987. Art2: Self-organization of stable category recognition codes for analog input patterns. Applied Optics 26, 4919 – 4930,.

Cherry, C., 1974. A Comunicação Humana: uma recapitulação, uma vista de conjunto e uma crítica. Cultrix - Editora da USP.

Clark, S. E., Gronlund, S. D., 1996. Global matching models of recognition memory: How the models match the data. Psychonomic Bulletin & Review 3 (1), 37–60.

CMU, 2011. The CMU pronouncing dictionary - a machine-readable pronunciation dictionary for north american english. On-line.

Deese, J., 1959. On the prediction of occurrence of particular verbal intrusions in immediate recall. Journal of Experimental Psychology 58 (1), 17–22.

Diana, R. A., Yonelinas, A. P., Ranganath, C., et al., 2007. Imaging recollection and familiarity in the medial temporal lobe: a three-component model. Trends in cognitive sciences 11 (9), 379–386.

Drever, J., 1964. A Dictionary of Psychology. Penguin.

Fletcher, P. C., Frith, C. D., Rugg, M. D., 1997. The functional neuroanatomy of episodic memory. Trends in Neurosciences 20 (5), 213–218.

Guyton, A. C., 1993. Neurociência Básica - Anatomia e Fisiologia.

Haykin, S., 1999. Neural Networks – A Comprehensive Foundation. Prentice Hall.

Henson, R. N., Rugg, M. D., Shallice, T., Josephs, O., Dolan, R. J., 1999. Recollection and familiarity in recognition memory: an event-related functional magnetic resonance imaging study. Journal of Neuroscience 19 (10), 3962–3972.

Hinton, G. E., Shallice, T., 1991. Lesioning an attractor network: investigations of acquired dyslexia. Psychological Review 98 (1), 74–95.

Israel, L., Schacter, D. L., 1997. Pictorial encoding reduces false recognition of semantic associates. Psychonomic Bulletin & Review 4 (4), 577–581.

Jacoby, L. L., 1996. Dissociating automatic and consciously controlled effects of study/test compatibility. Journal of Memory and Language 35 (1), 32–52.

Jakobson, R., Fant, G., Hall, M., 1952. Preliminaries to Speech Analysis. M.I.T., Acoust. Lab. Report, 13.

Knight, R. T., 1996. Contribution of human hippocampal region to novelty detection. Nature 383 (6597), 256–259.

Kohonen, T., 1989. Self Organization and Associative Memory, 3Âº Edição. Berlin: Springer-Verlag.

Kohonen, T., Makisara, K., Simula, O., Kangas, J., 1991. The hypermap architecture. in: Artificial neural networks, t. kohonen, k. makisara, o. simula, and j. kangas. Artificial Neural Networks 1, 1357–1360.

Mishkin, M., Appenzeller, T., 1987. The anatomy of memory. Scientific American 256 (6), 80–89.

Mishkin, M., Spiegler, B., Saunders, R., Malamut, B., 1982. An animal model of global amnesia. In Alzheimer's Disease: A Review of Progress,. Raven Press, 235–247.

Nelson, D. L., M. C. L., Schreiber, T. A., 1998. The University of South Florida word association, rhyme, and word fragment norms. On-line.

Pacheco, R. F., 2004. Módulos neurais para a modelagem de falsas memórias. Tese de Doutorado, Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2004.

Payne, D. G., Elie, C. J., Blackwell, J. M., Neuschatz, J. S., 1996. Memory illusions: Recalling, recognizing, and recollecting events that never occurred. Journal of Memory and Language 35 (2), 261–285.

Piaget, J., 1968. On the development of memory and identity. Vol. 14. Worcester, Mass., Clark University Press.

Proctor, 1995. Cambridge International Dictionary of English. Cambridge University Press, New York.

Reyna, V. F., Brainerd, C. J., 1995. Fuzzy-trace theory: Some foundational issues. Learning and Individual Differences 7 (2), 145 – 162.

Reyna, V. F., Lloyd, F., 1997. Theories of false memory in children and adults. Learning and Individual Differences 9 (2), 95–123.

Roediger, H., McDermott, K., 1995. Creating false memories: Remembering words not presented in lists. Journal of Experimental Psychology: Learning, Memory & Cognition 21, 803–814.

Rotello, C. M., Heit, E., 1999. Two-process models of recognition memory: Evidence for recall-to-reject? Journal of Memory and Language 40 (3), 432–453.

Schacter, D. L., 1996. Illusory memories: a cognitive neuroscience analysis. Proceedings of the National Academy of Sciences, U.S.A. 93 (24), 13527–13533.

Schacter, D. L., Norman, K. A., Koutstaal, W., 1998. The cognitive neuroscience of constructive memory. Annual Review of Psychology 49, 289–318.

Shallice, T., Burgess, P., 1996. The domain of supervisory processes and temporal organization of behaviour. Philosophical Transactions of the Royal Society B: Biological Sciences 351 (1346), 1405–11; discussion 1411–2.

Stadler, M. A., Roediger, H. L., McDermott, K. B., May 1999. Norms for word lists that create false memories. Mem Cognit 27 (3), 494–500.

Steyvers, M., Shiffrin, R. M., Nelson, D. L., 2004. Word association spaces for predicting semantic similarity effects in episodic memory. Experimental cognitive psychology and its applications: Festschrift in honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer, 237–249.

Taylor, J. G., Horwitz, B., Shah, N. J., Fellenz, W. A., Mueller-Gaertner, H. W., Krause, J. B., 2000. Decomposing memory: functional assignments and brain traffic in paired word associate learning. Neural Networks 13 (8-9), 923–940.

Wilson, F. A., Rolls, E. T., 1993. The effects of stimulus novelty and familiarity on neuronal activity in the amygdala of monkeys performing recognition memory tasks. Experimental Brain Research 93 (3), 367–382.