

UNIVERSIDADE FEDERAL DE GOIÁS
ESCOLA DE ENGENHARIA ELÉTRICA E DE
COMPUTAÇÃO
MESTRADO EM ENGENHARIA ELÉTRICA E DE
COMPUTAÇÃO

FABRICIA NERES BORGES

**Uso de Sistemas Tutores
Inteligentes na Compreensão de
Leitura**

Goiânia
2009

FABRICIA NERES BORGES

Uso de Sistemas Tutores Inteligentes na Compreensão de Leitura

Dissertação apresentada ao Programa de Pós-Graduação da Escola de Engenharia Elétrica e de Computação

Mestrado em Engenharia Elétrica e de Computação da Universidade Federal de Goiás, como requisito parcial para obtenção do título de Mestre no Programa de Mestrado em Engenharia Elétrica e de Computação.

Área de concentração: Engenharia de Computação

Orientador: Prof. Leonardo da Cunha Brito

Goiânia
2009

FABRICIA NERES BORGES

Uso de Sistemas Tutores Inteligentes na Compreensão de Leitura

Dissertação defendida no Programa de Pós-Graduação da Escola de Engenharia Elétrica e de Computação
Mestrado em Engenharia Elétrica e de Computação da Universidade Federal de Goiás como requisito parcial para obtenção do título de Mestre no Programa de Mestrado em Engenharia Elétrica e de Computação, aprovada em de Março de 2009, pela Banca Examinadora constituída pelos professores:

Prof. Leonardo da Cunha Brito

Escola de Engenharia Elétrica e de Computação
Mestrado em Engenharia Elétrica e de Computação – UFG
Presidente da Banca

Prof. Weber Martins

Escola de Engenharia Elétrica e de Computação – UFG

Prof. Solange da Silva

Departamento de Computação – PUC-Goiás

Dedico aos meus pais e ao meu padrinho Manuca Neres Brito

Agradecimentos

A Deus e aos espíritos superiores que me ajudaram nesta caminhada...

Ao Professor PhD. Weber Martins pela atenção e dedicação durante o trabalho.

Ao Professor. Dr. Rodrigo Lemos por sua atenção durante todo o meu mestrado.

Aos meus pais e ao meu padrinho Manuca Neres Brito.

Aos meus irmãos Douglas e Fábio que me apoiaram durante essa caminhada.

A todos os professores que me auxiliaram durante o curso, um grande obrigada.

A todos meus amigos que muito incentivaram. Agradeço especialmente minha amiga Livia que sempre me ouvia falar de informática e ainda tentava entender.

Aos novos amigos que eu fiz no mestrado. Em especial meu amigo Douglas que muito me auxiliou nos momentos que necessitei de sua presença e pretendo ter sua companhia por longos anos. Como não poderia faltar o meu amigo Geraldo, uma graça de pessoa. Agradeço a todos os outros que conheci no laboratórios Pireneus e Labsim.

Como não quero deixar ninguém de fora... um forte abraço para todos que torceram e acreditaram em mim, e que de uma forma ou outra contribuíram para o sucesso deste trabalho.

"Meus filhos terão computadores, sim, mas antes terão livros.
Sem livros, sem leitura, os nossos filhos serão incapazes de escrever
- inclusive a sua própria história."

Bill Gates,

.

Resumo

Borges, Fabricia Neres. **Uso de Sistemas Tutores Inteligentes na Compreensão de Leitura**. Goiânia, 2009. 89p. Dissertação de Mestrado. Escola de Engenharia Elétrica e de Computação
Mestrado em Engenharia Elétrica e de Computação, Universidade Federal de Goiás.

Os estudantes universitários brasileiros apresentaram baixos índices de leitura no Exame Nacional de Desempenho (ENADE) em 2006. O ENADE mostrou que o hábito de ler é pouco cultivado entre os universitários. O pouco interesse pela leitura é justificado pelo fato de que maioria dos estudantes são trabalhadores matriculados em cursos noturnos, com pouco tempo para se dedicar aos estudos. Esta pesquisa propõe o uso de sistemas tutores inteligentes para auxiliar universitários na compreensão de leitura. O objetivo do sistema tutor proposto é desenvolver a técnica de sublinhamento dos universitários para auxiliar na análise de textos acadêmicos. Dois grupos de estudantes, A e B, participaram da coleta de dados. A diferença entre os grupos reside na quantidade de exercícios realizados em cada grupo. O grupo A realizou 20 exercícios com quatro níveis de dificuldade. No grupo B a Rede Neural Artificial *Multilayer Perceptron* (MLP) decide a quantidade de exercícios que o estudante deve realizar em cada nível de dificuldade. A abordagem utilizada no grupo B adapta-se às características de retenção de conhecimento de cada estudante. Com isso, o sistema tutor se adapta ao grau de dificuldade ou facilidade do estudante. Por meio de estudo comparativo, a análise estatística dos dados indicou diferenças significativas entre os grupos A e B.

Palavras-chave

Sistemas Tutores Inteligentes, Leitura, Redes Neurais Artificiais, Sublinhamento

Abstract

Borges, Fabricia Neres. **Intelligent Tutoring Systems in Reading Comprehension**. Goiânia, 2009. 89p. MSc. Dissertation. Escola de Engenharia Elétrica e de Computação
Mestrado em Engenharia Elétrica e de Computação, Universidade Federal de Goiás.

Brazilian students have achieved poor results in the National Student Performance Exam (ENADE) in 2006. ENADE has shown reading is badly cultivated among undergraduates. The low interest on reading is justified by the fact that most of students have jobs and are enrolled in evening courses, without enough time to studies. The current research proposes the use of intelligent tutoring systems to improve student reading comprehension. The main goal is to develop the technique of underlining among undergraduates to assist in the analysis of academic texts. Two groups of students, A and B, participated in data collection. The difference between the groups is the amount of exercises performed in each group. Students of Group A have received 20 exercises with four levels of difficulty. In Group B, an Artificial Neural Network, Multilayer Perceptron (MLP), decides the amount of exercises that the student must perform at each level of difficulty by controlling what is the next exercise after each exercise is finished. The approach used in Group B adapts to the characteristics of knowledge retention of each student. Therefore, the tutoring system adapts the degree of exercise difficulty to the student. Statistical data analysis has indicated significant differences between groups A and B.

Keywords

Intelligent Tutoring Systems, Reading, Artificial Neural Network, Underlining

Sumário

Lista de Figuras	10
Lista de Tabelas	12
1 Introdução	13
1.1 Tema	13
1.2 Justificativa	13
1.3 Problema e Hipóteses	15
1.4 Visão Geral dos Capítulos	15
2 Fundamentação Teórica	17
2.1 Leitura	17
2.1.1 Compreensão de Leitura	18
2.1.2 Métodos para Avaliação da Compreensão de Leitura	21
2.1.3 Avaliação de Leitura com Adolescentes e Crianças	25
2.1.4 Avaliação de Leitura com Universitários	27
2.1.5 Leitura Proveitosa de Textos Acadêmicos	29
2.1.6 Índice de Legibilidade	30
2.2 Redes Neurais Artificiais	33
2.2.1 Breve Histórico	34
2.2.2 Neurônio Biológico	35
2.2.3 Neurônio Artificial	36
2.2.4 Arquiteturas das Redes Neurais	38
2.2.5 Processos de Aprendizagem	40
2.2.6 Perceptrons	41
2.2.7 Perceptrons de Múltiplas Camadas	42
2.3 Sistemas Tutores	44
2.3.1 Classificação dos Sistemas Tutores	45
2.3.2 Sistemas Tutores Computadorizados	47
2.3.3 Sistemas Tutores Inteligentes	48
3 Sistema Proposto	51
3.1 Estrutura do Sistema Tutor Inteligente	51
3.1.1 Ambiente Computacional	51
3.1.2 Tutoria	55

4	Experimentos e Resultados	58
4.1	Situação Experimental	58
4.1.1	Seleção das Variáveis	58
4.1.2	Sistema Tutor	60
4.1.3	Grupo de Controle	60
4.1.4	Treinamento da Rede Neural	61
4.1.5	Controle Inteligente	62
4.2	Análise dos Resultados	62
4.2.1	Análise Descritiva	63
4.2.2	Análise Inferencial	66
5	Conclusão	69
5.1	Considerações Gerais	69
5.2	Principais Contribuições	70
5.3	Trabalhos Futuros	70
	Referências Bibliográficas	71
A	Apêndice A	75
A.1	Sistema Tutor	75
B	Apêndice B	79
C	Apêndice C	84
D	Apêndice D	89

Lista de Figuras

2.1	Descrição do processo de compreensão de textos (Condemarín e Alliende 2005).	19
2.2	Ilustração do Teste de Cloze.	23
2.3	Neurônio biológico.	35
2.4	Neurônio artificial (Haykin 2001).	36
2.5	Função de limiar.	37
2.6	Função sigmóide.	38
2.7	Rede neural artificial com única camada.	39
2.8	Rede neural artificial com múltiplas camadas.	39
2.9	Rede neural artificial recorrentes (Haykin 2001).	40
2.10	Perceptron de única camada (Fausett 1994).	41
2.11	Perceptrons de múltiplas camadas.	42
2.12	Retropropagação (Haykin 2001).	43
2.13	Algoritmo de retropropagação.	44
2.14	Tutor clássico.	46
2.15	Tutor focado em atividades.	46
2.16	Tutor customizado.	46
2.17	Tutor de avanço por conhecimento.	47
2.18	Tutor exploratório.	47
2.19	Tutor gerador de lições.	47
2.20	Arquitetura tradicional dos STI.	49
3.1	Texto sublinhado de acordo com a técnica sugerida por (Marconi e Lakatos 2007).	52
3.2	Layout sistema tutor.	55
3.3	Etapas da tutoria.	56
3.4	Controle Inteligente dos Exercícios.	57
4.1	Composição do perfil.	59
4.2	Gráfico geral do treinamento.	62
4.3	Respostas sobre a quantidade de livros literários que estudante leu este ano de 2009.	63
4.4	Quantidade de exercícios na visão dos estudantes.	66
A.1	Tela de login do sistema tutor.	76
A.2	Fase 1: Uma frase.	76
A.3	Fase 2: Um parágrafo.	77
A.4	Fase 3: Dois parágrafos.	77
A.5	Fase 4: Quatro parágrafos.	78

A.6	Tela de encerramento do sistema tutor.	78
D.1	Primeira página: Termo de Consentimento Livre Esclarecido.	90
D.2	Segunda página: Termo de Consentimento Livre Esclarecido.	91
D.3	Terceira página: Termo de Consentimento Livre Esclarecido.	92

Lista de Tabelas

2.1	Índice de Flesch aplicado a textos brasileiros (Nunes et al. 1996).	31
2.2	Índice de Legibilidade (Squarisi e Salvador 2008)	32
4.1	Indicação do padrão de navegação.	59
4.2	Análise de resultados.	64
4.3	Análise descritiva dos dados coletados.	65
4.4	Comparativo das notas iniciais.	67
4.5	Comparativo das notas finais.	67
4.6	Comparativo dos ganhos normalizados.	68

Introdução

1.1 Tema

O hábito de ler do brasileiro é objeto de estudo de muitas pesquisas. Os estudantes do ensino fundamental, médio e superior apresentam dificuldades para realizar a análise de textos, ou seja, compreender a informação do texto. Esta pesquisa investiga a compreensão da leitura para adultos (alunos universitários). O sistema tutor inteligente proposto nesta pesquisa, procura auxiliar estudantes universitários na compreensão de leitura.

1.2 Justificativa

A leitura proporciona novos conhecimentos para o indivíduo e deve ser realizada com prazer. Em tempos de globalização, é preciso tornar-se bons leitores, realmente compreender o que é lido. Mais de 70% da população brasileira não lê revistas e nem jornais e o restante apresenta graus variados de compreensão de textos [Lima 2004].

Em 2000, um quarto da população com 15 anos eram analfabetos funcionais ¹ e apenas um adulto alfabetizado em cada três lê livros. O brasileiro lê, em média, 1,8 livros não acadêmicos por ano [Folha 2006]. Na Alemanha, esse índice é de 25 livros por habitante [Lima 2004].

O custo dos livros no Brasil desestimula a leitura, contudo a indiferença dos brasileiro pelos livros está ligado à história do país. Durante séculos de escravidão, nossos líderes descuidaram da Educação e o acesso às escolas estava restrito às elites. Na década de 1990, a escola primária tornou-se universal, enquanto o rádio e a televisão eram presença constante na vida do brasileiro [Folha 2006]. Com o uso de recursos de som e imagem, a televisão

¹ Pessoa que esteve na escola aprendeu a ler e a escrever, mas não é capaz de entender o que lê, e não consegue escrever um texto simples [Jack 2005]

estimula mais os jovens que os livros [Universia 2005]. Um exemplo do papel de nossos líderes pode ser visto na atitude do Presidente Luís Inácio Lula da Silva que, em entrevista à Revista Piauí [Congresso 2009], afirmou sua pouca afeição pela leitura. Lula declarou que não gosta de ler livros, jornais, sites de notícias ou blogs. Lula justificou que o contato com muitas pessoas durante o dia o mantém bem informado [Congresso 2009].

No cenário mundial, os estudantes brasileiros apresentam baixos índices de leitura, ciência e matemática no Programa Internacional de Avaliação de Alunos (PISA) [Goes e Pinho 2007]. Na edição de 2006, promovido pela Organização para a Cooperação e o Desenvolvimento Econômico (OCDE), o Brasil ocupou a 48^a em leitura (entre 56 países avaliados) e a 53^a em matemática (entre 57 países avaliados). Em leitura, 56 % dos jovens avaliados (15 anos) tiveram notas abaixo de um quando a nota máxima era cinco. Nossos representantes foram capazes de identificar apenas informações explícitas no texto e relações simples. Comparado com o PISA realizado em 2003, o desempenho em leitura de 2006 foi pior. O resultado do PISA confirma a necessidade de valorizar o papel da leitura na formação do brasileiro. Para solucionar o problema, não basta o governo comprar obras didáticas e aumentar a quantidade de bibliotecas públicas no país. É necessário incentivar a leitura em casa e na escola. Acredita-se que muito do prazer obtido na leitura está relacionado à compreensão do texto lido [Goes e Pinho 2007].

Durante a vida escolar, o jovem é forçado a ler textos didáticos ou livros alheios à sua realidade, gerando aversão à leitura. Segundo Perissé [Perissé 2002], as obras exigidas para realizar a prova de Literatura dos exames vestibulares exemplificam bem esse aspecto. O jovem gosta de ler histórias adolescentes (como o fenômeno Harry Potter, por exemplo), desconsiderando se o livro é apenas um volume de 500 páginas numa coleção de cinco ou mais livros. O jovem gosta de ler histórias envolventes, onde os personagens vivem situações diferentes e amores impossíveis com enredos atuais.

Oliveira [Oliveira 2007] registra que estudantes universitários participantes do Exame Nacional de Desempenho (ENADE) em 2006 apresentaram baixo índice de leitura [Oliveira 2007]. Segundo o ENADE, 43,6% dos alunos universitários estudam de uma a duas horas por semana além do horário de aula. Desses 43,6%, 34% lêem no máximo dois livros não didáticos por ano e 41,3% alegam que se informam mais pela televisão. Foi observado que a televisão está perdendo espaço para a Internet gradativamente [Universia 2005].

O desinteresse pela leitura é justificado pelo fato que a maioria dos

estudantes universitários brasileiros são trabalhadores matriculados em cursos noturnos, com pouco tempo para dedicar aos estudos. Do total, 68,2% dos universitários brasileiros estudam à noite e 73,2% trabalham durante o dia. Nesse contexto, é necessário e oportuno estabelecer políticas públicas para incentivar e facilitar a leitura entre os estudantes universitários [Universia 2005].

As universidades e escolas devem incentivar o prazer durante a leitura e preparar os alunos para além das provas, para a vida enfim. O hábito de ler pode auxiliar na expressão escrita e oral. O profissional afeito à leitura está geralmente informado sobre as novidades e tendências da sua área de interesse [Universia 2005].

A leitura abrange desde a codificação dos símbolos gráficos até a avaliação crítica do conteúdo. Para avaliar a leitura é necessário oferecer textos pequenos e completos de temas que despertem a atenção do leitor. Com crianças, é comum utilizar livros de histórias ilustradas, charadas ou desafios. Com adultos e adolescentes, pode-se utilizar reportagens de jornais e revistas [Ferreira 2004].

Neste trabalho é proposto a utilização de sistemas tutores inteligentes para auxiliar a compreensão de leitura de estudantes universitários. O sistema proposto procura desenvolver a habilidade de analisar textos, levando o leitor a diferenciar os elementos principais, secundários e a estrutura geral do texto.

1.3 Problema e Hipóteses

O **problema** abordado neste trabalho pode ser sintetizado na seguinte pergunta: Como desenvolver a habilidade de compreender textos acadêmicos?

Com base em estudos anteriores sobre compreensão de leitura de textos não-literários como o de Poersch e Rosa [Poersch e Rosa 2007], formulou-se a seguinte **hipótese principal**: Se a habilidade de sublinhamento é desenvolvida, então a capacidade de compreensão de textos acadêmicos é melhorada.

1.4 Visão Geral dos Capítulos

O Capítulo 2 apresenta a **Fundamentação Teórica** do presente trabalho sobre leitura abordando fatores influenciadores na compreensão de

leitura e a revisão bibliográfica sobre redes neurais artificiais e sistemas tutores inteligentes.

O Capítulo 3 dedica-se à descrição do **Sistema Proposto** para auxiliar na compreensão de leitura usando técnicas de inteligência artificiais.

O Capítulo 4 apresenta relato dos **Experimentos** realizados na pesquisa e **Resultados** alcançados durante a coleta de dados realizada com alunos universitários.

No Capítulo 5, a **Conclusão** do trabalho e as sugestões para trabalhos futuros.

Fundamentação Teórica

Este capítulo descreve a teoria de base do presente trabalho. A Seção 2.1 apresenta a teoria de base sobre compreensão de leitura, a Seção 2.2 trata-se das redes neurais artificiais e a Seção 2.3 sobre sistemas tutores inteligentes.

2.1 Leitura

Muitas pessoas acreditam que no futuro a leitura pode ser substituída pela imagem e o áudio aliados ao uso da tecnologia. Esta informação se fortalece porque muitas pessoas executam diversas atividades sem recorrer a leitura. Deixando a leitura apenas como necessidade no ensino escolar sem grande importância no cotidiano do indivíduo. A invasão súbita dos meios de comunicação de massa baseados em som e imagem (televisão, rádio e cinema) e o surgimento da Internet impõem uma mudança na situação e função da leitura no mundo contemporâneo. Nos países desenvolvidos os meios de comunicação de massa consolidaram a leitura, nos países menos desenvolvidos a realidade é outra, o hábito de ler foi afetado dentro e fora das escolas [Condemarín e Alliende 2005].

São numerosos os benefícios da leitura na vida das pessoas. A leitura expande o conhecimento, facilita a expressão de idéias oralmente e durante a produção de textos. O hábito de ler deve ser incentivado em casa, na escola e, até mesmo, nos meios de comunicação de massa [Condemarín e Alliende 2005].

A leitura é uma forma de entretenimento onde o indivíduo é livre para decidir uma série de parâmetros, dentre eles: lugar, tempo e a modalidade de leitura conveniente para o momento. Se o material é interessante, fácil ou conhecido, a leitura pode ser realizada de forma mais rápida, mas se existe o desconhecimento sobre o tema ou deseja realizar a leitura crítica o ideal é ler devagar [Condemarín e Alliende 2005].

O costume de ler enriquece e estimula o intelecto do estudante, repercutindo no processo de aprendizagem. A deficiência em leitura é apontado como fator complicador para a efetivação do processo de ensino-aprendizagem. Várias pesquisas comprovam a importância da compreensão da leitura para o bom desempenho escolar [Santos 2004].

O uso da linguagem escrita/impressa permitiu o registro e a recuperação da informação de forma eficiente, tornando o discurso autônomo e livrando de seu contexto de vivência. A organização de fatos foi iniciada pela escrita manuscrita, foi intensificada pela imprensa, e hoje é maximizada pelo uso de computadores. A cultura impressa e a informática facilitam o acesso, armazenamento e recuperação de textos ou livros de gêneros variados, proporcionando ao leitor grande variedade de conteúdo [Condemarín e Alliende 2005].

A produção de textos é estimulada pela leitura. Antes, a leitura e a produção de textos eram avaliadas de forma separada, a leitura como o processo de decodificação e a escrita à codificação do texto. Estudos comprovaram que escrita e leitura estão ligadas, apoiadas e envolvidas com o pensamento. Estudantes que participam de atividades de leitura e escrita em conjunto desenvolvem níveis de pensamento mais elevados e a produção de textos é beneficiada quando o leitor é submetido a vários gêneros literários [Condemarín e Alliende 2005].

O leitor pode modificar preconceitos, crenças, opiniões ou pontos de vista com a leitura de entrevistas, crônicas, artigos, documentos ou livros. Existem muitas possibilidades e pontos de vistas diferentes para o mesmo problema, através da leitura de diferentes textos ou livros o leitor pode confrontar idéias e formar sua opinião sobre o assunto. Segundo Condemarín e Alliende [Condemarín e Alliende 2005] as pessoas que não possuem o hábito de ler apresentam idéias fixas e dificuldade para aceitar o diferente, em oposição a esta realidade, as pessoas que possuem o costume de ler aceitam com maior facilidade idéias novas e pontos de vista distintos.

2.1.1 Compreensão de Leitura

Ler é a habilidade complexa que implica desde a decodificação dos símbolos escritos até a fase de análise reflexiva do seu conteúdo. A compreensão total do texto depende de muitas variáveis. Na perspectiva da teoria cognitiva é necessário que o leitor recupere o esquema mental que existe na memória com vocabulário e conhecimento prévio sobre as informações sin-

táticas e semânticas do texto [Joly e Istome 2008].

Segundo Condemarín e Alliende [Condemarín e Alliende 2005], a compreensão da linguagem escrita é a meta fundamental da leitura. Para que isso aconteça os leitores determinam relações entre o seu conhecimento anterior e a nova informação que o texto lhe oferece, fazendo inferências, comparações e formulando perguntas relacionadas ao conteúdo.

Muitos fatores que determinam a compreensão de leitura estão misturados entre si e mudam continuamente. O problema de compreensão de leitura é apontado como o processo de reconstrução por parte do leitor do sentido dado pelo escritor ao texto. O autor deve escrever de acordo com seu público alvo. Por exemplo: os contos infantis usam palavras e contextos de acordo com os níveis de compreensão das crianças. O texto depois de criado se torna independente de seu autor e o leitor aplica seus esquemas cognitivos e conhecimento prévio sobre o assunto durante a leitura (patrimônio de conhecimentos no contexto denominado "circunstâncias da leitura"). A Figura 2.1 realiza descrição do processo de compreensão de textos considerando três aspectos: autor, texto e leitor [Condemarín e Alliende 2005].

Emissor (autor escritor)	Mensagem (texto escrito)	Receptor (leitor)
UTILIZA - Código do emissor. - Esquemas cognitivos do emissor. - Patrimônio de conhecimentos do emissor. EM Circunstância da emissão (escrita).	Texto da expressão (reproduz a expressão do autor) SEUS COMPONENTES SÃO: - Características físicas do texto. - Componentes lingüísticos oracionais e textuais. - Componentes referenciais. TRANSFORMA-SE EM: Texto da interpretação. Permite a interpretação do receptor (leitor).	APLICA AO TEXTO A INTERPRETAÇÃO: - Códigos do receptor (leitor). - Esquemas cognitivos do receptor (leitor). - Patrimônio de conhecimento do receptor (leitor). EM Circunstâncias da recepção (leitura).

Figura 2.1: Descrição do processo de compreensão de textos [Condemarín e Alliende 2005].

Fatores da compreensão provenientes do autor

Os fatores da compreensão provenientes do autor estão relacionados:

- códigos usados - saber os códigos usados pelo autor pode auxiliar na compreensão de elementos que aparecem nos seus textos;
- esquemas cognitivos - entender os esquemas cognitivos usados pelo autor no texto, ou seja, o que o conjunto de palavras usadas no contexto deseja expressar;
- patrimônio cultural - os costumes de determinado tempo ou ciência podem interferir no significado das palavras no contexto;
- circunstâncias que o texto foi elaborado - lugar e tempo que o texto foi escrito, idade ou profissão do autor.

Fatores da compreensão derivados do texto

Os mais importantes fatores derivados do texto podem ser divididos em referenciais, linguísticos e físicos. Eles podem de certa forma definir a legibilidade do texto. A leitura é fácil quando não existem lacunas entre o conhecimento do leitor e o tema do texto. As provas da avaliação de leitura procuram determinar a capacidade do aluno de reconhecer:

- idéia principal e idéias secundárias;
- percepção de detalhes;
- relações de causa e efeito;
- caracterização de lugares e pessoas;
- opiniões;
- ou acontecimentos reais e imaginários.

A lingüística textual estuda as características, as regras de construção, relação e reconhecimento das partes do texto. O leitor a partir do texto pode identificar:

- se o texto está completo;
- as partes divergem ou concordam entre si;
- se o desenvolvimento de idéias pode ser antecipado.

O leitor experiente encara, sem maiores dificuldades problemas de lingüística textual, mas durante o processo de aprendizagem muitas dificuldades tem origem em fatores ligados a estrutura lingüística.

A identificação da estrutura textual avalia como os componentes do texto se relacionam com a compreensão do texto, analisando:

- elementos pré-textuais - disposição das palavras, orações e parágrafos no texto;
- elementos estruturais - indica o tipo de texto (dissertativo, narrativo ou poema);
- elementos supra-estruturais - conjunto de relações que o texto possui com outros textos.

Fatores da compreensão provenientes do leitor

A compreensão da leitura vai depender de fatores provenientes do leitor, são eles:

- código lingüístico - vocabulário, estruturas lingüísticas;
- esquemas cognitivos - conhecimento prévio do leitor sobre o assunto do texto;
- patrimônio cultural - contexto de vivência do leitor;
- interesse e valores - escolher assuntos que sejam atrativos para o leitor;
- circunstâncias da leitura - tempo e local da leitura.

A compreensão de leitura depende de vários fatores complexos e inter-relacionados. O conhecimento desses fatores favorece a identificação das dificuldades de compreensão e as formas de aperfeiçoá-la [Condemarín e Alliende 2005].

2.1.2 Métodos para Avaliação da Compreensão de Leitura

A leitura é a reconstrução dos significados da linguagem representado por símbolos gráficos. Para avaliar a compreensão de leitura é imprescindível examinar com atenção o nível de habilidade leitora e o grau de complexidade do texto [Condemarín e Alliende 2005]. Para se obter a compreensão total do texto na perspectiva da teoria cognitiva, é essencial que o leitor recupere o esquema mental existente na memória com o vocabulário e conhecimento que se relacionem com as informações, semânticas e sintáticas contidas no texto [Joly e Istome 2008].

A avaliação da compreensão da linguagem escrita implica na leitura e interpretação de frases, enunciados e textos para verificar as habilidades sintáticas, semânticas e lexicais do leitor. As habilidades sintáticas referem-se a

ordem das palavras na frase, concordâncias, localização de morfemas, as construções gramaticais e as relações entre as sentenças. Nas habilidades semânticas e lexicais é importante identificar a quantidade de palavras (vocabulário passivo e ativo) que o leitor classifica ordena e utiliza, avaliam também o reconhecimento das relações entre palavras locuções e frases com os respectivos objetos, fatos e conceitos que representam [Joly e Istome 2008].

Para avaliar de forma eficiente a leitura é necessário conhecer a realidade dos alunos avaliando alguns parâmetros, tais como: idade cronológica, grau de escolaridade, nível sócio-econômico, dentre outros. Conhecendo essas habilidades de leitura os professores podem selecionar textos adequados a realidade dos estudantes. Durante o processo de elaboração do material é necessário reconhecer quais as habilidades específicas exigem aprimoramento [Condemarín e Alliende 2005]. Este trabalho descreve três formas de avaliação de leitura: o Teste de Cloze, avaliações taxonômicas e o programa de computador *Summary Street*.

Teste de Cloze

O Teste de Cloze foi proposto por Taylor em 1953 como recurso de avaliação de compreensão de leitura [Santos 2004]. Através do Teste de Cloze é possível estimar o nível de compreensão de leitura. Também pode ser usado como estratégia de desenvolvimento da compreensão da leitura e como técnica para avaliar a legibilidade dos textos.

O Teste de Cloze apresenta as seguintes vantagens:

- é fácil de elaborar, administrar e interpretar;
- não é necessário ser um especialista para construir, administrar e interpretar os resultados;
- as respostas dos alunos estão baseadas apenas em indícios dados pelo texto e não são induzidas ou enfraquecidas pelas perguntas do examinador;
- podem ser usados textos breves de matérias científicas, prosas, notícias, dentre outros.

A análise das respostas dos alunos avalia se eles são capazes de usar os indícios semânticos e sintáticos dados pelo contexto. Se eles possuem dificuldade com algum tipo particular de indício que necessite de apoio direto ou explícito. A análise também contribui para conscientizar os alunos de suas estratégias de antecipação e adivinhação e da função da repetição de informações proporcionada pelo contexto [Condemarín e Alliende 2005].

O Teste de Cloze consiste em solicitar ao estudante que acrescente palavras suprimidas no texto. As omissões são feitas em intervalos regulares que geralmente variam entre 5 e 10 palavras. Deve-se evitar suprimir nomes próprios e palavras da primeira e última oração do texto [Condemarín e Alliende 2005]. Pode-se suprimir seletivamente substantivos, verbos ou algumas combinações de classes de palavras e pedir para o aluno adivinhar a palavra suprimida adequada para o contexto. A palavra suprimida é substituída por um traço de tamanho sempre igual ou de tamanho proporcional a palavra suprimida [Santos 2004]. A Figura 2.2 ilustra o texto de Luiz Fernando Verissimo chamado Desentendimento, usado no estudo de Santos [Santos et al. 2002], as palavras sublinhadas correspondem as palavras omitidas do texto.

“A comunicação com garçons pode ser uma provação em qualquer parte do mundo. Costumo tomar leite frio, de manhã. Em Tóquio, na primeira vez em que nos aventuramos a tomar café fora do hotel, fui munido da palavra japonesa para leite. Miruku, ou coisa parecida. Experimentei-a no garçom, que me devolveu um olhar cheio de perplexidade. Tentei de novo, em várias flexões. Finalmente acertei: era preciso dizer a palavra rapidamente. Veio o leite. Quente. Não sei que mímica usei para dizer que o queria frio - duvido que tenha me abraçado e simulado uma tremedeira, o que só faria o garçom sair atrás de alguma corrente de ar perdida dentro do restaurante -, mas ele entendeu, levou o copo e o trouxe de volta. Com pedras de gelo dentro do leite. Um japonês que observava nosso desentendimento de uma mesa ao lado se ofereceu para ajudar. Traduziu meu pedido para o garçom. Tudo esclarecido. Agradecei a intermediação enquanto o garçom voltava para a cozinha com o copo e pouco depois o trazia de volta. Afastada a ameaça à paz mundial e à amizade entre os povos, levantei o copo em direção do tradutor num brinde agradecido. O leite estava morno.

Em Budapeste entramos num restaurante em que o menu era em húngaro e alemão, mas o húngaro é tão difícil que entendemos em alemão. Não tive dificuldade em me fazer entender pelo garçom. Apontei com o dedo para a única palavra em alemão que qualquer pessoa precisa saber, além de bite e danke: forelle. Quer dizer truta. Eu poderia andar pela Alemanha comendo trutas em todas as refeições - salvo no café da manhã, claro. São as melhores do mundo, e não havia razão para as trutas húngaras não serem parecidas. Não fiquei sabendo. Não havia forelle em qualquer língua no restaurante. Acabei apontando com o dedo para a palavra com menos consoantes que vi, e que acabou sendo o nome de uma almôndega do tamanho aproximado de uma granada. Sobrevivi.”

Figura 2.2: Ilustração do Teste de Cloze.

O Teste de Cloze pode ser modificado para:

- múltipla escolha - várias palavras são colocadas ao lado do espaço em branco;
- lista de palavra - as palavras omitidas são listadas em ordem alfabética na margem;
- indícios fônicos - algumas letras da palavra omitida são colocadas no espaço em branco.

É importante aplicar testes pilotos com os alunos antes de aplicar o Teste de Cloze. Alguns autores acreditam que só deve ser pontuada se a palavra estiver exata, porque aceitando sinônimos pode reduzir a confiabilidade, mas outros autores acreditam que palavras da mesma família semântica devem ser aceitas desde que não altere o significado [Condemarín e Alliende 2005].

Avaliação Taxonômica

Muitos psicólogos e professores acreditam que as habilidades de leitura podem ser identificadas através da avaliação taxonômica. Durante as avaliações taxonômicas a compreensão de leitura é considerada como conjunto de habilidades que devem ser desenvolvidas separadamente. Elas apresentam a seguintes vantagens: permite fixar limites dos aspectos críticos da compreensão de leitura; usar a estratégia de compreensão de bons leitores para ensinar leitores imaturos.

Segundo Condemarín e Alliende [Condemarín e Alliende 2005] a Taxonomia de Barret inclui dimensões afetivas e cognitivas da compreensão de leitura de textos narrativos. Os principais pontos da taxonomia de Barret são:

- compreensão literal de idéias explícitas no texto, dividida em reconhecimento (identificar e localizar idéias explícitas no texto) e lembrança (o estudante reproduz fatos, idéias e informações do texto);
- reorganização - organizar e sintetizar as idéias ou informações do texto;
- compreensão inferencial - exige do estudante o uso de idéias e informações explícitas do trecho e sua intuição e experiência pessoal como base para as conjecturas e hipóteses;
- leitura crítica - comparar idéias do texto com outros critérios do leitor ou de outros autores;
- avaliação - tenta avaliar o impacto psicológico ou estético que o texto provocou no leitor.

Summary Street

O programa de computador *Summary Street* avalia a compreensão de leitura através da produção de resumos. O *Summary Street* fornece *feedback* se o resumo elaborado pelo aluno apresenta pouco ou excesso de informações. Ele foi desenvolvido para falantes nativos da língua inglesa.

O *Summary Street* utiliza a Análise Semântica Latente (ASL), modelo estatístico que procura explicar como o significado das palavras pode ser derivado a partir de dados estatísticos originados pela experiência. A teoria pressupõe a existência de padrões do uso das palavras em determinado contexto. O ASL é modelo associativo linear de alta dimensionalidade e não usa o conhecimento humano além do pressuposto pelo seu mecanismo de aprendizado. O modelo ASL obtém experiência com a grande quantidade de textos eletrônicos fornecidos durante a fase de aprendizagem e a partir deles extrai aspectos matemáticos de caráter estatístico e indutivo. A semelhança entre o programa e o homem é a capacidade de lidar com dados textuais e de inferir novos dados a partir de conhecimento adquirido pela experiência com textos escritos.

Para analisar o texto o modelo ASL gera uma matriz de ocorrência de cada palavra na frase ou parágrafo. As linhas da matriz representam tipos de eventos unitários (palavras), enquanto as colunas representam os contextos que elas acontecem. As células apresentam a frequência de cada palavra no contexto. Depois de estabelecido o peso das células, a matriz é submetida à decomposição de valor-singular. O ASL capta os efeitos de semelhança entre as palavras através dos padrões de uso delas em diferentes contextos, estabelecendo de forma automática a correlação de ordem contextual do uso das palavras sem a necessidade de uma análise da ordem nas frases [Poersch e Rosa 2007].

2.1.3 Avaliação de Leitura com Adolescentes e Crianças

Esta seção realiza uma breve revisão bibliográfica sobre avaliação de compreensão de leitura com adolescentes e crianças no Brasil e nos Estados Unidos da América. Descrevendo os resultados encontrados utilizando técnicas como o Teste de Cloze.

O trabalho de Phelippe e Witter [Phelippe e Witter 2007] comparou a compreensão de leitura entre meninos e meninas estudantes da quinta série do Ensino Fundamental e a influência de estímulos a leitura em casa e na escola. Participaram da pesquisa 19 (seis meninos e treze meninas) alunos de uma escola pública com idade média de nove anos e seis meses. O estudo usou

questionários para verificar o estímulo a leitura em casa e na escola, também foi aplicado o Teste de Cloze para avaliar o nível de compreensão de leitura dos participantes. Os resultados revelam semelhanças entre os gêneros masculino e feminino e nível de compreensão de leitura abaixo do esperado, com poucos incentivos em casa e na escola.

O objetivo do estudo de Nikaedo, Kuriyama e Macedo [Nikaedo, Kuriyama e Macedo 2007] foi verificar a correlação dos resultados em leitura e escrita entre dois procedimentos de avaliação: quantitativos e qualitativos. Participaram da pesquisa 52 alunos da 1^o série do Ensino Fundamental de uma escola pública municipal localizada na zona leste da cidade de São Paulo. Do total, 22 meninas e 30 meninos com idade média de 7,1 anos. O desempenho escolar foi mensurado através de avaliações quantitativas e qualitativas. A avaliação quantitativa usou três testes de leitura e escrita da BALE-On-line (teste de competência de leitura de palavras, teste de nomeação de figura escrita e teste de compreensão de sentença escrita). A avaliação qualitativa é realizada pelo professor da turma e descreve o estágio de alfabetização dos alunos. Os resultados do estudo mostraram correlações positivas entre os três instrumentos quantitativos e destes com a avaliação qualitativa dos professores. Nas avaliações quantitativas professores tenderam a superestimar o desempenho de alguns alunos.

O trabalho de Santos [Santos 2004] usou o Teste de Cloze como método para diagnosticar e remediar a compreensão de leitura entre alunos da 5^o série de uma escola pública no interior de São Paulo. Participaram da pesquisa 24 alunos (11 do gênero masculino e 13 do gênero feminino) com idade média de 13 anos. Inicialmente foi aplicado o questionário de caracterização, com itens referentes à identificação do aluno, informações sobre seus hábitos e dificuldades de leitura e estudo. No pré-teste e no pós-teste foi aplicado o Teste de Cloze tradicional com o quinto vocábulo omitido, buscando diagnosticar a compreensão de leitura. Nas outras etapas do procedimento os trechos foram elaborados com a omissão do décimo vocabulário, inserindo dificuldades crescentes para o seu preenchimento. O estudo verificou a baixa frequência de leitura entre os alunos e apontou a mãe e a professora como as maiores incentivadoras da leitura. Para se calcular o nível de compreensão os pesquisadores usaram a classificação de Bormuth ¹. Os dados demonstraram que dos 24 participantes do programa de remediação, 11 continuaram

¹ Percentual de acertos até 40% no Teste de Cloze corresponde ao nível de frustração, de 41% a 56% ao nível institucional e maior ou igual 57% corresponde ao nível de independência.

no nível de frustração (não apresentaram progressos significativos), quatro alunos passaram para o nível institucional e dez alunos saltaram do nível de frustração para o nível independente de leitura.

O estudo de Martins e Pennington [Martins e Pennington 2001] procurou investigar a correlação entre, por um lado, a nomeação seriada rápida ² e a consciência de fonemas ³ e, por outro, diversas medidas de habilidade de leitura e escrita. Participaram da pesquisa 146 crianças e adolescentes norte-americanos, entre sete e 18 anos de idade. Aproximadamente metade dos participantes da pesquisa apresentava dificuldades de leitura. Os resultados indicam que a nomeação seriada rápida contribui para a variação na habilidade de leitura e escrita de forma independente da consciência de fonemas. Análises adicionais sugerem que a habilidade subjacente à nomeação seriada rápida é importante para o desenvolvimento da habilidade de ler textos de forma rápida e eficiente. Por outro lado, a consciência fonêmica é importante para o desenvolvimento da habilidade de ler através da conversão das letras ou grupos de letras em seus sons correspondentes. O trabalho concluiu a pouca evidência de que a nomeação seriada rápida e a consciência de fonemas variem em função da idade ou da habilidade de leitura.

2.1.4 Avaliação de Leitura com Universitários

Esta seção dedica-se a avaliação de compreensão de leitura entre os universitários brasileiros. Descreve também os métodos e resultados obtidos por diversos pesquisadores.

O hábito de ler é fundamental para o bom desempenho do universitário nas disciplinas do curso superior. A leitura é importante para o contexto universitário, sendo interessante à manutenção e desenvolvimento de programas de atendimento aos alunos. A primeira etapa desses programas é avaliar o desempenho em leitura de todos os universitários, e com base nos resultados obtidos eles poderão ser encaminhados para programas: remediativos (corrigir falhas ou superar limitações em leitura); desenvolvimento (aperfeiçoamento da habilidade de ler); especiais (recomendados para qualquer tipo de leitor) [Silva 2004].

É importante estimular os universitários a desenvolver o hábito da leitura e escrita como atividades rotineiras na Universidade. O ensino da

²O participante nomeia de forma rápida e corretamente o quanto possível, o conjunto de estímulos visuais apresentado serialmente (letras, números, cores ou objetos).

³Avalia a capacidade de manipular e segmentar fonemas.

Língua Portuguesa nos primeiros anos de Universidade é essencial e deve ser direcionado para as necessidades do aluno na vida acadêmica e profissional. As instituições de ensino superior devem incentivar e auxiliar a qualificação de professores para que eles possam estimular a leitura dentro e fora da sala de aula [Silva 2004].

Nos últimos 20 anos, a clientela escolar brasileira mudou. Os alunos atualmente não pertencem apenas aos grupos sociais dominantes. Esses universitários fazem parte das classes populares e desconhecem a variedade escrita e oral utilizada na Universidade, dificultando o processo de aprendizagem. Atualmente estudantes com deficiências em habilidades consideradas básicas (ler e escrever, por exemplo) ingressam nas Universidades brasileiras [Silva 2004].

O objetivo da pesquisa de Tessaro [Tessaro 2004] foi obter conhecimento sobre a formação de leitores entre os alunos do quinto ano do curso de Psicologia de três instituições de ensino superior. Participaram da pesquisa 72 alunos do quinto ano de Psicologia de Universidades nos estados de Mato Grosso do Sul, Paraná e São Paulo. Os pesquisadores utilizaram questionários com diversas perguntas sobre leitura. Os resultados mostraram que os livros infantis são a lembrança mais antiga de leitura entre os universitários pesquisados. A escola e o professor foram os itens mais relevantes para aprendizado da leitura. Os participantes aprenderam ler entre os seis e sete anos de idade e tiveram boas experiências com a leitura durante sua vida e evidenciaram a mãe e a professora como incentivadoras da leitura.

O trabalho de Oliveira e Santos [Oliveira e Santos 2005] analisa a relação entre compreensão de leitura e o desempenho acadêmico na avaliação de aprendizado em disciplinas de diferentes cursos superiores. Participaram da pesquisa 270 alunos calouros dos cursos de direito, administração e psicologia de uma universidade particular no interior do estado de São Paulo. Os instrumentos usados na pesquisa foram: questionário de caracterização dos tipos de avaliação e o Teste de Cloze. A pesquisa evidenciou que o método mais utilizado por professores para avaliar o desempenho foi a prova dissertativa individual e menos usado foi a prova objetiva de grupo. Os resultados encontrados apontaram correlação entre a compreensão de leitura, desempenho acadêmico e o tipo de avaliação utilizada pelo professor.

O estudo de Cunha e Santos [Cunha e Santos 2006] teve como meta verificar a relação entre a compreensão de leitura e a produção escrita em universitários. Participaram da coleta de dados 134 ingressantes de cursos das áreas exatas, humanas e biológicas. O Teste de Cloze foi usado para avaliar

a compreensão de leitura e uma redação produzida pelos alunos foi usada para avaliar a escrita dos universitários. Os resultados indicaram estudantes que apresentaram baixos índices de compreensão de leitura, também tiveram maior quantidade de erros gramaticais, evidenciando a relação entre essas atividades.

2.1.5 Leitura Proveitosa de Textos Acadêmicos

A leitura proporciona a ampliação e aquisição de novos conhecimentos e a abertura de novos horizontes. É necessário ler muito e sempre, porque grande parte do conhecimento é adquirido através da leitura. Segundo Marconi e Lakatos [Marconi e Lakatos 2007] ler significa interpretar, conhecer, compreender algo obscuro e classificar os elementos importantes e secundários do texto. Durante a leitura é importante se atentar para os seguintes aspectos: ler com um determinado objetivo e avaliar o conteúdo do texto; consultar dicionários e outras fontes se não conhece todas as palavras usadas pelo autor; interromper a leitura se julgar que o conteúdo não é importante.

Uma leitura proveitosa requer:

- atenção - concentração;
- intenção - interesse em adquirir novos conhecimentos através da leitura;
- reflexão - consideração sobre o que se lê;
- espírito crítico - avaliar o texto, julgar;
- análise - fase mais importante da leitura, identifica o que é importante no texto;
- síntese - reconstrução do texto, elaboração do resumo.

Os níveis de leitura são: *scanning* (varredura no texto procurando identificar palavras-chaves ou frases importantes); *skimming* (captação das tendências gerais, olhando figuras e gráficos ou alguns parágrafos do texto); do significado (visão geral do conteúdo deixando de lado aspectos secundários, fazendo tudo de uma vez sem voltar); de estudo ou informativa (compreensão do texto, sublinhamento e elaboração de resumos); crítica (estudo e formação do ponto de vista sobre o texto, requer do leitor conhecimento prévio sobre o assunto) [Marconi e Lakatos 2007].

A leitura de **estudo ou informativa** é o nível necessário para a compreensão de textos acadêmicos. As fases da leitura informativa são: pré-

leitura (visão global do assunto abordado e seleção das obras que contém o assunto de interesse. Nesta fase são analisados folha de rosto, índices, referências bibliográficas, notas ao pé da página, prefácio, introdução e conclusão); seletiva (eliminação do que não é importante e concentração nas informações realmente interessantes para o leitor); reflexiva (avaliação das informações do texto e as intenções do autor); interpretativa (analisar resultados e conclusões do autor) [Cervo, Bervian e Silva 2007]

É importante discutir sobre o sublinhamento e sua utilidade durante a leitura. Muitas pessoas possuem dificuldades de decidir o que é realmente importante no texto e acabam sublinhando muito mais que o razoável. Para solucionar o problema, emprega-se o sublinhamento à lápis durante a primeira leitura. Segundo Soares e Santos [Soares e Santos 2007], para que se identifique realmente as partes mais importantes do texto é necessário vê-lo globalmente. Portanto, na segunda leitura, deve-se consolidar o sublinhamento (possivelmente à caneta), eliminando-se repetições e mal julgamentos de importância. Para evitar sublinhamento excessivo, sugere-se usar pequenas régua (cartões de crédito, inclusive) e lembrar os leitores de que os termos sublinhados devem ser memorizados.

Esta pesquisa adverte a inadequação da técnica utilizada durante o ensino fundamental, quando se tenta ler devagar para compreender completamente o texto. Textos acadêmico-científicos exigem muitos conhecimentos prévios, sendo destinados à divulgação de resultados científicos. Para atender a esse público, convém utilizar a técnica de aproximações sucessivas, isto é, aumentar a compreensão gradativamente. Mais especificamente, quatro passos são sugeridos:

1. Reconhecimento inicial permite o planejamento da leitura, noção do tamanho, complexidade e estrutura do texto.
2. Primeira leitura rápida sem parar e sublinhamento a lápis.
3. Segunda leitura mais lenta e consolidação do sublinhamento, se a compreensão foi insuficiente outra leitura mais lenta.
4. Síntese do texto (esquema antes do resumo).

2.1.6 Índice de Legibilidade

Os textos informativos devem ser elaborados para todos os tipos de leitores. O primeiro passo para elaborar texto legível é imaginar o tipo e o nível de conhecimento do leitor sobre o assunto. O segundo passo é organizar

as idéias no texto para facilitar a compreensão do leitor. Muitos escritores ignoram a importância da legibilidade, utilizando linguagem confusa e palavras pouco usadas [Fulgencio e Liberato 2007].

Fórmulas de legibilidade são usadas a mais de cinquenta anos na língua inglesa para avaliar e adequar os textos ao público de interesse. Existem mais de trinta e cinco fórmulas para calcular a legibilidade, geralmente considerando o comprimento das frases, quantidade média de palavras e de sílabas ou letras [Nunes e Júnior 2000].

O conceito de índice de legibilidade surgiu no trabalho de Flesch em 1948 para a língua inglesa. Flesch procurou evidenciar a relação entre tamanhos médios de palavras e sentenças e facilidade de leitura. O índice de legibilidade não inclui aspectos de compreensão de texto. O trabalho do Núcleo Interinstitucional da Linguística Computacional foi o primeiro a elaborar um índice para a língua portuguesa no Brasil. Foi realizado um estudo comparativo de textos originais em inglês e traduzidos para o português. O estudo concluiu que os parâmetros da equação do índice Flesch precisavam ser adaptados para o português porque as palavras da língua portuguesa possuem em média mais sílabas [Nunes e Júnior 2000].

O método de facilidade de leitura de Flesch classifica a legibilidade de um texto em uma escala que varia de zero (difícil) a 100 (muito fácil). O método considera a quantidade de palavras por frase e a quantidade de sílabas por palavra. Frases e palavras longas reduzem a legibilidade do texto. A fórmula para calcular o Índice de Legibilidade de Flesch (ILF) é apresentada na Equação 2-1 [Cavique 2008].

$$ILF = 206,835 - ((1,0 * CMF) + 0,846 * (MSP)) \quad (2-1)$$

onde:

CMF = Quantidade média de palavras por frase.

MSP = Quantidade média de sílabas por palavra.

A Tabela 2.1 apresenta a relação entre o Índice de Flesch e o nível de instrução necessário para leitura dos textos.

Tabela 2.1: *Índice de Flesch aplicado a textos brasileiros*
[Nunes et al. 1996].

Índice Flesch	Legibilidade	Nível de instrução
100 - 75	muito fácil	da 1 ^a à 4 ^a série do ensino fundamental
74 - 50	fácil	da 5 ^a à 8 ^a série do ensino fundamental
49 - 25	difícil	ensino médio e nível superior
24 - 0	muito difícil	textos acadêmicos

Testes realizados com textos dirigidos ao público dos quatro níveis de instrução apresentaram índices satisfatórios. Jornais de grande circulação como o Folha de São Paulo apresentam bons índices de legibilidade Flesch para leitores com escolaridade equivalente ao final do ensino fundamental [Nunes et al. 1996].

O índice de legibilidade de Flesch é calculado para trechos com mais de 100 palavras [Nunes e Júnior 2000]. O presente trabalho procura analisar a legibilidade de parágrafos e frases tornando-se inviável a utilização do índice de Flesch. O teste de legibilidade citado por Squarisi e Salvador [Squarisi e Salvador 2008] foi adaptado por Alberto Dines para o português e avalia parágrafos isolados. O cálculo é realizado em seis etapas:

1. Conte as palavras do parágrafo
2. Conte as frases do parágrafo
3. Cálculo da média palavra / frase
4. Some a média palavra / frase ao número de polissílabos
5. Multiplique o resultado por 0,4
6. O produto da multiplicação é o índice de legibilidade

Artigos, cartas podem ser avaliados pelo índice de legibilidade. Se resultado ficou acima de 15 é necessário revisar o texto. A solução é reduzir o tamanho das frases e a quantidade de palavras proparoxítonas. Os resultados possíveis do índice de legibilidade estão na Tabela 2.2:

Tabela 2.2: *Índice de Legibilidade*
[Squarisi e Salvador 2008]

Índice de Legibilidade	Resultados
1 - 7	história em quadrinhos
8 - 10	excepcional
11 - 15	ótimo
16 - 19	pequena dificuldade
20 - 30	muito difícil
20 - 30	linguagem técnica
acima de 41	nebulosidade

As fórmulas de legibilidade levam em conta as estruturas morfossintáticas das orações e as palavras usadas como o mais importante. Por comprovações estatísticas chegaram a conclusão de que existe uma relação entre a quantidade de palavras da frase e a dificuldade de compreensão. As orações

com estruturas sintáticas complexas, com grande quantidade de elementos subordinados, ou longas e complicadas dificultam a compreensão de leitura [Condemarín e Alliende 2005].

A supersimplificação também pode ser obstáculo para a compreensão. O texto pode conter apenas orações simples sem coesão que indiquem relações espaciais e temporais ou de causa [Condemarín e Alliende 2005]. Alguns cuidados durante processo de elaboração do texto podem auxiliar na compreensão do leitor. Podem-se citar algumas dicas importantes: a redação de textos de forma clara e simples; uso de palavras simples e frases médias (com duas ou três linhas); Eliminação de anbiguidades nas frases e verbos no gerúndio; Frases na ordem lógica (sujeito, verbo e complemento ou contexto, sujeito, verbo e complemento); Consistência lógica [Squarisi e Salvador 2008].

2.2 Redes Neurais Artificiais

A inteligência artificial é a área da ciência da computação que procura desenvolver sistemas que sejam capazes de simular faculdades humanas. O sistema para ser considerado inteligente deve ser capaz de fazer três coisas: guardar o conhecimento adquirido; aplicar esse conhecimento para resolver problemas; e aprender com as experiências [Haykin 2001].

As redes neurais artificiais são sistemas de processamento de informações com características semelhantes as redes neurais naturais. Os modelos de redes neurais artificiais são baseados em generalizações de modelos matemáticos da biologia neural [Fausett 1994]. Segundo Haykin [Haykin 2001] a rede neural é um processador paralelo distribuído composto por unidades de processamento simples, que possuem tendência natural para armazenar conhecimento experimental e torná-lo disponível para o uso. Elas apresentam comportamento semelhante ao cérebro em dois pontos: 1) o conhecimento é adquirido pela rede através do processo de aprendizagem 2) os pesos sinápticos são usados para armazenar o conhecimento adquirido.

Os elementos que identificam as Redes Neurais são: 1) os padrões de conexões entre os neurônios (denominados arquitetura), 2) os métodos que determinam os pesos das conexões e 3) a função de ativação. As redes neurais são utilizadas em diversas áreas do conhecimento, dentre elas pode-se citar: processamento de sinal, controle de processos industriais, medicina, classificação de dados e reconhecimento de voz [Fausett 1994].

Pode-se ser destacadas algumas vantagens do uso de redes neurais artificiais tais como: não linearidade; adaptabilidade (alta capacidade de se

adaptar seus pesos sinápticos as modificações do meio ambiente); informação contextual (todos os neurônios da rede são afetados pela atividade de todos os outros neurônios da rede); tolerância a falhas (quando implementada em hardware é capaz de ser pouco vulnerável a falhas) [Haykin 2001].

2.2.1 Breve Histórico

O primeiro modelo artificial de um neurônio foi resultado do trabalho de Warren McCulloch e Walter Pitts em 1943. McCulloch era psiquiatra e Pitts matemático. No artigo *A logical calculus of the ideas immmanent in nervous activity* McCulloch e Pitts descrevem o modelo artificial do neurônio, expondo suas capacidades computacionais [Haykin 2001].

Em 1949, Donald Hebb escreveu o livro chamado *The Organization of Behavior* onde foi apresentado pela primeira vez a lei de aprendizagem específica para as sinapses dos neurônios. Hebb propôs que as conexões do cérebro são alteradas conforme o organismo aprende tarefas diferentes e os agrupamentos neurais são criados por essas modificações [Haykin 2001].

Frank Rosenblatt em 1958 introduziu e desenvolveu a classe de redes neurais artificiais chamadas perceptrons. O típico modelo de perceptron consiste em camada de entrada (a retina), camada intermediária (recebe os impulsos da camada de entrada e envia para a camada de saída) e camada de saída [Fausett 1994].

Em 1960, Widrow e Hoff desenvolveram a regra delta ou algoritmo do mínimo quadrado médio e o utilizaram para formular o Adaline (elemento linear adaptativo, do Inglês *adaptive linear element*). A diferença entre o perceptron e o Adaline é o procedimento de aprendizagem. O objetivo do treinamento no Adaline é minimizar o erro quadrático médio entre a saída da rede e a saída desejada [Haykin 2001]. No ano de 1969 Minsky e Papert usaram a matemática para demonstrar que o perceptron de uma camada era capaz de solucionar apenas problemas linearmente separáveis [Fausett 1994].

Muitos fatores contribuíram para o pouco interesse em Redes Neurais durante os anos 70, dentre eles podem ser citados, a falta de financiamento e dificuldades tecnológicas. Muitos pesquisadores durante esse período abandonaram as pesquisas. Na perspectiva de engenharia os anos 70 é considerado uma década de adormecimento das redes neurais [Haykin 2001].

Em 1982, Hopfield publicou o artigo "*Neural networks and physical systems with emergent collective properties*" que chamou atenção para as propriedades associativas das redes neurais. Em 1986 foi desenvolvido o algo-

ritmo de retropropagação. Em 1987 aconteceu em São Francisco a primeira conferência de redes neurais a ICANN (*International Conference on Artificial Neural Networks*) e foi formada a INNS (*International Neural Networks Society*). A partir dos anos 80 muitos pesquisadores iniciaram pesquisas que aperfeiçoaram as redes neurais e ofereceram soluções para diversos problemas.

2.2.2 Neurônio Biológico

As redes neurais biológicas foram a inspiração para as redes neurais artificiais. O elemento de processamento (neurônio artificial) é uma analogia ao neurônio biológico. O neurônio biológico apresenta três tipos de componentes que são de particular interesse para a compreensão dos neurônios artificiais: os dendritos, soma e axônio. Os dendritos recebem sinais de outros neurônios. A soma é a responsável por reunir e combinar as informações vindas de outros neurônios. O axônio envia os estímulos para outras células. Os impulsos elétricos são transmitidos pelas sinapses através de processos químicos. A Figura 2.3 ilustra típico neurônio biológico [Fausett 1994].

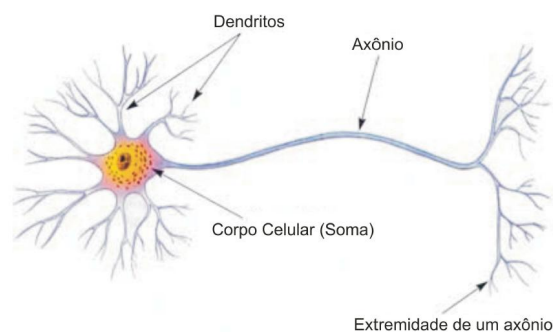


Figura 2.3: *Neurônio biológico.*

Diversas características do neurônio artificial foram inspiradas nos neurônios biológicos, dentre elas podemos citar: capacidade de receber vários sinais; os sinais podem ser modificados pelos pesos das sinapses de recepção; o neurônio artificial soma os pesos da entrada; a saída do neurônio pode ser a entrada para vários neurônios; o processamento da informação é local; a memória é distribuída; as conexões sinápticas podem ser alteradas pela experiência; tolerância a falhas [Fausett 1994].

2.2.3 Neurônio Artificial

Segundo Haykin [Haykin 2001] o neurônio é a unidade de processamento de informação necessário para o funcionamento da rede neural artificial. O neurônio artificial é dividido em três partes:

1. Conjunto de sinapses: o sinal x_j na entrada da sinapse j conectada ao neurônio k é multiplicada pelo peso sináptico w_{ij} .
2. Somador: realiza a soma dos sinais de entrada ponderados pelas sinapses do neurônio.
3. Função de ativação: usada para delimitar a amplitude da saída do neurônio. Geralmente o intervalo normalizado da amplitude da saída do neurônio é o intervalo fechado $[0,1]$ ou $[-1,1]$ [Haykin 2001]

O neurônio k ilustrado na Figura 2.4 recebe o sinal de saída (a função de ativação) dos neurônios $x_1, x_2, \dots, x_m$. Os pesos para as conexões de $x_1, x_2, \dots, x_m$ para o neurônio k são $w_1, w_2, \dots, w_m$. A saída de k será a soma ponderada dos sinais dos neurônios $x_1, x_2, \dots, x_m$.

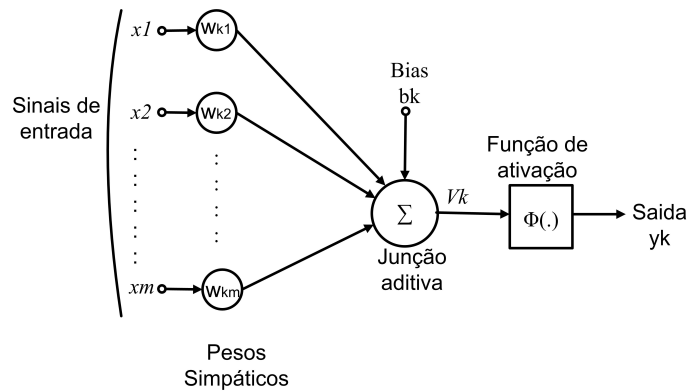


Figura 2.4: Neurônio artificial [Haykin 2001].

O neurônio artificial k ilustrado na Figura 2.4 pode ser descrito pelas Equações 2-2 e 2-3:

$$u_k = \sum_{j=1}^m w_{kj} x_j \quad (2-2)$$

$$y_k = \varphi(u_k + b_k) \quad (2-3)$$

onde:

$x_1, x_2, \dots, x_m$ são os sinais de entrada;

$w_{k1}, w_{k2}, \dots, w_{km}$ são os pesos simpáticos do neurônio k ;

u_k é saída do combinador linear;
 b_k é o bias;
 $\varphi(.)$ é a função de ativação;
 y_k é a sinal de saída do neurônio.

A **Função de ativação** representada por $\varphi(.)$ na Figura 2.4 define a saída do neurônio k em função do campo local induzido v . O funcionamento do neurônio envolve a soma ponderada dos sinais de entrada aplicados ao sinal de saída (função de ativação). As funções de ativação mais conhecidas são:

1) Função de limiar: apresentam muitas vantagens para o treinamento de redes recorrentes. Seu uso reduz o esforço computacional durante o treinamento [Fausett 1994]. A Equação 2-4 corresponde a saída do neurônio k com função de ativação limiar.

$$y_k = \begin{cases} 1 & \text{se } v_k \geq 0 \\ 0 & \text{se } v_k < 0 \end{cases} \quad (2-4)$$

A equação 2-5 expressa o campo local induzido (v_k) do neurônio k .

$$v_k = \sum_{j=1}^m w_{kj}x_j + b_k \quad (2-5)$$

A Figura 2.5 descreve a função de ativação limiar do neurônio de McCulloch-Pitts, onde a saída do neurônio assume valor 1 se campo local induzido é não-nulo e 0 quando é nulo [Haykin 2001].

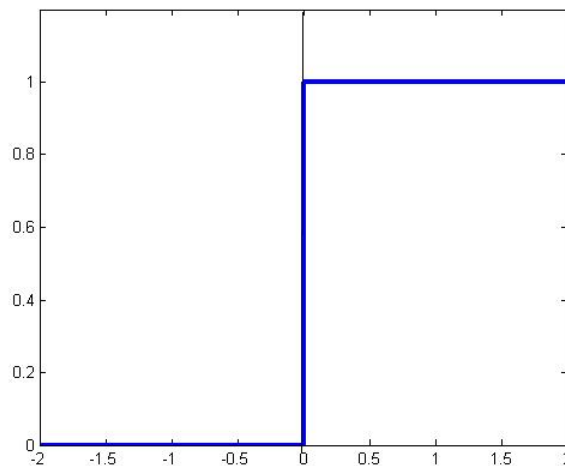


Figura 2.5: Função de limiar.

2) Função Sigmóide: é a função de ativação mais utilizada pelas redes neurais e seu gráfico apresenta a forma de s. Segundo Haykin [Haykin 2001]

ela é definida como função estritamente crescente e fornece equilíbrio entre o comportamento não - linear e linear. A Figura 2.6 ilustra a função de sigmóide com parâmetro de inclinação a variável.

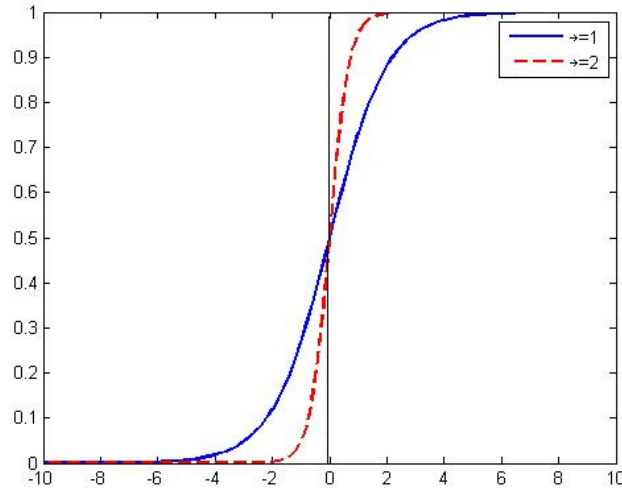


Figura 2.6: Função sigmóide.

A função de ativação sigmóide pode ser definida pela Equação 2-6, onde a é o parâmetro de inclinação.

$$y_k = \frac{1}{1 + \exp(-ax)} \quad (2-6)$$

Variando o parâmetro de inclinação é possível encontrar funções sigmóides com diferentes inclinações. A função de sigmóide apresenta intervalos de valores contínuos entre 0 e 1 conforme ilustra a Figura 2.6 [Haykin 2001]. Para intervalos de ativação entre -1 e 1 é usada a função tangente hiperbólica, definida por:

$$h(x) = \frac{1 - \exp(-2x)}{1 + \exp(-2x)} \quad (2-7)$$

Outra função de ativação citada em Haykin [Haykin 2001] é a limiar por partes. Ela é vista como aproximação do amplificador não-linear.

2.2.4 Arquiteturas das Redes Neurais

O arranjo dos neurônios nas camadas e os padrões de conexão com outras camadas é definido como arquitetura da rede. As redes neurais artificiais podem ser classificadas em redes de única camada e de múltiplas camadas.

Essas redes neurais podem apresentar realimentação, ou seja, os neurônios de saída podem fornecer sinais para a entrada da rede [Fausett 1994].

As **redes alimentadas a diante** são a forma mais simples de redes neurais artificiais. Elas são divididas em uma camada de entrada e outra de saída. É importante ressaltar que na camada de entrada não é realizado nenhum processamento, portanto ela não é considerada uma camada da rede [Haykin 2001]. A Figura 2.7 ilustra a rede neural com única camada.

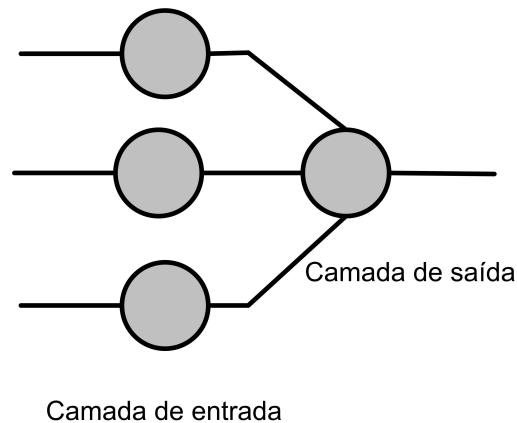


Figura 2.7: Rede neural artificial com única camada.

As **redes alimentadas a diante com múltiplas camadas** possuem uma ou mais camadas ocultas. Os neurônios ocultos se localizam entre a camada de entrada e a de saída. Os conjuntos de sinais nos neurônios de saída são a resposta global da rede para determinadas entradas [Haykin 2001]. A Figura 2.8 ilustra rede alimentada a diante com uma camada de neurônios ocultos.

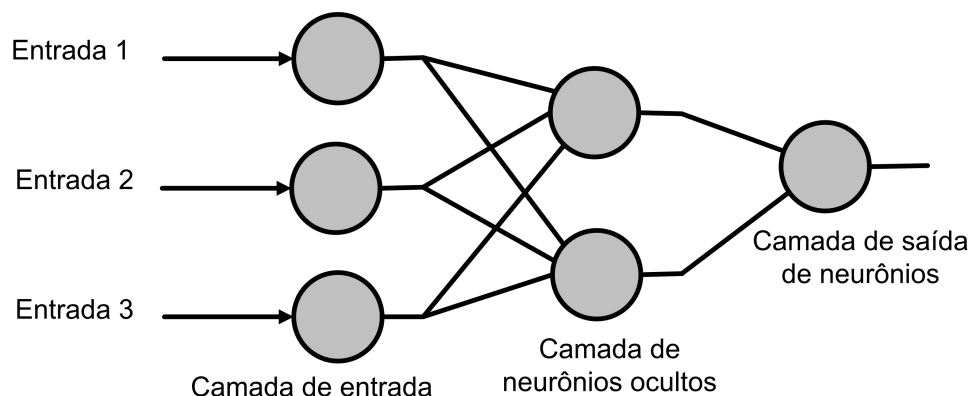


Figura 2.8: Rede neural artificial com múltiplas camadas.

As **redes neurais recorrentes** possuem pelo menos um laço de realimentação em sua estrutura. Esses laços influenciam na capacidade de

aprendizado e desempenho da rede [Fausett 1994]. Os laços de realimentação são representado na Figura 2.9 por Z^{-1} .

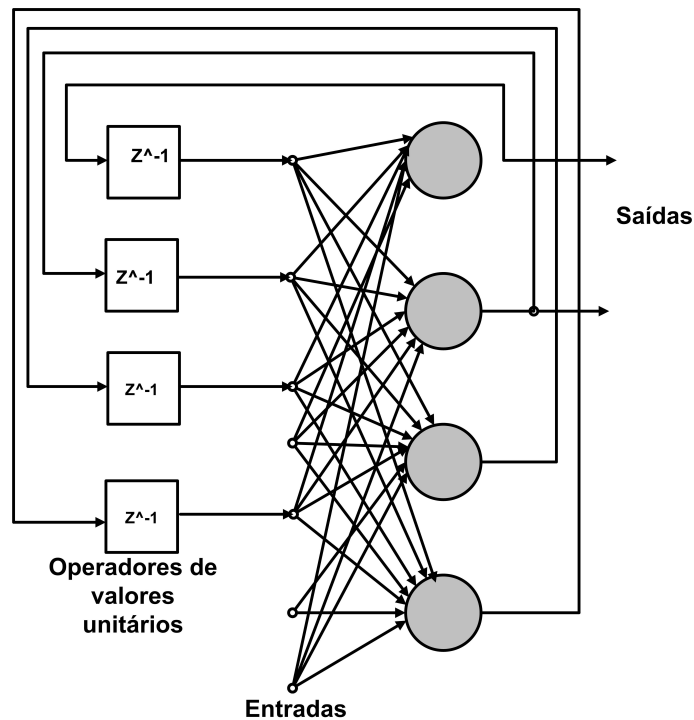


Figura 2.9: Rede neural artificial recorrentes [Haykin 2001].

2.2.5 Processos de Aprendizagem

O processo de aprendizagem é a habilidade de aprender a partir de exemplos do ambiente com objetivo de melhorar o desempenho da rede neural. Segundo Haykin [Haykin 2001] aprendizagem é o processo onde os parâmetros livres da rede neural são adaptados através de estímulos do ambiente. As regras que definem a forma de aprendizado das redes neurais são denominados algoritmos de aprendizagem. Fausett [Fausett 1994] define dois tipos de treinamento (aprendizagem): supervisionado e não-supervisionado.

Durante a **aprendizagem supervisionada** a rede neural recebe do professor ⁴ o conjunto de exemplos de entradas e saídas desejadas. Os parâmetros (pesos) da rede neural são ajustados de acordo com o sinal de erro e os padrões de treinamento. O sinal de erro é definido como a diferença entre o saída desejada e a resposta da rede. O processo é realizado diversas vezes até que o sinal de erro seja minimizado (se aproxime de zero) [Haykin 2001].

⁴Fornecer a resposta desejada para o vetor de treinamento.

Na **aprendizagem não-supervisionada** é desnecessária a figura do professor indicando a resposta desejada para os padrões de entrada. Após o ajuste das regularidades estatísticas dos dados de entrada a rede neural desenvolve a capacidade de formar representações internas para codificar as características da entrada [Haykin 2001].

2.2.6 Perceptrons

O perceptron é a forma mais simples de redes neurais usado para classificar dados linearmente separáveis ⁵. Ele é composto de único neurônio com pesos sinápticos ajustáveis e bias. O perceptron é dividido em unidades sensoriais, associativas e resposta [Haykin 2001]. A Figura 2.10 ilustra o perceptron de única camada com três entradas e bias.

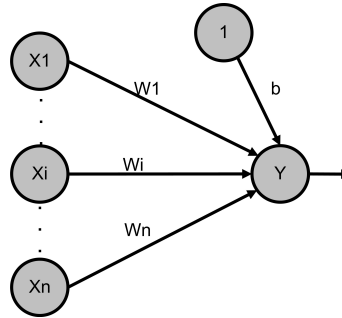


Figura 2.10: Perceptron de única camada [Fausett 1994].

No perceptron da Figura 2.10 os pesos sinápticos são representados por $w_1, \dots, w_i, \dots, w_n$; as entradas são representadas por $x_1, \dots, x_i, \dots, x_n$; e o bias é representado por b . A Equação 2-8 ilustra a função de ativação do neurônio Y com resposta variando de -1 a +1. Onde, y_{in} é o campo local induzido no neurônio Y :

$$Y = \begin{cases} 1 & \text{se } y_{in} > 0 \\ 0 & \text{se } -0 \leq y_{in} \leq 0 \\ -1 & \text{se } y_{in} < -0 \end{cases} \quad (2-8)$$

$$y_{in} = \sum_{i=1}^m w_i x_i + b \quad (2-9)$$

Durante o treinamento é calculado uma saída para cada entrada da rede. O teorema de convergência do perceptron altera os valores dos pesos para que a rede neural responda de forma correta todos os padrões de

⁵Padrões se encontram em lados opostos no hiperplano.

treinamento. Os pesos sinápticos são encontrados depois de um número finito de etapas de treinamento [Fausett 1994].

2.2.7 Perceptrons de Múltiplas Camadas

As redes neurais com única camada apresentam limitações na representação de padrões de treinamento, o que dificulta seu uso em problemas complexos. Em 1986, Rumelhart, Hilton e Williams demonstraram que é possível treinar redes neurais com camadas internas. O resultado ficou conhecido como redes neurais com múltiplas camadas [Haykin 2001].

As redes neurais com múltiplas camadas são compostas por uma camada de entrada, uma ou várias camadas ocultas e uma camada de neurônios de saída. Os **Perceptrons de Múltiplas Camadas - MLP** (do inglês, *Multilayer Perceptron*) apresentam resultados satisfatórios na solução de problemas complexos usando treinamento supervisionado com algoritmo de retropropagação do erro [Haykin 2001]. A Figura 2.11 ilustra a rede neural com múltiplas camadas.

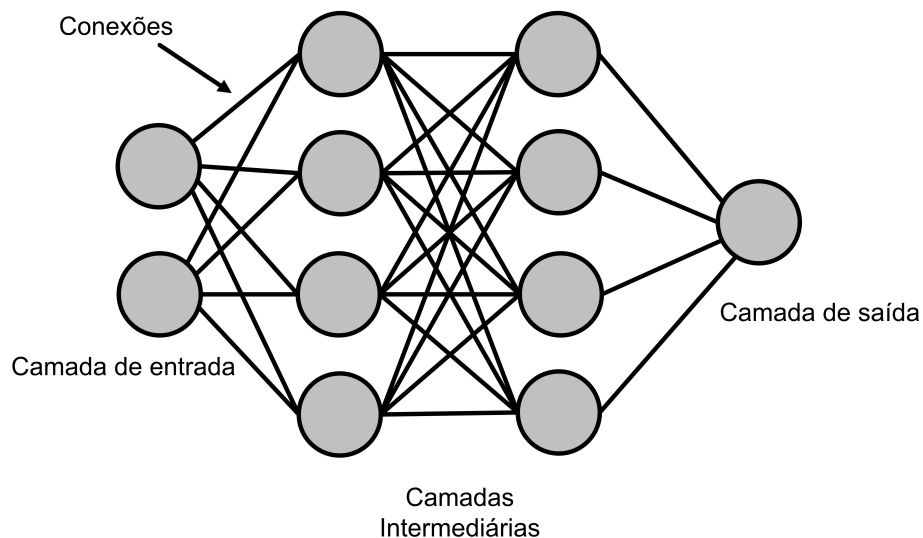


Figura 2.11: *Perceptrons de múltiplas camadas.*

Na rede neural ilustrada na Figura 2.11 cada camada possui função específica. A **camada de saída** recebe estímulos da camada intermediária e monta o padrão que será resposta da rede. As **camadas intermediárias** retiram as características, seus pesos são codificações de características apresentadas na **camada de entrada** e permitem que a rede elabore uma representação mais eficiente do problema [Haykin 2001].

O poder computacional das redes neurais MLP é derivado juntamente com a habilidade de aprender, da combinação de três características:

1) As funções de ativação tipicamente utilizadas em redes MLP são não-lineares e a mais usada é a função logística sigmóide ilustrada na Equação 2-10.

$$y_j = \frac{1}{1 + \exp(-v_j)} \quad (2-10)$$

onde:

v_j = campo local induzido;

y_j = saída do neurônio j .

2) A rede possui uma ou várias camadas ocultas que não fazem parte da camada de entrada ou de saída e capacitam a rede a aprender tarefas complexas.

3) A rede apresenta alto grau de conectividade, determinado pelas sinapses da rede. Uma mudança na conectividade exige alteração nos pesos e conexões [Haykin 2001].

Existem diversos algoritmos para o treinamento de redes MLP. Grande parte deles é do tipo aprendizado supervisionado. As redes MLP são muito usadas para solucionar problemas complexos utilizando treinamento supervisionado com o algoritmo de retropropagação. O treinamento das redes MLP com retropropagação é dividido em três etapas: **propagação** (são inseridos os padrões de treinamento); **retropropagação** do erro (diferença entre a saída desejada e saída real da rede); e **ajuste dos pesos**. Essas etapas são repetidas diversas vezes durante o treinamento da rede [Fausett 1994]. O funcionamento do algoritmo de retropropagação é ilustrado na Figura 2.12, as linhas contínuas representam a propagação e as linhas tracejadas representam a retropropagação.

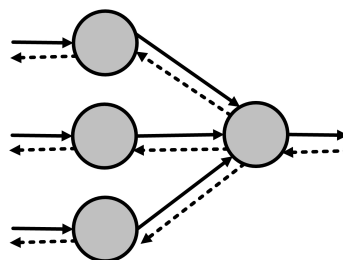


Figura 2.12: Retropropagação [Haykin 2001].

Na fase de propagação os parâmetros de entrada são apresentados a camada de entrada da rede. Em cada valor de entrada é aplicado o peso sináptico do neurônio ou neurônios da camada seguinte, produzindo determinado efeito. Esse procedimento é repetido nas camadas ocultas da rede até a camada de saída onde é gerado a resposta da rede para os padrões de entrada

inseridos na rede. Esse procedimento acontece na fase de treinamento e uso da rede.

Durante a fase de treinamento a saída da rede é comparada a saída desejada. A resposta da rede é subtraída da resposta desejada, produzindo o sinal de erro. Se o erro for considerado grande o algoritmo de retropropagação realiza o ajuste dos pesos. A Figura 2.13 ilustra o algoritmo de retropropagação.

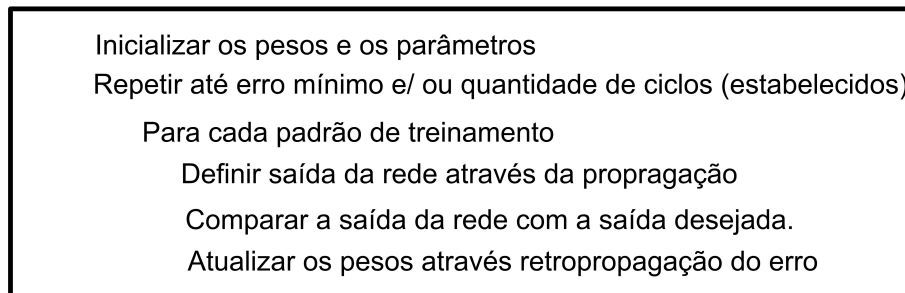


Figura 2.13: *Algoritmo de retropropagação.*

Na fase de treinamento é necessário definir de forma adequada o critério de parada. A amostra de dados usada durante o treinamento deve ser dividida em dados de treinamento e validação. Os dados de treinamento são usados para treinar a rede e os de validação são usados para verificar a eficiência do treinamento, ou seja, a capacidade de generalização da rede. Três fatores influenciam na generalização da rede: 1) a quantidade de parâmetros usados no treinamento que representa o ambiente; 2) a arquitetura da rede; 3) a complexidade do problema [Haykin 2001].

O uso de redes neurais artificiais neste trabalho é justificado por sua capacidade de generalização e aprendizado. As redes MLP apresentam bons desempenhos em sistemas tutores inteligentes. O trabalho de Melo [Melo 2003] usou Redes Neurais Artificiais MLP e resultados considerados satisfatórios foram encontrados. A rede neural MLP foi usada nesta pesquisa para controlar a quantidade de exercícios propostos ao aluno⁶.

2.3 Sistemas Tutores

O processo de ensino e aprendizagem é objeto de estudo de muitas pesquisas. O método de ensino mais antigo e utilizado até hoje é denominado "presencial". Ele é caracterizado pela presença do professor e alunos no

⁶Maiores informações vide capítulo 3