

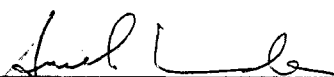
Análise Sintática Baseada em Regras

Usando um Modelo Conexcionista: Uma Visão Crítica

Márcia Aparecida Fernandes

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

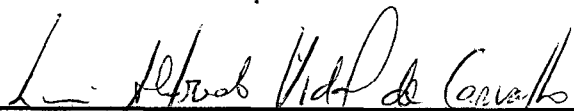
Aprovada por:



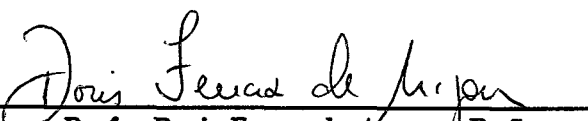
Prof. Sueli Bandeira Teixeira Mendes, Ph. D.
(presidente)



Prof. Valmir Carneiro Barbosa, Ph. D.



Prof. Luís Alfredo Vidal de Carvalho, D. Sc.



Prof. Doris Ferraz de Aragon, D. Sc.

RIO DE JANEIRO, RJ - BRASIL
SETEMBRO DE 1989

FERNANDES, MÁRCIA APARECIDA

Análise Sintática Baseada em Regras Usando um Modelo Conexionista:

Uma Visão Crítica [Rio de Janeiro] 1989

**X, 90 pp., 29.7 cm, (COPPE/UFRJ, M. Sc., ENGENHARIA DE SISTEMAS
E COMPUTAÇÃO, 1989)**

TESE - Universidade Federal do Rio de Janeiro, COPPE

1 - Linguagem Natural 2 - Análise Sintática 3 - Conexionismo

4 - Redes Neurais I. COPPE/UFRJ II. Título (Série).

À memória de minha mãe

Agradecimentos

Agradeço aos meus pais, familiares e amigos que mesmo distantes estiveram sempre presentes através do carinho e do apoio que me dedicaram.

Expresso a minha profunda gratidão aos professores Sérgio de Mello Schneider, Sandra e Victor Bojarczuk, Antônio Costa Pereira e Gilberto Arantes Carrijo da Universidade Federal de Uberlândia. Alguns pela compreensão e críticas pessoais e outros pelo estímulo e orientação.

Indistintamente, agradeço a todos os colegas que durante o período deste trabalho permitiram um ambiente agradável e facilitaram a minha locomoção pela cidade. Também as conversas sempre tão enriquecedoras que me farão pensar por um longo tempo. Aos funcionários e técnicos do Programa expresso meu reconhecimento pela atenção e prestatividade. Também aos monitores e consultores do grupo IBM, instalado no Núcleo de Computação Eletrônica, NCE/UFRJ, agradeço a cordial disponibilidade.

A toda família Inácio da Silva que me recebeu e ofereceu os mais diversos auxílios e especialmente um lar, registro o meu mais sincero agradecimento. E também a outras pessoas que pelo convívio diário, favoreceram a minha adaptação neste período.

Por fim, agradecimentos a toda orientação recebida em vários níveis de professores da COPPE. Em especial, à professora Sueli B. T. Mendes (orientadora), ao professor Valmir Carneiro Barbosa, à professora Leila Eizirik, todos da COPPE/Sistemas e ao professor Luis Alfredo Vidal de Carvalho da COPPE/Mecânica.

Resumo da Tese apresentada à COPPE como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M. Sc.)

Análise Sintática Baseada em Regras
Usando um Modelo Conexionista:
Uma Visão Crítica

Márcia Aparecida Fernandes

Setembro de 1989

Orientadora: Sueli Bandeira Teixeira Mendes

Programa: Engenharia de Sistemas e Computação

Através de um modelo para análise sintática baseada em regras são analisadas as influências da abordagem conexionista no Processamento das Linguagens Naturais. Neste modelo, pretende-se formular regras gerais para síntese de uma rede construída segundo as regras de uma gramática livre de contexto. Utilizando um esquema computacional similar ao da Máquina de Boltzmann para o processamento da rede, deseja-se obter as árvores de derivação sintáticas das sentenças da gramática. Além disso, pretende-se retirar ambigüidades entre substantivo e verbo.

Nesta tese, é visto que isto nem sempre ocorrerá, pois a síntese da rede que é proposta não permite que a Máquina de Boltzmann funcione bem. Assim, a retirada de ambigüidade poderá falhar em alguns casos e também para algumas entradas corretas.

Com esta análise é possível ver que o procedimento conexionista ainda não pode resolver os fenômenos da cognição humana. Contudo, conexionismo talvez possa oferecer melhores recursos para a compreensão dos processos cognitivos, desde que tenta reproduzir a atividade mental humana.

Abstract of Thesis presented to COPPE as partial fulfillment of the requirements for the degree of Master of Science (M. Sc.)

Rule-based Parsing
Using a Connectionist Model:
A Critic View

Márcia Aparecida Fernandes

September, 1989

Thesis Supervisor: Sueli Bandeira Teixeira Mendes

Department: Systems Engineering and Computer Science

The influences of connectionist theory in the Natural Language Processing is evaluated through a model of rule-based parsing. The aims of this proposal is to formulate general rules for setting weights and thresholds in a network to parsing, whose topology is based on context-free grammar rules. So, using the computational scheme similar to Boltzmann Machine is intended to obtain correct parse trees of the grammar sentences and disambiguate words that can be noun or verb.

In this thesis, it is shown this will not always occur due to the setting weights and thresholds that sometimes do not allow the successful processament of the Boltzmann Machine. Consequently, the disambiguation can fail in some cases.

This analysis shows that the connectionist approach can not yet solve all the cognition phenomena. However, maybe the connectionism can offer best conclusions, since it is based on biologic properties of brain and this is closer to reality.

Índice

Capítulo I - INTRODUÇÃO	1
Capítulo II - O PROCESSAMENTO DAS LINGUAGENS NATURAIS (PLN)	3
II.1 - A Abordagem Simbolista	3
II.2 - Evolução e Sistemas Clássicos	7
II.3 - A Análise Sintática	17
II.3.1 - Revisão Geral	17
II.3.2 - Alguns Analisadores	23
II.3.3 - Ambigüidades	26
Capítulo III - CONEXIONISMO E PLN	29
III.1 - Conexionismo: Aspectos Gerais	29
III.2 - Evolução e Principais Modelos	33
III.3 - A Nova Perspectiva para PLN	46
III.4 - Primeiros Modelos	48
Capítulo IV - O MODELO	54
IV.1 - Projeto da Rede	54
IV.1.1 - Topologia	55
IV.1.2 - Síntese	61
IV.2 - Processamento	69

Capítulo V - CONCLUSÕES	76
REFERÊNCIAS BIBLIOGRÁFICAS	79
APÊNDICE	85

Lista de Figuras

II.1 Modelo Geral segundo Abordagem Simbolista.	12
II.2 Gramática Livre de Contexto.	19
II.3 Árvore de derivação sintática.	20
II.4 Gramática de Cláusulas Definidas.	21
II.5 Uma rede de Transição. SV é o estado inicial e SV1, SV2 e SN2 os finais.	22
III.1 Algoritmo para a Máquina de Boltzmann.	44
III.2 Um modelo para um sistema conexionista.	49
IV.1 Primitivas conexionistas com duas (a) e três (b) unidades.(c) simplificação de (b).	56
IV.2 (a) Compartilhamento para as primitivas da figura IV.1 e (b) sua simplificação.	57
IV.3 Um compartilhamento para as derivações da gramática da figura II.2.	58
IV.4 A rede construída para gramática da figura II.2.	60
IV.5 Atribuições de pesos às ligações excitatórias.	62
IV.6 Quatro grupos básicos.	63
IV.7 Configurações do grupo (a) da figura IV.6.	64
IV.8 Configurações do grupo (c) da figura IV.6.	66
IV.9 Configurações do grupo (d) da figura IV.6.	68

IV.10 Contra-exemplo para o modelo. 73

IV.11 Gramática para o contra-exemplo. 74

Capítulo I

Introdução

A forma simbólica e seqüencial com que Inteligência Artificial e Ciência Cognitiva evoluíram, durante estes quarenta anos, vem sendo agora questionada tanto teorica como computacionalmente. Para mostrar isto, faz-se no capítulo II, uma revisão geral do Processamento e Compreensão das Linguagens Naturais, analisando as bases teóricas da *Ciência Cognitiva* em que se apoiou até hoje e alguns modelos dessa visão clássica. Também são revistos os métodos e as estratégias que delinearam o processamento sintático dentro desta abordagem. Em particular, as *Gramáticas de Cláusulas Definidas* e as *Redes de Transição Aumentada*, que de forma eficiente, tratam esta parte do processamento de linguagens de acordo com a abordagem simbolista.

Este questionamento tem incentivado o crescimento de uma linha de pesquisa quase totalmente abandonada anteriormente, a abordagem *conexionista*. Esta propõe uma aproximação ao comportamento humano para realizar tarefas cognitivas, através de um sistema de unidades de processamento, similares ao neurônio humano. Atualmente, com avanços teóricos e tecnológicos, o conexionismo tem introduzido vários modelos para a diversificada área da cognição humana.

No capítulo III, são apresentadas e discutidas as principais características e modelos da linha conexionista. Em especial, o modelo que serviu de base para o sistema que será apresentado. A união Conexionismo e Processamento das Linguagens Naturais é o assunto das últimas seções deste capítulo, com uma análise de seus primeiros modelos.

No capítulo IV é apresentada uma crítica ao modelo de uma rede neuronal para análise sintática baseada em regras segundo a abordagem conexionista de acordo com SELMAN [46]. Este descreve um sistema que reconhece a estrutura sintática dos constituintes de uma sentença, sendo que as regras estão baseadas numa gramática livre de contexto. O objetivo principal é mostrar a possibilidade de processamento de regras utilizando um modelo conexionista. É ainda objetivo do modelo retirar alguma ambigüidade neste nível. Neste modelo, utilizam-se duas técnicas, a Máquina de Boltzmann e o "Simulated Annealing". De algumas características da primeira tenta extrair regras gerais para síntese da rede neuronal que constrói. No entanto, através de exemplos e de um contra-exemplo, é mostrado que seu modelo não tem o alcance que sugere. Isto pode também ser visto numa revisão apresentada em SELMAN et. alia. [55].

Uma das principais discussões, existente não só entre as duas linhas, mas também entre os seguidores da linha conexionista, é a utilização do processamento explícito de regras. Por exemplo, COTTRELL et. alia. [36,51] não utiliza uma árvore de derivação sintática em seu modelo. WALTZ et. alia. [52], por outro lado, utiliza. McCLELLAND et. alia. [49], no entanto, diz que regras deveriam ser uma consequência dos processos de aprendizagem. Por isso, é objetivo desta tese, avaliar o grau de necessidade e o desempenho de um analisador sintático baseado em regras para modelagem do Processamento das Linguagens Naturais, segundo a abordagem conexionista.

Além disso, discute-se aqui o uso do seqüenciamento para a entrada das sentenças num sistema baseado em modelos conexionistas. No capítulo V são vistas as conclusões deste trabalho.

Capítulo II

O Processamento das Linguagens Naturais

II.1 A Abordagem Simbolista

Nas últimas décadas, desde o advento do computador, pesquisas direcionadas à compreensão da mente humana têm sido altamente estimuladas. Este, que não é um interesse novo para o homem, tem atualmente atingido o seu auge. No decurso deste tempo, pesquisadores passaram de um paradigma a outro tentando descobrir e explicar os princípios básicos de funcionamento da mente e do cérebro humanos. A base destas pesquisas são os processos mentais que possibilitam aos humanos exibirem habilidades ou capacidades ligadas à inteligência.

A forma com que humanos utilizam estas habilidades, tais como, compreensão de linguagem e imagem, memória, percepção, pensamento e visão é tão simples e natural para qualquer observador comum, pode-se até dizer que é quase inconsciente, que à primeira vista é possível supor a inexistência de complexidade nos processos mentais necessários a quaisquer destas habilidades. No entanto, a dificuldade em explicá-los e compreendê-los reside justamente nestas simplicidade e naturalidade pois, apesar disso, é evidente que exigem um alto grau de abstração, conhecimento e aprendizagem, características que os tornam não observáveis e que permitem denominá-los *processos cognitivos*. Assim, quase

toda a pesquisa desenvolvida à respeito da mente humana está baseada em hipóteses.

Muitas são as áreas da ciência envolvidas nesta pesquisa, mas a Filosofia e a Psicologia são obviamente duas delas. A primeira por ter como seu espectro de interesse todo o conhecimento humano, organizá-lo é seu objetivo. A segunda, por extrair suas teorias de observações do comportamento humano através de experiências que testam habilidades e reações humanas, como as acima citadas. Por isso, pode-se dizer que constituem a base teórica para as pesquisas dos processos cognitivos. Assim, os pesquisadores destas áreas buscam respostas para questões que de um ponto de vista geral poderiam ser: "Como são os processos pelos quais passa a mente humana quando faz uso da linguagem, visão, percepção, pensamento e memória?", "Como são organizados e quais são os conhecimentos necessários para exibir tais habilidades?".

Com o desejo de unificar os esforços relativos a estas e outras ciências, alguns pesquisadores postularam fundamentos básicos e hipotéticos para orientar as diferentes áreas de pesquisa. STILLINGS et alli. [1] discutem quatro hipóteses em que se apóia tal unificação, uma ciência interdisciplinar chamada *Ciência Cognitiva*. Sendo esta uma das hipóteses, pois da Psicologia absorvem a idéia de que a mente humana é um poderoso sistema de processamento da informação por receber, acumular, recuperar, transformar e transmitir informação. Então, seria possível uma ciência básica do processamento da informação.

A posição assumida por filósofos como Descartes e Hobbes, também apresentada em [2], de que a compreensão consiste em formar e manipular representações do conhecimento, símbolos relacionados a objetos do mundo real, mas desprovidos de significado, é a segunda e talvez mais forte hipótese adotada. No entanto, uma teoria para tratamento de significados não é descartada, pois se não fosse possível retirar alguma significação destas relações, elas seriam inúteis. Dessa maneira, identificando a informação como um símbolo, é possível traduzir os processos cognitivos em algoritmos, que supõem um conjunto de operações formais, que podem adquirir formas variadas e ainda assim manter as relações de representação entre os símbolos. Tem-se então a terceira hipótese.

A quarta e última diz que é dispensável conhecer os sistemas que suportam tais algoritmos, estes devem ser definidos independentemente de plausibilidades neuro-fisiológicas. Portanto, o que foi exposto pode ser generalizado no que é conhecido como

abordagem simbolista, devido principalmente à segunda hipótese. Este é o paradigma no qual toda a pesquisa teórica e pragmática sobre processos cognitivos que, neste caso podem ser interpretados como processos da informação, tem-se desenvolvido.

Além da Psicologia e da Filosofia, como mencionado, o advento do computador acelerou as pesquisas nesta área. A Ciência da Computação, essencialmente pragmática, também ofereceu suporte teórico às formulações da Ciência Cognitiva.

Através das hipóteses nota-se a profunda relação destas com os princípios gerais de organização do computador. Esta máquina é vista como um sistema de processamento da informação, que executa algoritmos expressos em operações formais sobre símbolos, preservando as relações representacionais entre eles. Reforçando as três primeiras hipóteses. Além disso, uma implicação que é mais geral é que o computador se mostrou capaz de tarefas impossíveis para humanos, porém (ainda?) não é capaz de outras, isto justifica a quarta hipótese, que conhecer o sistema, quer seja físico ou biológico, não é relevante, pois os cientistas cognitivistas, adeptos desta abordagem, não acreditam haver características importantes compartilhadas por cérebro e computador.

De um ponto de vista mais específico, a área de Inteligência Artificial, subdivisão da Ciência da Computação, desenvolve pesquisa sobre processos cognitivos, visando resultados práticos efetivos. Seu objetivo principal é adaptar a máquina ao homem pela construção de sistemas inteligentes, isto é, sistemas que inspirados no comportamento humano realizem as habilidades citadas. Também é esperado que exibam aprendizagem, ou seja, capacidade de atuarem frente a situações não previstas. Por isso, se utilizam das teorias psicológicas, hipoteticamente estabelecidas tanto para fornecer subsídios ao desenvolvimento da Ciência Cognitiva, através da verificação que viabiliza, como para atingir seu próprio objetivo.

Por este fato alguns a consideram interdisciplinar. Mas, MINSKY em [3] refuta esta colocação argumentando que apesar de extrair métodos de outras áreas, a Inteligência Artificial tem seus próprios conceitos, técnicas e jargões que progressivamente formam uma especialidade organizada. Na breve evolução histórica que apresenta entre 1955 e 1962 constata-se que seu ponto de vista está fundamentado na postura formal e matemática dos primeiros trabalhos, no pouco desenvolvimento das teorias psicológicas e na idéia de ser esta área uma tentativa de construir máquinas inteligentes sem questionar se os sistemas

eram simples, biológicos ou humanóides. Contudo, será visto que é crescente a influência das ciências que investigam o comportamento humano nos trabalhos que se seguiram. Assim, a definição de Inteligência Artificial hoje caminha para aquela dada anteriormente.

Em meados dos anos 50, a força do computador tinha sido testada para sistemas não-numéricos como jogos, redes neuronais e transformações de figura. Mas, a programação séria de processos de manipulação de símbolos consistia em experimentos para tradução de linguagem. Os métodos empregados eram heurísticos, isso garantia eficiência devido à busca exaustiva (tentativa-e-erro) mas, era inviável para problemas grandes e difíceis. O computador nesta época foi visto como rápido mas estúpido, comprometendo os programas heurísticos.

Por volta de 1961, havia bons programas para resolver provas de teoremas e alguns jogos, vários apresentavam formas simples de aprendizagem. No entanto, nenhum manuseava bem linguagem natural ou desenvolvia conceitos de mais alto nível. Um argumento que logo se mostrou errôneo dizia que isso realmente não seria possível. Sendo, o computador uma máquina "lógica" era bom apenas para problemas formais e lógicos. Basicamente, antes de 1962, a idéia central era encontrar dispositivos heurísticos para controlar a largura da busca exaustiva e técnicas de aprendizagem. Após 1962 a preocupação se concentrava em representação do conhecimento.

O que se deu depois desta época e o que se tem ainda não são soluções gerais e abrangentes para estes e outros problemas da cognição. No entanto, avanços significativos em todas as áreas da ciência relacionadas a processos cognitivos, especialmente Psicologia e Computação, permitiram que algumas alternativas para tratamento destes problemas fossem realizadas. Os métodos heurísticos não foram completamente abandonados e em alguns casos de aplicação numérica, como jogos, tornaram-se mais eficientes. Aplicações não-numéricas cresceram consideravelmente, um exemplo é o Processamento das Línguas Naturais (PLN). Porém, alguns acreditam que o paradigma simbolista não pode responder satisfatoriamente questões como visão, audição e PLN.

II.2 Evolução e Sistemas Clássicos

Dentre as habilidades cognitivas mencionadas, a compreensão e o uso de linguagem natural foi a que mais estimulou pesquisas sobre processos cognitivos. Houve correntes firmadas na crença de que linguagem e pensamento eram inseparáveis ou que esta permearia o comportamento e desenvolvimento humanos. É também considerada uma das linhas delimitadoras entre humano e não-humano, ou seja, caracteriza o ser dotado de inteligência. Teoricamente, o estudo desta habilidade conta em particular com estudiosos da Lingüística, empenhados na descrição formal da língua[†], e da Psicolingüística [4], interessados pelos processos mentais relacionados ao uso e a aprendizagem de linguagem.

Sob o ponto de vista prático, PLN é a subárea da Inteligência Artificial que investiga os processos cognitivos relacionados à linguagem. Também a motivação tecnológica sustenta o seu desenvolvimento, pois com o objetivo de adaptar a máquina ao homem, a construção de sistemas para tradução automática de máquina ou Tradução Mecânica, análise de textos, compreensão de fala (diálogo), aprendizagem auxiliada por computador e "interface" para linguagem natural, vêm sendo largamente incentivada ALLEN [5].

Entre esta estreita relação da Ciência Cognitiva com Inteligência Artificial, em particular, sob o ponto de vista do tratamento de linguagem natural, se encontra a Lingüística Computacional. Os lingüistas computacionais pretendem criar sistemas que manuseem linguagem semelhantemente a humanos. Através da análise evolutiva desta área dada em SCHANK [6] e dos paradigmas da Lingüística apresentados em WINOGRAD [7], é possível traçar os pontos marcantes que influenciaram a evolução de PLN em seus primeiros anos.

Dois destes paradigmas, *Gramática Prescritiva* e *Lingüística Comparativa*, em ordem cronológica, faziam analogias da Lingüística com a Advocacia e com a Biologia, respectivamente. Mas, foram refutados, não sendo inclusive considerados paradigmas, e sim metáforas por tratarem da organização social da ciência e não de suas formas conceituais.

A preocupação da primeira metáfora estava na estrutura correta e pura da lin-

[†] Há divergências quanto ao uso das palavras "linguagem" e "língua". Os teóricos concordam que a segunda é a correta.

guagem. O ensino da gramática refletia as regras de bom comportamento impostas pela sociedade. Porém, percebeu-se que linguagem é um fenômeno natural e que está em constante mutação. A segunda se motivou na teoria da evolução de Darwin, por isso a classificação das linguagens era o seu objetivo, dadas as várias semelhanças percebidas, seria possível construir uma árvore de famílias de linguagens. Porém, esta tarefa tornou-se tediosa e repetitiva, sendo portanto, quase totalmente abandonada.

O que veio a seguir e predominou os anos 50 tinha suas bases na teoria "behaviorista" da Psicologia mas, principalmente nas técnicas experimentais da Química para descobrir estruturas de substâncias. A analogia feita pelos lingüistas incentivou o estudo da sintaxe (estrutura) de várias linguagens, e este ficou conhecido como paradigma da *Lingüística Estrutural* ou *Descritiva*.

Pouco tempo antes desta época, os lingüistas computacionais acreditavam que apenas um grande dicionário léxico seria suficiente para realizar tradução de uma língua para outra, ou seja, a Tradução de Máquina ou Mecânica. Eram as primeiras tentativas de escrever sistemas com comportamento inteligente para PLN. Muitas sentenças não podiam ser compreendidas, desde que a tradução direta nem sempre é possível ou significativa, por isso os resultados não foram promissores.

Influenciados pelo fracasso da tradução palavra a palavra e possivelmente pelo paradigma da lingüística descritiva, os lingüistas computacionais concluíram que seria necessário analisar a estrutura das sentenças. A gramática de estrutura de frase ou sintagmática foi o formalismo empregado para verificar a combinação correta dos constituintes de uma sentença. Todos os esforços se concentraram nesta direção para resolver o problema de Tradução Mecânica. Porém, esta gramática apenas permitia uma única descrição de uma frase, incluía duplicações desnecessárias, sendo bastante limitada para expressar conhecimento lingüístico.

Por volta de 1957 surge o paradigma da *Lingüística Gerativa*. Com a publicação de 'Syntatic Structures' Noam Chomsky revoluciona o estudo de Lingüística [4,7]. Refuta a teoria psicológica subjacente ao paradigma anterior dizendo que é necessário considerar a criatividade e a intuição do usuário de uma linguagem. Isto implicaria no estudo de processos mentais, o que era negado pelos "behavioristas" por não serem observáveis e passíveis de experimentação e, portanto qualquer afirmação a este respeito não poderia

ser dita científica. Assim, introduz os conceitos de *desempenho*, uso real da linguagem em situações concretas e *competência*, conhecimento que se tem da linguagem.

Apesar de toda polêmica causada apresenta ainda a teoria da *Gramática Transformacional*, também em [8], que é uma elaboração de regras mais complicadas que permitiam transformar estruturas de superfície em estruturas subjacentes mais significativas. Através destas transformações possibilita relacionar sentenças de mesmo significado, como por exemplo, a voz ativa e passiva. Esta teoria incentivou os pesquisadores de Tradução Mecânica que chegaram a crer na solução do problema. No entanto, este não foi o desfecho.

A Gramática Transformacional, enquanto uma oposição à teoria descritiva, permitia maior desenvoltura na compreensão e geração de sentenças, mas dava pouca atenção ao tratamento de significados. Chegou a incluir algum significado [9], como +humano, -animal, -contável ou +comum, os chamados traços sintáticos, que em algum sentido poderia ser pensado como inclusão de informação semântica numa entrada léxica, embora isso não fosse completamente definido e claro. Também as transformações sugeriam algum conhecimento de semântica devido ao apagamento ou substituições de alguns constituintes. Mas, ainda assim, a sintaxe era o ponto forte da teoria gerativa. O próprio Chomsky anos depois reconhece a falha da postura excessivamente sintática de sua teoria.

Eram insuficientes os resultados com Tradução Mecânica, mas parecem ter permitido alguma especulação linguística informal a respeito dos problemas de ambigüidade que se originaram a partir das substituições e transformações. Além disso MINSKY em [3], afirma que o erro estava na esperança de que análise sintática adequada poderia ser feita sem a atribuição de significado. Parece óbvio que os significados devem ser considerados para resolver ambigüidades mesmo dentro de discursos coerentes numa linguagem. Também, foi vista a necessidade de métodos de representação e capacidade de inferência lógica para fazer deduções. Vale a pena notar que psicologicamente, o modelo se mostra insatisfatório, pois não parece que humanos se preocupem em tal grau com a estrutura de uma sentença num processo de comunicação.

Possivelmente, Tradução Mecânica serviu para expor claramente os dois problemas com que lingüistas computacionais tinham que tratar, a *análise* e *geração* de linguagem. Os resultados desencorajadores obtidos mostraram que esta aplicação exigia tanto análise quanto geração, como toda aplicação que pretendesse realmente tratar lin-

guagem de forma similar a humanos. Inicialmente, a geração de linguagem era feita de forma aleatória, utilizando uma gramática gerativa. Isto permitia que sentenças agramaticais e sem significados fossem produzidas. Com a Gramática Transformacional, os traços sintáticos ajudaram a restringir a geração de sentenças agramaticais. No entanto, era necessário uma boa análise de significados para evitar as outras inadequações.

Assim, mais do que reconhecer os constituintes de uma sentença, humanos analisam significados das palavras e da sentença como um todo. Então não só um dicionário, nem só sintaxe, mas os dois e também processos para tratar significados. Dessa forma, a intenção de fazer Tradução Mecânica estanca, ressurgindo longo tempo depois [7], mas fica evidente que seria necessário mais um tipo de conhecimento, o *conhecimento semântico*. Embora, não fosse claro em que consistia uma teoria semântica, as atenções se centralizaram na tentativa de fazer análises mais profundas, envolvendo significados. No final dos anos 60 e década de 70 teorias semânticas como a Gramática de Casos e a Semântica Gerativa começaram a ser difundidas.

Na primeira, um exemplo de gramática semântica, são estabelecidas relações entre o verbo e seus casos, isto é, classificam-se os complementos de um verbo de acordo com as suas possíveis interpretações. Simplificando, estipula as relações funcionais entre os constituintes de uma sentença. FILLMORE em [10] apresenta um sistema de casos, mas como era de se supor, não supera a complexidade e a diversidade da linguagem. Poderia ser necessário para alguns verbos acrescentar outras definições de casos.

Foi uma tentativa de acomodar significado dentro da Gramática Transformacional. Em vez de tratar a questão em termos estritamente sintático como sugere sujeito e objeto, admite funções semânticas subjacentes como agente, instrumento, dativo e objetivo. No aspecto psicológico casos possuem noções de ordem mental mais profunda se comparados a sintagmas [4].

Os modelos da Semântica Gerativa que surgiram tentavam produzir estruturas semânticas subjacentes mais complexas que a Gramática de Casos [4]. Teoricamente, foi uma tentativa de transferir o foco da teoria gerativa para semântica, como é argumentado em [9], os linguistas desejavam mostrar que a semântica poderia e deveria ser vista como gerativa, e não apenas interpretativa como os transformacionalistas gerativistas a encaravam. Porém, os resultados obtidos com esta teoria mostraram ser muito abstrato o

conceito da nova estrutura subjacente que era proposta e assim, de difícil representação.

Com as teorias semânticas o problema da representação do conhecimento e significado fica exposto. As discussões envolviam a classificação do conhecimento lingüístico em procedimental (“procedural”) ou declarativo. Alguns o classificaram como procedimental, mas outros como [6] acreditavam que isso era uma má interpretação entre o que consistia conhecimento e os processos para utilizá-los, e que na verdade, conhecimento deveria ser declarativo. As controvérsias persistem e também o problema central, construir modelos de representação que evitassem ambigüidades, permitissem deduções e inferências, também a inclusão de novas representações, no caso de conhecimento ainda não adquirido pelo sistema.

Argumenta PITRAT [11], que usar conhecimento declarativo é benéfico devido às suas conveniência e eficiência. É mais fácil ter conhecimento declarativo que não indica como deve ser usado, dado que a quantidade de conhecimento é menor. Os componentes do conhecimento são independentes, pode-se então remover, adicionar ou modificá-los independentemente uns dos outros.

“Frames” e Redes Semânticas foram, e ainda são, os formalismos mais utilizados. Estes são modelos que tratam o conhecimento de forma declarativa, pois as inferências ou deduções são extraídas utilizando processos que não estão implícitos nestas formas representacionais. “Frames” de acordo com [5] são simplesmente um conjunto de fatos e objetos que descrevem alguma situação típica, adicionando estratégias de inferências específicas para o raciocínio sobre a situação. Redes Semânticas, é uma rede de hierarquias, onde os nós representam os conceitos ou objetos e as suas relações são arcos rotulados. Uma rede assim construída pode ser atravessada de forma ascendente ou descendente, com a primeira encontram-se os conceitos mais gerais e a outra os mais específicos.

Um outro aspecto a se considerar, que é inconveniente no uso de tais formalismos, é a dificuldade de representar conhecimento de forma completa. A escolha das características dos objetos impõe esta limitação. No caso de “Frames”, o conjunto de descritores para os objetos ou conceitos e nas Redes Semânticas, os rótulos para os arcos, que em geral, são identificados como *é-um*, *parte-de*, *feito-de*, dependendo da criatividade do construtor.

Considerações lingüísticas mostram haver conhecimentos além dos já citados. Porém estes são os mais explorados em aplicações de linguagem natural. Além disso, [5,6] afirmam que mais que todos estes conhecimentos, humanos utilizam o *conhecimento do mundo* para se comunicarem. Sendo este, portanto um dos principais conhecimentos para PLN. Atualmente, uma descrição geral [5] para os vários tipos de conhecimento, que é compartilhada pela maioria dos pesquisadores, que inclui, além dos já citados, outros seria:

- **Conhecimento Fonológico:** relacionado aos sons das palavras.
- **Conhecimento Morfológico:** relaciona-se com a formação das palavras a partir dos morfemas.
- **Conhecimento Sintático:** reconhece as relações entre as palavras numa sentença.
- **Conhecimento Semântico:** estabelece os significados possíveis para a sentença.
- **Conhecimento Pragmático:** como utilizar contexto para interpretação da sentença.
- **Conhecimento do Mundo:** informações necessárias para que os falantes mantenham um processo de comunicação.

A maioria dos sistemas que surgiram a partir daí eram então divididos em dois processos principais que utilizavam os conhecimentos *sintático* e *semântico*. Em [11] se diz que os primeiros utilizaram conhecimento procedimental. A análise sintática era feita



Figura II.1: Modelo Geral segundo Abordagem Simbolista.

primeiro em seguida significados eram atribuídos às estruturas extraídas da sintaxe e

possivelmente inferências eram feitas para permitir deduções, como em referências anteriores. Tais sistemas poderiam ser representados numa forma geral como na figura II.1, em CHARNIAK [12]. Nota-se a interação entre sintaxe e semântica. Isso possibilita a sintaxe rever a estrutura que forneceu à semântica, quando não fosse possível atribuir significado correto ou aceitável.

Alguns argumentaram que não era definido e nem claro, psicológica ou biologicamente, seqüencialidade na análise e geração de linguagem por humanos. Nem tão pouco a ordenação dos processos segundo este modelo. Uma grande discussão que envolve este modelo e que atualmente tem encontrado um grande número de adeptos é que análise sintática deveria ser feita de forma integrada em outro conhecimento. ALLEN [5], SCHANK [6] e CHARNIAK [12] entre outros compartilham desta idéia, o primeiro sugere um modelo diferente do apresentado na figura II.1. ALLEN argumenta que esta divisão dentro de fases é mais uma divisão conceitual que de seqüenciamento. A maioria dos sistemas atuais combinam pelo menos duas dessas fases. Por exemplo, fases de interpretação sintática e semântica para produzir um analisador sintático semanticamente dirigido, ou fases de processamento contextual e semântico, separando a fase sintática.

A questão sintaxe-semântica parece profundamente relacionada à postura da lógica em dividir nitidamente estes dois conhecimentos [13]. No entanto, em termos de linguagem natural isso não deveria ser feito, pois a lógica trabalha com domínio específico, ou seja, o micromundo, estritamente desaconselhável quando tratando linguagem natural, que não pode ser expressa amplamente por formalismos tão restritos.

De acordo com as teorias que surgiram vários sistemas foram criados para as diversas aplicações de PLN. Para se verificar a potencialidade e validade de alguns sistemas seria necessário implementação e teste, mas os exemplos a seguir serão analisados apenas através de suas descrições, mesmo porque são sistemas que se popularizaram já tendo recebido várias críticas. O objetivo é apenas mostrar como as teorias foram empregadas e qual o alcance obtido.

ELIZA [14] é um exemplo dos sistemas pergunta-resposta, que é citado não por sua teoria lingüística, que na verdade é quase inexistente, mas pelos resultados obtidos. Desenvolvido nos anos 60 no MIT, foi codificado para diversas máquinas. Basicamente, possui uma base de dados com palavras-chave associadas a especificações de saída. Assim,

dada uma sentença de entrada é pesquisada até que se encontre uma palavra-chave, sendo então gerada uma sentença de saída; caso contrário, gera sentenças vagas. De fato o que se tem é um conjunto de sentenças pré-estabelecidas para responder determinadas perguntas, não usando o sistema qualquer processo inteligente. O usuário é que deve usar sua inteligência para manter um diálogo coerente (ALLEN [5]).

Sistemas clássicos [7] como LUNAR e SHRDLU se inspiram na teoria da Gramática Transformacional, embora utilizando um formalismo diferente, as Redes de Transição Aumentada[†] [15]. Este formalismo é a base para a análise sintática dos dois sistemas, uma extensão das redes de transição recursiva para tratamento de gramáticas. A inovação que apresentam é a elevação da capacidade de análise que não é verificada na teoria gerativa de Chomsky. Utilizando esta muitos sistemas foram construídos visando melhorar este aspecto.

LUNAR[‡] é um sistema de pergunta-resposta como ELIZA. Estruturado de forma modular, contém um recurso geral para interpretação semântica. A maior motivação é explorar a interação entre sintaxe e semântica, cuja a ênfase está no uso de informação semântica para guiar a análise sintática. Inclui procedimentos para os problemas de conjunção, análise léxica e morfológica. SHRDLU [7] foi criado por WINOGRAD para simular um robô simples. Enfatiza a integração sintaxe, semântica e processos de raciocínio. Verificações de significado, envolvendo conhecimento geral do mundo e conhecimento real do robô, são espalhadas dentro do processo sintático.

No que diz respeito à teoria lingüística subjacente ao aparato computacional, tem-se MARGIE, sistema criado por SCHANK e outros pesquisadores em meados dos anos 70. Em [6] apresentam uma descrição detalhada de toda teoria e parte da implementação realizada em MLISP. Fundamentalmente, está dividido em duas partes PARAPHRASE e INFERENCE, a primeira é um processo que procura todas as interpretações (paráfrases) possíveis para uma sentença; e a segunda faz inferências para produzir uma saída em linguagem natural.

A grande contribuição deste sistema é a teoria da Dependência Conceitual empregada no mapeamento da sentença de entrada numa representação conceitual, da qual são

[†] Será vista com maiores detalhes na seção II.3.2

[‡] Em [15] é apresentada suas principais características, embora estivesse em fase inicial e nem mesmo o nome é citado.

extraídas as paráfrases. Nesta teoria, melhor que uma representação para uma palavra, esta é expressa através de características, unidades de representação. Por exemplo, 'João' seria representado por algumas unidades como 'homem', 'altura', entre outras. Assim, também seriam representados verbos, por meio de uma classificação prévia de acordo com a ação ou estado que representassem. A sintaxe, entretanto, não recebeu o tratamento habitual, a estrutura da sentença não foi considerada como na maioria dos sistemas, apenas uniria os conceitos, e nisto consistiria a dependência.

Dependência Conceitual representou um avanço significativo para lingüística computacional, por oferecer uma teoria com algum grau de consistência. Em especial, o tratamento dado a sintaxe bastante influenciado pelas teorias psicológicas. Mas, apesar de afirmar a necessidade de uma representação não ambígua, muito possivelmente isto ocorreria tanto na representação de verbos, dada que a classificação era arbitrária e incompleta, quanto de substantivos, caso o sistema tomasse grandes proporções.

A aplicação em análise de texto é exemplificada pelo sistema de QUILLIAN. Em [16] apresenta uma implementação para compreensão de linguagens, baseado em sua teoria psicológica (realizada em anos anteriores e discutida em [17]) que sugere a idéia de que humanos ativariam todas as possíveis interpretações para uma sentença simultaneamente, a de caráter mais usual seria considerada. Por isso utiliza busca em largura, que simula mecanismo de busca paralela. Entre outras vantagens, como análise sintática dirigida por semântica, o TLC ("Teachable Language Comprehender"), inclui capacidade de aprendizagem pelo sistema.

Dentre os sistemas apresentados, MARGIE e o TLC são os que mais se preocuparam em modificar a estrutura do modelo da figura II.1, através da inclusão de representações de significado mais sofisticadas. No sistema MARGIE o que se tem é uma rede semântica conectada por relações sintáticas. Em TLC é vista estrutura similar, no entanto, a sintaxe é construída separadamente.

A complexidade e o detalhamento das linguagens naturais foram gradativamente observados na evolução de PLN e nestes poucos exemplos. Em se tentar construir um sistema capaz de tal habilidade verifica-se que os conhecimentos, suas interações e representações são por demais intrincados. Há vários pormenores a considerar, ambigüidades, referências através de pronomes, quantificação, conjunções, advérbios, subcategorizações,

verbos auxiliares, entre outros.

Os sistemas construídos mostram-se limitados para capturar este emaranhado de subproblemas que surgem na análise e geração de linguagem. Alguns são bons para tratar certos tipos de ambigüidades, outros pela forma com que representam conhecimento, havendo aqueles que permitem aprendizagem. O contexto é considerado em alguns casos. Mas ainda não existe um que trabalhe com linguagem como seres humanos parecem trabalhar.

Há três fatores que contribuem para este progresso parcial de PLN. É sabido que há pouco conhecimento sobre os processos cognitivos e mente humana, apesar dos avanços de áreas intimamente ligadas a este estudo. A maioria das teorias existentes são hipóteses, isto pode implicar facilmente nas suas violações ou até mesmo na não verificação.

Também as divergências que obviamente ocorrem quando se tratando de assuntos interdisciplinares, como os casos da Ciência Cognitiva e da Inteligência Artificial. Focalizando em PLN, tem-se os lingüistas teóricos ou lingüistas, os psicolingüistas e os lingüistas computacionais. Observa-se que os primeiros estão mais inclinados a construir uma teoria geral para lingüística. Há algum tempo a hipótese da existência de uma gramática universal ou universais lingüísticos tem encontrado adeptos.

Este, no entanto, não é o objetivo dos lingüistas computacionais e psicolingüistas. Os primeiros precisam considerar o alto grau de detalhamento das linguagens naturais para obterem sistemas mais eficientes e reais. Os psicolingüistas, dentro do contexto de suas especialidades, apoiam a posição destes, desde que também pretendem explicar os fenômenos relacionados à linguagem considerando a mente humana e os processos cognitivos. Capturar as várias plausibilidades numa única teoria tem se tornado um grande gargalo.

Um último fator, que já superficialmente questionado, é a abordagem simbolista. Esta que está mais ligada a processos computacionais que aos mentais, se forem considerados aspectos biológicos, não tem conseguido desenvolver na prática teorias defendidas durante tanto tempo. Volta-se a preocupação de alguns pesquisadores nos primeiros anos de Inteligência Artificial, não seria possível tratar habilidades mentais, não formais, numa máquina lógica e formal como o computador.

II.3 A Análise Sintática

II.3.1 Revisão Geral

Como já dito, nos primeiros trabalhos de PLN, a sintaxe recebeu grande incentivo e chegou-se até a acreditar que esta resolveria todos os problemas. Passado o entusiasmo, hoje se discute qual seria realmente seu papel e sua importância para sistemas de linguagem natural. O foco desta discussão está em duas questões relacionadas à plausibilidade psicológica. Uma, “Até que ponto humanos fazem análise sintática num processo de comunicação?” e a outra, “Esta análise é feita através de regras?”.

Alguns concordam que é necessária, embora não tenha a importância da semântica, e que poderia ser excluída em alguns casos, como afirma CHARNIAK em [12] “semântica exige relações funcionais, relações funcionais exigem sintaxe, mas, semântica pode ser feita sem sintaxe.” Outros dizem que deve ser feita, mas sem grandes detalhes, ou ainda, integrá-la em outro conhecimento. SCHANK em [6] diz que “há pouca evidência de que pessoas usem noções de sintaxe padrão ...”, ou que, “regras não são particularmente úteis”. No entanto, PINKER et. alia. [18] dizem que os pesquisadores concordam que conhecimento lingüístico está baseado na forma de regras e princípios.

O exemplo clássico da Psicologia da Aprendizagem de que as crianças a princípio usam as conjugações verbais irregulares erradas, mas a partir de determinado momento começam a fazer as construções corretas é sempre tomado para fazer referência ao uso das regras. Sob o aspecto psicolingüístico [4], acredita-se que os indivíduos utilizem algo psicologicamente equivalente a um sistema de regras.

Mas, em que consiste a análise sintática? FREDERKING [13] discute dois pontos de vista, o tradicional e o intuitivo. Tradicionalmente, pode-se dizer que é o processo que identifica a estrutura dos constituintes de uma sentença de acordo com um conjunto de regras. Dessa maneira, é possível distinguir entre sentenças gramaticais e agramaticais.

Intuitivamente, parece aquela porção do conhecimento do usuário de uma língua que indica como combinar os significados dos morfemas, palavras e sintagmas, usando

informação posicional para conseguir significados de unidades maiores: palavras, sintagmas e sentenças, respectivamente.

A diferença entre estes dois enfoques é que o primeiro se fundamenta na estrutura pura de uma sentença mas, o segundo parece introduzir algo mais. Precisamente estabelece relações funcionais entre os constituintes da sentença, ou seja, mais que um sintagma nominal, o constituinte é identificado através do papel que desempenha na frase. É fácil admitir que esta é mais interessante em termos práticos, atribuir papel semântico ao constituinte claramente pressupõe conhecimento semântico.

Talvez, a princípio se tenha dado tanta atenção à sintaxe devido à existência de uma especificação formal capaz de expressar as regularidades de um subconjunto de sentenças da linguagem, a *gramática*. Esta é um conjunto de regras de reescrita, onde o lado direito das regras contém as partes em que pode ser decomposta uma sentença ou parte desta, que é caracterizada pelo lado esquerdo. A gramática vem de encontro a hipótese da abordagem simbolista, por ser uma manipulação de símbolos através de regras.

Os primeiros trabalhos trataram sintaxe sob o enfoque tradicional, a questão era distinguir a gramaticalidade das sentenças. Porém esta questão é algo difícil de definir [13], já que existem sentenças, que embora agramaticais são perfeitamente compreensíveis. Portanto, não parece que essa visão tradicional se molde a aspectos psicológicos. Contudo, muitas gramáticas surgiram, algumas mais simples, como gramática sintagmática, que serviu de base para a teoria gerativa. Outras mais complexas, como por exemplo a Gramática Transformacional. Além disso, aquelas que tentavam incluir conhecimento explícito em suas regras.

Desde a introdução da teoria gerativa, discussões sobre as capacidades de geração e análise das gramáticas aumentaram. A Gramática Transformacional não era suficientemente boa para análise por não permitir inversão do processo, embora várias tentativas tivessem sido feitas. No entanto, esta é uma vantagem das gramáticas livre de contexto, terem o mesmo poder para gerar e analisar sentenças. Mas, está completamente voltada para a estrutura dos constituintes, nenhuma informação semântica está explícita ou implícita em suas regras.

A gramática livre de contexto foi e ainda é bastante utilizada, como em AL-

LEN [5]. Muitos estudiosos acreditam ser este um formalismo poderoso para expressar sentenças de uma linguagem e que é restrito bastante para permitir construções de analisadores sintáticos eficientes. Mas, WOODS em [15] diz que não pode analisar constituintes relacionados que não sejam adjacentes e por isso não é adequada para tratar linguagem natural. Na verdade, em WINOGRAD [7] é conhecida também como gramática dos constituintes imediatos pelos lingüistas tradicionais ou forma normal de Backus pelos construtores de linguagens de programação. É um tipo particular das gramáticas de estrutura de frase.

A figura II.2 exemplifica uma gramática livre de contexto. Os não terminais ou variáveis, que representam categorias sintáticas estão em letras maiúsculas e os terminais[†] representando classes gramaticais em minúsculas, onde det, prep e subst são respectivamente, abreviações para determinante, preposição e substantivo. O símbolo S (de sen-

S	→	SN SV
		SV
SN	→	det SN1
		SN1
SV	→	verbo SN
		verbo SNP
		verbo
SNP	→	prep SN
SN1	→	subst

Figura II.2: Gramática Livre de Contexto.

tença) é o símbolo inicial, toda sentença que pode ser reconhecida ou gerada por esta gramática deve ter a presença de S. Aqui, S, ou seja, uma sentença pode ser decomposta num sintagma nominal (SN) mais um sintagma verbal (SV) ou, apenas em SV. Por exemplo, as regras $SN \rightarrow \text{det SN1}$ e $SN \rightarrow \text{SN1}$, dizem respectivamente que, um sintagma nominal (SN) pode ser decomposto em um determinante seguido de um SN1 ou simplesmente decomposto em SN1, que é um substantivo. Então, nesta gramática os sintagmas nominais podem ser apenas de dois tipos.

[†] O uso do termo "terminais" para símbolos como 'verbo', 'subst', 'det' e 'prep', não é exatamente próprio, mas devido a sua utilização no modelo que será analisado no Capítulo IV, será empregado desta forma. O uso correto seria para designação dos itens das sentenças.

A derivação de uma sentença é a aplicação de uma seqüência de regras que devem ser utilizadas para gerar ou analisá-la (reconhecê-la). Uma árvore de derivação sintática é a representação mais comum utilizada para expressar a estrutura de uma sentença. Considere a representada na figura II.3 para a sentença: 'A garota correu pelo bosque'.

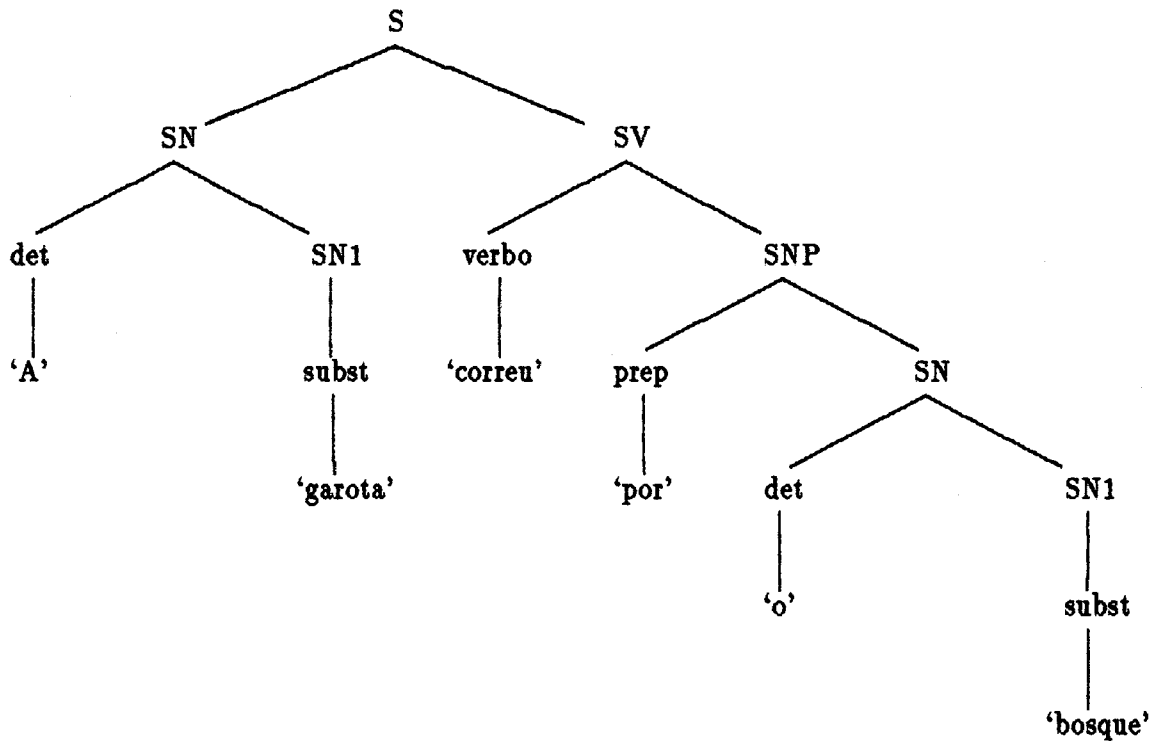


Figura II.3: Árvore de derivação sintática.

A Gramática de Cláusulas Definidas de PEREIRA et. alia. [19] é uma extensão das gramáticas livre de contexto, por transformar cada regra numa cláusula do cálculo de predicados de primeira ordem. Assim, é utilizada num provador de teoremas ou método da resolução da Lógica, através da linguagem PROLOG. Supera as inadequações das gramáticas livre de contexto por fornecer dependência de contexto, permitir inclusão de significados e condições extras nas regras. Um exemplo é mostrado na figura II.4, de acordo com as regras da figura II.2. Outra versão utilizando apenas informação posicional dos constituintes é possível, no entanto, a que é apresentada representa explicitamente as

árvores de derivação sintática.

sentença(s(SN,SV))	→ sintagma_nominal(SN), sintagma_verbal(SV).
sentença(s(SV))	→ sintagma_verbal(SV).
sintagma_nominal(sn(SN1))	→ substantivo(SN1).
sintagma_nominal(sn(Det,SN1))	→ determinante(Det), substantivo(SN1).
sintagma_verbal(sv(V))	→ verbo(V).
sintagma_verbal(sv(V,SN))	→ verbo(V), sintagma_nominal(SN).
sintagma_verbal(sv(V,SNP))	→ verbo(V), sintagma_preposicional(SNP)
sintagma_preposicional(snp(Prep,SN))	→ preposição(Prep), sintagma_nominal(SN).

æ

Figura II.4: Gramática de Cláusulas Definidas.

Um modelo de gramática descrito por WOODS em [15], Redes de Transição Aumentada, baseia-se em redes de transição recursiva, que por sua vez são extensões de redes de transição simples. Estas são grafos direcionados com estados e arcos rotulados, onde se distinguem estados iniciais e finais. Cada arco é rotulado com uma classe gramatical e os nós são categorias sintáticas. Um arco é seguido se a palavra está na categoria deste. No entanto, não podem descrever as linguagens que uma gramática livre de contexto descreve.

A introdução de recursão supera esta desvantagem, tornando as gramáticas livre de contexto e as redes de transição equivalentes em capacidade gerativa. As redes de transição recursiva permitem que os rótulos nos arcos sejam tanto estados quanto símbolos terminais. Um exemplo de uma rede de transição é como na figura II.5, a rede SV. Uma frase é um SV correto se há um caminho do nó SV para um estado final avaliando toda palavra no sintagma. Esta rede reconhece os mesmos sintagmas que as regras $SV \rightarrow verbo$ e $SV \rightarrow verbo SN$ da gramática da figura II.2.

A Rede de Transição Aumentada adiciona a cada arco da rede de transição recursiva uma condição arbitrária que deve ser satisfeita para que o arco seja atravessado e, caso isto aconteça, tem um conjunto de ações a serem tomadas. É capaz de apresentar

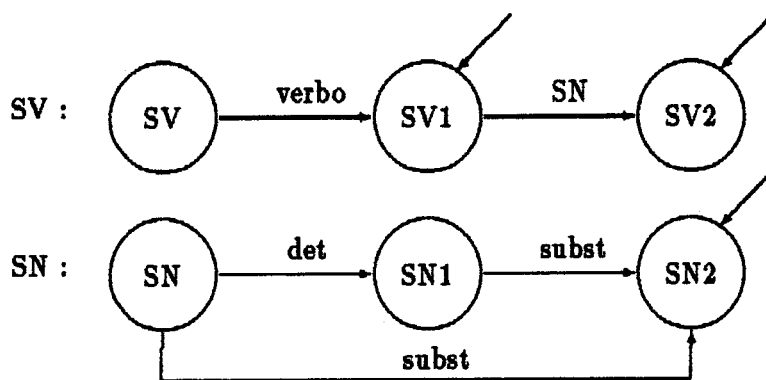


Figura II.5: Uma rede de Transição. SV é o estado inicial e SV1, SV2 e SN2 os finais.

o equivalente computacional da Gramática Transformacional sem precisar separá-lo num componente, uma das objeções em relação à teoria de Chomsky.

Há outras gramáticas como a léxico-funcional [20,21] e a gramática de MARCUS [22]. A primeira é constituída de regras livre de contexto que geram várias estruturas de frases e regras lexicais expressas no padrão lógico, argumento e predicado, que podem alterar a função do argumento numa dada estrutura. Como cada item léxico tem uma representação, não utiliza as transformações como na teoria de Chomsky. Por exemplo, a voz ativa e passiva de uma sentença estão representadas no léxico. A segunda é melhor compreendida através de seu analisador, dado que é subentendida neste. Pode-se dizer que é um conjunto de regras padrão/ação. Cada regra é formada de um padrão, que é instanciado com algum subconjunto de constituintes, e uma ação, seqüência de operações que atuam nestes constituintes.

Em geral, ALLEN [5] oferece duas técnicas principais de análise sintática simples e comuns, a análise sintática ascendente e descendente. A primeira começa com as palavras individuais e as substitui por suas categorias sintáticas. Usando regras de reescrita é possível substituir a análise obtida por uma outra de mesmo comprimento ou maior. A segunda começa pelo símbolo inicial (S) reescrevendo as demais que surgirem através das regras de reescrita. Assim, usa-se o lado direito para reescrever o símbolo do lado esquerdo. É fácil entender estas técnicas observando a figura II.3.

Métodos de análise sintática envolvem análises ascendentes e descendentes, redes de transição recursiva e gramáticas livre de contexto. Tem-se análise sintática descendente

em redes de transição recursiva ou em gramáticas livre de contexto e métodos mixtos.

O que atualmente todos concordam é que usar sintaxe explicitamente, ou seja, considerar a estrutura de uma sentença não é absolutamente necessário. Mas, não se pode dispensar este conhecimento, pois há casos em que sintaxe é decisiva, exatamente quando existe ambigüidade estrutural, isto é, uma sentença gera várias representações ou estruturas sintáticas. E a melhor forma de proceder é integrá-la num outro conhecimento, no semântico por exemplo, a conhecida análise sintática semanticamente dirigida.

II.3.2 Alguns Analisadores

Entre os vários aspectos considerados no projeto de um analisador sintático, o não-determinismo e ambigüidade são as principais preocupações da maioria dos analisadores existentes. Estes dois aspectos estão profundamente relacionados como mostra WINOGRAD [7], a inexistência do não-determinismo poderia conduzir a uma descrição estrutural incorreta, mas a sua existência pode produzir várias paráfrases estruturais, ou seja, dificultar a resolução da ambigüidade. Além destes, FREDERKING em [13] apresenta a vinculação do sintagma preposicional, integração com outro processo e tratamento de sentenças mal formadas.

Uma outra diferença é o grau de conhecimento semântico que contém, devido aos argumentos em favor da junção sintaxe-semântica, ou melhor um analisador sintático semanticamente dirigido. Dessa maneira, essa junção oferece não só a estrutura da frase mas também informações sobre as relações funcionais entre os constituintes, que são mais úteis em PLN.

Então as principais características que poderiam distinguir analisadores sintáticos de acordo com [13], além do grau que utiliza conhecimento semântico e tratamento de ambigüidades, também estratégia de controle e a modularidade, esta significa o nível de organização interna, que é desejável desde que a flexibilidade não seja afetada, o que não ocorre frequentemente.

O analisador sintático flexível específico por construção, FlexP [23] é um exemplo de utilização de multi-estratégia. Está baseado na análise ascendente e instanciação de padrões ("pattern-matching"). A escolha da primeira derivou da necessidade de reconhecer fragmentos de sentenças isoladas, mas como só esta conduziria a alternativas espúrias, a descendente também foi utilizada. Uma gramática de padrões lineares foi escolhida devido à segunda estratégia. A grande vantagem deste analisador é que cada sentença é tratada pela estratégia mais apropriada. DYPAR e CASPAR são dois analisadores baseados neste procedimento.

As principais estratégias que podem ser usadas num analisador multi-estratégia, segundo o autor de FlexP, seriam instanciação por "frames" de casos ou instanciação parcial utilizando uma gramática semântica, sistemas de casos ou transformação de forma canônica (em geral não acontece em Português). Um exemplo desta última seria: "John's house" se transformaria em "House of John". Argumenta ainda, analisadores sintáticos com multi-estratégia são melhores que os de estratégia única, pois permitem mais flexibilidade, redundância e habilidade no tratamento de conhecimento específico. Além disso trata ambigüidade semântica e estrutural sem ter que representar tais ambigüidades com a duplicação de informação não ambígua. O princípio utilizado é que conhecimento apropriado deve ser utilizado no tempo certo e não interferir em outros estágios do processamento.

Os analisadores sintáticos construídos a partir das Redes de Transição Aumentada [14], implementados em LISP, tiveram algumas vantagens em relação aos baseados nas gramáticas livre de contexto e Gramática Transformacional. Uma delas é a mencionada na seção II.2, a capacidade de análise. Tem o poder de uma Máquina de Turing, e acrescenta WOODS, possui eficiência de representação, capacidade de capturar regularidades das linguagens, legibilidade das gramáticas livre de contexto, flexibilidade de experimentação, eficiência de operação, pois o determinismo pode ser introduzido por eliminar (não totalmente) a recursão. O uso de registradores temporários permite tentar decisões sem o uso do "backtracking".

Um analisador sintático segundo o formalismo das Gramáticas de Cláusulas Definidas [19] está direcionado à construção das árvores de derivação sintática das sentenças

da gramática que traduz. O dicionário[†] léxico que apresenta é bastante rico em conhecimento de concordância sintática, como gênero e número para substantivos, modo, tempo e transitividade verbal. Utilizando o poder do “backtracking” e a estratégia descendente, chega-se à árvore de derivação correta.

Entretanto, acrescentar informação semântica através de uma gramática de casos ou de uma rede semântica é possível e relativamente simples. Para torná-lo capaz de executar, por exemplo, uma gramática de casos basta adicionar argumentos às cláusulas representando os casos de cada verbo. Além de apresentarem uma tradução de Redes de Transição Aumentada em Gramáticas de Cláusulas Definidas, PEREIRA et. alia. [19] discutem cada um dos pontos que WOODS apresentou como vantagem de seu formalismo.

O que parece é que os dois formalismos são capazes de incluir informação semântica através das condições empregadas. Utilizam a mesma técnica de controle. Talvez, alguma diferença esteja entre o uso de “backtracking” e de registradores temporários, que em termos computacionais podem refletir o mesmo processo.

FININ et. alia. [20] acreditando que a vantagem das Gramáticas de Cláusulas Definidas estava apenas no uso de variáveis lógicas, construíram um analisador substituindo os registradores das Redes de Transição Aumentada por variáveis lógicas. E assim, apostaram na superioridade do modelo de WOODS, dizendo que variáveis lógicas e unificação são técnicas poderosas que podem ser adaptadas para analisadores sintáticos sem o inconveniente do paradigma de prova de teoremas.

PARSIFAL, o analisador de MARCUS [22] está fundamentado no seu princípio de que é desnecessário analisar linguagem natural por um mecanismo que simule não-determinismo. Assim, constrói o seu interpretador de gramática baseado na hipótese do determinismo. Seu modelo preserva as possibilidades de desempenho de Redes de Transição Aumentada, mas difere por usar predição para substituir o uso dos registradores, e regras de produção que são mais transparentes que o código LISP, afirma.

Os analisadores também podem ser distingüidos quanto à forma e ao momento em que são tratadas as ambigüidades [13]. O FlexP, por exemplo, tenta evitar o que geralmente ocorre em sistemas interativos. Isto é, a maioria dos sistemas ou retornam

[†] Não consta na gramática apresentada na figura II.2.

todas as paráfrases possíveis ou informam da impossibilidade de interpretação. FlexP analisa a sentença mesmo que haja pontos que não possa resolver, estes são retornados ao usuário para uma possível solução.

PARSIFAL por utilizar o determinismo retorna apenas uma solução. Mas, as Gramáticas de Cláusulas Definidas pode gerar todas devido ao não determinismo do “backtracking”, ocorrendo isto também com as Redes de Transição Aumentada.

Assim, a divisão está entre os que adotam o “backtracking”, ou seja, a possibilidade de retornar e tentar nova interpretação, aqueles que utilizam paralelismo, manter todas as paráfrases, e os que agem deterministicamente em busca da interpretação correta. Além dessas, [13] discute a possibilidade de correção de erro inteligente. Contudo, todos estes analisadores são totalmente sintáticos e, atualmente concordam com a necessidade de maior introdução de semântica, dado que este conhecimento mais facilmente conduzirá às soluções de PLN.

II.3.3 Ambigüidades

Como várias vezes citado, ambigüidade é um dos grandes problemas para o PLN. É uma questão de difícil tratamento, pois é uma característica que é inerente às línguas naturais. Os seres humanos lidam muito facilmente com isto, devido principalmente à presença de informação de contexto ou conhecimento do mundo. Porém, coletar e representar toda informação necessária para tratar a diversidade de ambigüidades possíveis também é um grande problema.

Ambigüidades ocorrem quando existe a possibilidade de atribuir várias interpretações distintas a uma palavra, sintagma ou sentença, sendo que pode decorrer de diferentes fontes do conhecimento lingüístico. Para HANKMER et. alia. [24] estas fontes podem ser léxicas, funcionais, morfológicas ou sintáticas.

As léxicas produzem dois tipos de ambigüidade, a *polissemia*, que é a existência de vários significados diferentes para o mesmo item léxico, e a *homofonia*, a identificação

do mesmo som para itens semanticamente distintos. Aquelas que dizem respeito ao aspecto gramatical ou sintático do item léxico, ou seja, a propriedade de pertencer a mais de uma classe gramatical caracterizam as fontes funcionais.

As ambigüidades devidas às fontes morfológicas ocorrem quando novas palavras são formadas a partir de unidades básicas que já são ambíguas. Além disso, palavras formadas por acréscimo de sufixos podem ser atribuídas a categorias sintáticas diferentes de acordo com o sufixo. Por exemplo, há sufixos, como 'mente', que originam advérbios, e outros como 'oso', adjetivos.

As principais ambigüidades de origem sintática são classificadas em simples, atribuição de diferentes descrições estruturais; derivacionais, existência de estruturas profundas distintas. Estas se relacionam às transformações introduzidas pela teoria da Gramática Transformacional, que permitem interpretações diferentes para tais estruturas. Além destas, as introduzidas por processos anafóricos, isto é, atribuição de substantivo a um pronome oblíquo também são discutidas.

Podem ainda, serem classificadas quanto ao momento em que as ambigüidades são retiradas. Ambigüidades locais são aquelas que podem ser resolvidas dentro do âmbito da sentença, as globais, entretanto, exigem contexto e conhecimento do mundo. Em geral, as locais podem ser melhor resolvidas, já que são mais simples, embora não sejam exatamente fáceis.

Segundo DAHLGREN [25] a retirada de ambigüidade léxica computacional divide-se em atribuição de categoria sintática, como por exemplo, identificar se uma palavra é um substantivo ou um verbo, e retirada de ambigüidade de sentido de palavra dentro da categoria sintática. Nota-se uma ligeira diferença entre a divisão de HANKAMER (lingüista) e de DAHLGREN (lingüista computacional), que em geral, é assumida por grande parte dos lingüistas computacionais. Estes tratam todas as ambigüidades citadas como léxicas e assim são resolvidas a nível dos processos léxico e sintáticos, pouca representação do conhecimento do mundo é exigida, segundo este ponto de vista.

Os analisadores acima descritos apresentaram algumas alternativas para resolução de ambigüidades, mas podiam tratar apenas aquelas que estão especialmente envolvidas em descrições estruturais ou significados individuais das palavras, dado que são sintáticos.

Estas são ambigüidades locais, existem apenas dentro de uma sentença. Porém, as globais, que necessitam de contexto ou conhecimento do mundo, não são tratadas, sendo na verdade as mais interessantes e difíceis, pois exigem uma representação de conhecimento realmente eficiente e abrangente.

Um usuário nativo e comum de uma linguagem por certo não supõe a complexidade do conhecimento lingüístico que tem guardado em sua mente, porém o utiliza de forma extraordinariamente eficaz e rápida. Assim, um sistema só pode ser dito eficiente se é capaz de resolver todos ou grande parte dos problemas relacionados à ambigüidade. É impossível que PLN reproduza o comportamento de humanos frente a linguagem sem que esta parte esteja devidamente solucionada.

Capítulo III

Conexionismo e PLN

III.1 Conexionismo: Aspectos Gerais

Na década de 50, paralelamente à realização dos primeiros trabalhos na área de Inteligência Artificial, alguns pesquisadores, também interessados em reproduzir o comportamento humano num computador, defendiam um procedimento diferente do normalmente utilizado. Esta linha para tratamento dos processos cognitivos introduzia um outro paradigma, que buscava suas bases teóricas nos aspectos estruturais e funcionais do cérebro humano. Por isso, a Neurociência, subárea da Biologia dedicada à explicação e ao estabelecimento dos fundamentos básicos da estrutura e do funcionamento do cérebro, desempenhava papel fundamental dentro deste contexto. As pesquisas desta ciência mostraram que o cérebro consiste de bilhões de células nervosas interconectadas por milhares de ligações ($\cong 10.000$ para cada célula) formando uma rede [26]. E as informações em forma de estímulos ou impulsos nervosos seriam transmitidos às diversas áreas do cérebro através destas ligações.

Os poucos defensores deste paradigma, acreditavam então, que comportamento inteligente emergiria de interações entre elementos similares a estas células, que constituiriam um grande conjunto (sistema), no qual estariam altamente conectados. No entanto, os primeiros resultados não foram muito encorajadores, provocando várias críticas dos cientistas cognitivistas adeptos da abordagem simbolista.

Vale lembrar a grande polêmica causada por MINSKY e seu colega PAPERT. Em 1965, publicam um livro que é o resultado de uma análise detida do procedimento de ROSENBLATT para reconhecimento e classificação de padrões utilizando um modelo inspirado na abordagem biológica. Esta publicação que inclusive leva o nome do modelo, PERCEPTRONS, causou divisões devido às conclusões apresentadas. Para os autores era apenas uma demonstração da limitação e da incapacidade de tratar a complexidade do problema proposto [27,28]. Mas, para a comunidade científica, isto representou a quase destruição da área, pois os argumentos levaram a crer que esta linha não explicaria os processos da cognição no nível em que Inteligência Artificial já estava desenvolvendo.

Como diz MINSKY [3], que passou a ser um dos maiores críticos desta linha, estes sistemas não eram capazes de atividade cognitiva de mais alto nível. Por isso foram julgados não promissores ou prematuros. Mas, como ele mesmo afirma, pouco se sabia sobre comportamento de alto nível. Portanto, resultados expressivos realmente ainda não seriam possíveis.

Numa análise recente da repercussão destas conclusões, PAPERT [28] tenta justificá-las argumentando que naquela época devido a interesses de várias ordens, as pesquisas foram desfavorecidas economicamente e, assim definharam, culminando na conclusão citada.

Assim, em meio a tanta turbulência, as pesquisas nesta área foram, por assim dizer, abandonadas, ocorrendo o predomínio da abordagem simbolista para estudo da cognição humana. Como visto anteriormente, a Ciência Cognitiva e Inteligência Artificial direcionaram seus trabalhos de acordo com as hipóteses desta linha.

Com efeito, as bases teóricas do paradigma simbolista estavam melhores esclarecidas e conseqüentemente, os resultados práticos eram mais estimulantes e promissores. O outro paradigma, entretanto, estava essencialmente baseado em teorias biológicas, que eram menos evidentes nesta época. Mas, apesar disso, alguns filósofos apoiavam esta corrente, argumentando que o objetivo computacional de criar sistemas inteligentes seria mais facilmente atingido se fosse utilizado um modelo do cérebro humano. Isto devia-se a crença de que computador fosse um meio para modelagem do cérebro e assim, estaria mais próxima a compreensão dos processos de aprendizagem, a grande preocupação desta ala de filósofos [2].

Apesar da Neurociência ser a base de um paradigma conflitante com o simbolista, os cientistas cognitivistas seguidores deste último concordam que, saber como o sistema nervoso realiza processamento de informação complexa é fascinante, possibilita testar algumas teorias, explicitar o poder e as limitações de humanos para lidar com certas habilidades e, contribui para o projeto de ambientes de aprendizagem [2]. E portanto, é obviamente uma ciência básica da Ciência Cognitiva.

Alguns adeptos da linha biológica, atribuíram o seu insucesso inicial ao pouco conhecimento sobre a atividade mental humana. Porém, já era constatado a sua indubitável complexidade. Ainda não é claro, por exemplo, como ocorrem os processos de recuperação da informação, embora fenômenos como aprendizagem e memorização sejam mais bem conhecidos. Também, há que se considerar que o tipo de processamento para sistemas inspirados nesta abordagem, dificilmente ocorreria com eficiência nas máquinas disponíveis na década de 50.

Além destes, um outro argumento que deve ser citado é a simplicidade dos modelos computacionais da época. A complexidade biológica conhecida era evitada, desde que o grande número de elementos exigidos para uma aproximação com o comportamento cerebral, tornaria o modelo inviável se estes fossem por demais complexos. Alguns neurofisiologistas acreditam no entanto, que tal procedimento é um equívoco, uma maior aproximação com a realidade seria a melhor tentativa de se obter bons resultados.

VEMURI [29] diz que, para alcançar os objetivos cognitivistas de acordo com esta abordagem, é necessário compreender como são organizadas e executadas as operações ou funções do cérebro, que são diferentes da manipulação formal de expressões simbólicas.

Atualmente, vem se tentando superar as dificuldades de conhecer e entender tais funções. Com a introdução de novos procedimentos e o aprimoramento dos já existentes, os pesquisadores simpatizantes da abordagem *conexionista*, como é agora conhecida, estão incitando o progresso desta linha. Alguns atacam este interesse, classificando-o de modismo, devido ao desenvolvimento tecnológico. Mas, a levar em consideração críticas deste nível, a própria Inteligência Artificial já teria estancado.

Assim, nos últimos quinze anos, vigora uma notável expansão do *conexionismo*, que tem agora, encontrado novos adeptos e mais incentivos. Segundo FELDMAN et.

alia. [30], pertencentes a um grupo que desenvolve pesquisas nesta área, conexionismo é a idéia de que computações possam ser efetuadas pelas ligações existentes entre um grande conjunto de unidades de processamento. Isto deve-se à crença de que o cérebro organize o conhecimento ou inteligência nas ligações (conexões) entre as células nervosas. COWAN et. alia. [31], caracterizam este fato como *neconexionismo*, pois a área é antiga e o que está agora ocorrendo é apenas um ressurgimento, com renovação dos modelos e técnicas.

Uma justificativa para pesquisas nesta direção citada por VEMURI [29], é que as habilidades cognitivas estão classificadas dentro de categorias (pesquisa, representação e aprendizagem) profundamente relacionadas a duas propriedades importantes do cérebro, a auto-associação e a auto-organização. A primeira é a capacidade de reconstruir uma informação completa sem que se tenha todos os dados desta. A segunda, é a possibilidade de adquirir conhecimento através de processos de aprendizagem que envolvem organização frente a influências externas.

Mas, a possibilidade de tornar os modelos conexionistas massivamente paralelos é a maior motivação para os seguidores desta teoria. Em [30], FELDMAN et. alia argumentam que os procedimentos atuais de Inteligência Artificial exigem muito tempo de processamento, enquanto o cérebro expressaria comportamentos complexos em menos tempo. Assim, utilizar paralelismo seria conveniente, e mesmo porque muito tem-se investido em computadores e computação paralela.

McCLELLAND et alli. [32], apesar de explorarem o processamento paralelo, dada a hipótese da proximidade deste com a fisiologia cerebral, dizem que este não é o principal motivo. Por se julgarem, principalmente cientistas cognitivistas, esperam oferecer com tal procedimento mecanismos que sejam computacionalmente eficientes e psicologicamente plausíveis para elucidar a cognição humana.

Os resultados com os modelos conexionistas têm estimulado novas pesquisas, embora questões resolvidas pela abordagem simbolista não sejam suficientemente bem tratadas por este procedimento. Mas, não parece que avaliação do alcance destas duas correntes possa ser feito apenas analisando resultados e procedimentos. Como dito em [33], o principal ponto de discussão deveria ser o objetivo final, que se resume na seguinte questão: Pretende-se construir máquinas inteligentes ou compreender comportamento do cérebro? Pois, parece que estes objetivos têm se misturado, fazendo com que várias divergências

ocorram quanto ao alcance das propostas.

Seria precipitado fazer qualquer previsão quanto ao desenvolvimento e a capacidade da teoria conexionista para explicar a atividade cognitiva humana. Porém, a perspectiva de atingir este objetivo, parece estar mais próxima desta, pois o que se tem tentado com tais modelos é utilizar o conhecimento biológico, até então pouco explorado, para entender comportamento inteligente. É fácil encontrar aspectos mais atrativos nesta postura, já que é pelo menos mais próxima do real.

As duas abordagens contudo, compartilham algumas dificuldades, desde que têm o mesmo objeto de estudo. No entanto, hoje, dividem entre si tanto o desenvolvimento prático de teorias psicológicas quanto as base para a formulação destas. São inclusive, consideradas “irmãs” por DREYFUS et. alia. [2], sendo, por motivos óbvios, a conexionista, a “irmã natural” e Inteligência Artificial, a “irmã artificial”.

III.2 Evolução e Principais Modelos

Como anteriormente discutido, há apenas algumas semelhanças entre a modelagem computacional e aspectos biológicos. Isso será constatado a partir das descrições resumidas destes dois enfoques.

De acordo com a Neurofisiologia [26], todo o funcionamento do cérebro está basicamente vinculado a atividade de células nervosas ou *neurônios*. Em linhas gerais, nestas células se distinguem três partes: o corpo da célula (soma) e dois tipos de prolongamentos que possibilitam a ligação entre as várias células. Estes dois tipos de estrutura, estão divididos em um *axônio* e vários *dendritos*. A principal diferença entre eles é que o axônio é único, possui as suas próprias ramificações (colaterais) e se liga a outras células nervosas ou também a células musculares ou glandulares. Os dendritos podem ser em grande número e, assim como o soma, são pontos terminais dos axônios de outras células. Há várias formas de neurônios de acordo com as diferentes características dos dendritos, pois estes podem não existir, e no caso de estarem presentes podem formar uma estrutura

regular ou irregular.

As ligações entre os neurônios ocorrem através das chamadas *sinapses*, que são pontos de ligação entre uma terminação axônica e outra célula. A princípio, acreditava-se (errôneamente) que o estímulo conduzido por um axônio se transferia diretamente para a outra célula, o que é conhecido como sinapse elétrica. Verificou-se que esta, em geral, não ocorria, e sim as sinapses químicas, ou seja, liberação de substância química de uma terminação axônica que atua sobre a célula vizinha.

A atividade de uma célula nervosa está relacionada à concentração de substâncias no meio intra e extracelular. Se não há agentes externos atuando no neurônio tem-se seu *potencial de repouso* (negativo), que representa o desequilíbrio de cargas negativas dentro da célula e cargas positivas no meio extracelular, cujo o potencial é 0 (zero) por convenção. Diz-se então que a célula está polarizada, pois há excesso de cargas negativas em seu interior. A partir de estímulos, que são conduzidos à célula por sinapses ou vizinhanças da membrana, esta se despolariza até atingir um *limiar*. Neste inicia-se o *potencial de ação*, que ao ultrapassar o potencial zero permite a *excitação* do neurônio, ou seja, decréscimo de cargas negativas dentro da célula. Depois disso a célula se repolariza, podendo reiniciar o processo de excitação.

Além dos processos de excitação também existem os que possibilitam *inibição*, isto é, a redução da atividade do neurônio. Em geral, a forma mais importante destes processos são de natureza sináptica, neste caso a ativação da sinapse conduz a um decréscimo da excitabilidade do neurônio. Assim, é possível distinguir as sinapses que conduzem excitação ou inibição.

A inibição pode ocorrer nos dois extremos da sinapse. Na inibição pré-sináptica (extremo do axônio) há redução ou extinção da liberação de substâncias, não alterando a célula no outro extremo. A pós-sináptica (extremo da célula vizinha), entretanto, diminui a excitabilidade da célula. Há varios tipos de inibição, a *retroativa* ("feedback") por exemplo, ocorre quando os neurônios atuam inibitoriamente sobre as células que os excitaram. Uma forma variada deste tipo é a *inibição lateral*, nesta os neurônios atuam não só nas células que os excitaram como também em todas que tenham função idêntica. Na *inibição antecipada* ("feedforward"), os neurônios são inibidos sem que tenham sido excitados.

Estas inibições servem para suprimir excitações supérfluas, ou seja, há uma realimentação negativa sobre a excitação inibindo-a com intensidade maior que a sua ocorrência. A realimentação positiva, entretanto, é considerada uma questão aberta.

Tem-se ainda, que as sinapses são modificáveis, quanto maior a freqüência de utilização de uma sinapse, melhor transmitirá a excitação. Aliada a esta certa *plasticidade*, isto é, capacidade de adaptação (observada em todo o cérebro), existem as funções de *aprendizagem e memorização*. Também é dito que pelo conceito da especificidade funcional, os neurônios podem ser apenas excitatórios ou inibitórios.

Em suma, a estrutura do cérebro consiste de um grande conjunto de células nervosas denominada *rede neuronal* ou *neural*[†], onde cada elemento se comporta da forma geral descrita acima. Nos impulsos conduzidos pelos axônios são transmitidas informações a todo cérebro através de excitações ou inibições, assim quando ligações adequadas surgem da atividade neuronal, tem-se o processamento da informação. A recepção, armazenamento e liberação são algumas das propriedades da rede, sendo que estas permitem que funções elementares, como aprendizagem, memória e recordação existam, tornando possível aos humanos desempenharem habilidades inteligentes ou cognitivas.

É observável na maioria dos modelos computacionais para esta teoria biológica, características comuns. Cada elemento ou unidade de processamento, que representa um neurônio esta relacionado a um valor chamado *limiar*. O estado de ativação, ou seja, o estado em que se encontra a unidade, em excitação ou em inibição, também estado ativo ou inativo, em geral, é estabelecido por valores discretos e binários, ou seja, um valor fixo para o estado ativo e outro para o inativo. Contudo, há modelos que tomam estes valores dentro de um intervalo fixo. Assim, um estado ou configuração da rede é determinado por distinguir quais as unidades ativas e inativas.

As unidades representam os conceitos envolvidos no universo de discurso do problema que se deseja tratar. Por exemplo, quando tratando Linguagem Natural, estas unidades podem representar categorias sintáticas, palavras ou até mesmo sentenças. São distingüidas duas formas de representação, a *localizada*, onde um conceito é representado

[†] Em [34] o uso do termo neuronal se relaciona com as propriedades de neurônios reais e, neural apenas quando similaridades gerais estão envolvidas. Alguns acreditam ser apenas estrangeirismo, fazendo com que se prefira neural. Contudo será usada neuronal, pois acreditam outros que este é o termo biologicamente correto.

apenas por uma unidade e a *distribuída*, na qual um conceito é expresso por um conjunto de unidades, onde cada uma representa um conceito mais simples que o constitui.

As sinapses são interpretadas como as ligações entre as unidades, às quais são atribuídos valores discretos, que representam a força ou intensidade sináptica e são comumente chamados *pesos*. Pesos positivos indicam as sinapses excitatórias, negativos as inibitórias e nulos a inexistência de sinapse.

A *topologia* de uma rede destas unidades, ou seja, o número e a disposição das unidades e as suas ligações, depende do problema e do modelo. Mas, a quantidade destas, obviamente, é inferior ao que supostamente existe no cérebro. Normalmente, são colocadas em forma de camadas, onde unidades de uma mesma camada têm funções idênticas, sendo portanto, possível a utilização dos processos de inibição citados.

O estímulo ou sinal transmitido por uma unidade, denominado *saída*, é uma função dos estados das unidades a ela ligada, dos pesos destas ligações e de seu limiar. Esta função de saída, em geral, é a soma ponderada das entradas para a unidade, onde estas, são influências recebidas das outras unidades da rede ou externas. Há hipótese de que este também seja o processo da rede neuronal humana, segundo RECCE et. alli. em [35].

Assim, o processamento ou atividade de uma rede com tais características é realizado tomando-se um estado inicial da rede, em geral, aleatório, e a partir deste deixa-se que interações ocorram entre as unidades. As interações provocam mudanças nos estados das unidades. Isto é, dependendo das entradas para uma unidade, esta poderá mudar seu estado, atualizar-se. Passando do estado ativo para o inativo e vice-versa.

A rede evolui dentro deste processo de atualizações sucessivas até atingir um estado em que mudanças não sejam mais possíveis. Este estado é denominado *estado estável*, e o processo pelo qual passa a rede para atingí-lo, ou seja, a estabilização da rede, *relaxação*.

Superficialmente, parece que a analogia feita pela modelagem computacional está próxima à realidade. Por exemplo, o uso de um limiar, o processo de excitação, a função de saída e a evolução da rede fazem pensar na reprodução da atividade biológica. No entanto, estas observações podem apenas ser consideradas semelhanças. Além disso, uma

analogia perfeita não seria possível. Como dizem REEKE et. alia. [34], modelos conexionistas não reproduzem a rede neuronal humana, não sendo de fato este o objetivo destes modelos. Há várias diferenças, o reduzido número de unidades e ligações, as simplificações da atividade de uma unidade e um processo de relaxação que não ocorre, pois o cérebro está continuamente exposto a mudanças. Pode-se justificar estas divergências através de comparações entre cérebro e computador. Por exemplo, como dito por COTTRELL et. alia. [36] a natureza mais lenta de um neurônio para responder a um estímulo, a imprecisão e o paralelismo do cérebro, que não são observadas na máquina.

Além destas diferenças, existem outras que são notadas nas adaptações feitas para atender as particularidades de cada modelo. A forma com que ocorrem as atualizações das unidades é uma delas. Alguns modelos adotaram o processo síncrono, em que uma unidade pode mudar seu estado em espaços de tempo regulares, outros utilizam o assíncrono, onde uma unidade é escolhida aleatoriamente para atualizar-se. Sob o ponto de vista biológico, o primeiro é menos provável de se verificar, pois não parece que um neurônio seja excitado ou inibido em intervalos de tempo determinados. Há uma maior plausibilidade para o segundo.

Uma outra é a regra utilizada para efetuar as atualizações. Em alguns isto é feito deterministicamente. Isto é toma-se a soma das entradas para uma unidade, que é comparada ao seu limiar. No caso desta soma ser maior que seu limiar, a unidade pode tornar-se ativa, se ainda não estiver neste estado. Caso contrário, limiar menor que a soma, a unidade não se ativará. Em outros faz-se o uso de probabilidades. Então, uma unidade se ativará se houver alta probabilidade para que isso aconteça, em caso diferente, não. Assim, os modelos quanto à regra de atualização podem ser determinísticos ou probabilísticos. Estes últimos decorrem mais da proximidade dos sistemas conexionistas com sistemas dinâmicos do que com características biológicas. A utilização de procedimentos deste tipo deve-se à necessidade de simulação dos modelos, já que não se encontram disponíveis em máquinas.

Pode-se também distinguir os modelos que incluem métodos ou regras de aprendizagem. Será visto, que este importante processo ocorre através de mudanças efetuadas nos pesos das ligações, apoiado na hipótese de que as sinapses são mutáveis. No entanto, não é observada na maioria dos modelos.

Como estão representados os conceitos é uma importante diferença entre os modelos. Psicologicamente, acredita-se que a representação distribuída seja a mais plausível, pois expressa melhor a hipótese de que há predominância de conceitos simples no cérebro. Os mais complexos seriam em menor número. Contudo, não é a mais utilizada. A dificuldade em escolher as características necessárias e suficientes para representação de um conceito, especialmente quando tratando conceitos similares, é a principal razão para se preferir a localizada. Também, nesta tem-se menor número de unidades que torna menos complexo o controle do processamento.

A base dos modelos que, atualmente utilizam boa parte destas características, foi o modelo desenvolvido por McCULLOCH e PITTS em 1943. É considerado a primeira maior contribuição para esta área. Neste, os neurônios foram classificados como neurônios formais, pois desempenhavam operações lógicas simples [31,34].

Um neurônio i de McCULLOCH e PITTS [29] se caracterizava por ter peso $w_i = +1$ para a ligação excitatória e $w_i = -1$ para a inibitória; um limiar L (qualquer inteiro positivo); saída y e valor do estado ativo $+1$ e inativo 0 (zero). Entradas x_i para $i = 1, \dots, n$, onde n é o número de neurônios e a saída é expressa em termos de suas entradas conforme a expressão,

$$y = g \left(\sum_i w_i x_i - L \right), \quad (\text{III.1})$$

onde, $g(p) = 0$ se $p < 0$ e $g(p) = 1$ se $p > 0$, sendo p a expressão entre parênteses.

Uma rede destes neurônios formais evolui sincronamente, segundo a expressão abaixo, em que w_{ij} representa a ligação entre os neurônios i e j .

$$y_j = g \left(\sum_j \sum_i w_{ij} x_i - L_j \right). \quad (\text{III.2})$$

Com este modelo, seus autores mostraram como seria aplicação da lógica simbólica a modelos neuronais. Entre outras limitações, o sincronismo e a impossibilidade de aprendizagem, propriedade desejável, principalmente para estudos dos processos cognitivos.

Depois que McCULLOCH e PITTS mostraram como redes similares às neuronais poderiam computar, o problema tornou-se a compreensão de como tais redes poderiam aprender. O primeiro trabalho realmente significativo foi a regra de HEBB [37]. Antes

disso, acreditava-se que mudanças físicas deveriam ocorrer na rede para suportar aprendizagem, contudo não era claro quais seriam estas mudanças [27].

De acordo com o postulado de DONALD HEBB, que data de 1949, um organismo aprende quando ocorrem mudanças nas sinapses, contudo não havia evidência definitiva para esta proposição [31]. Assim, lançou a hipótese da ativação repetida, dessa maneira, grupos de neurônios fracamente conectados se organizariam dentro de uma forte conectividade, caso fossem sincronamente ativados.

Baseado nestas hipóteses, surgem procedimentos e modelos que tentam introduzir aprendizagem, que é vista como um treinamento. Porém, nota-se que a forma como HEBB colocou seu postulado, restringe o processo de aprendizagem a um reforçamento de conexões fracas, que alguns acreditam estar bastante próximo às idéias "behavioristas". No entanto, RUMELHART et. alia. [27] dizem que a hipótese de que devem haver mudanças nas sinapses continua sendo empregada.

No modelo apresentado acima esta propriedade não está incluída, pois os pesos são fixos. Assim, em 1957, outro pioneiro nesta área, ROSENBLATT, cria as redes denominadas PERCEPTRONS, cujo o objetivo era mostrar como as redes anteriores poderiam ser treinadas a reconhecer e classificar padrões, que foi considerado por McCULLOCH e PITTS um problema central em teorias de comportamento inteligente [31,34].

Um dos procedimentos de ROSENBLATT, descrito em [35], é uma forma simples de aprendizagem, porém muito utilizado, é denominado atualmente *aprendizagem competitiva* em [27]. A aprendizagem seria feita a partir de um conjunto de entradas, em dois estágios. No primeiro, as unidades que representam a entrada para a rede e as que representam a saída permanecem no estado desejado, é como se já tivesse terminado um processamento bem sucedido daquelas entradas. Tenta-se então, ajustar os pesos para esta configuração da rede, atribuindo-lhes valores. No segundo estágio, apenas as entradas estão representadas na rede, e assim novamente ajusta-se os pesos das ligações para que a rede comporte adequadamente ao tentar responder a estas entradas. Este processo é repetido várias vezes. Assim, depois disso, espera-se que para entradas iguais ou próximas às do conjunto, a rede proceda de maneira a atingir o resultado desejado. Este procedimento foi comprovado no reconhecimento e classificação de alguns caracteres, embora não pudesse distinguir caracteres complexos.

Isto foi principalmente, demonstrado por MINSKY e PAPER, como já mencionado. Uma das grandes críticas era que PERCEPTRONS de uma camada não representaria nenhum dos problemas a que se destinava, sem que utilizasse um grande número de unidades [35]. Entretanto, VEMURI em [29] diz que várias aplicações decorrentes utilizaram este modelo com três camadas, que é a sua forma mais geral.

Mas ficou claro que PAPER em [28], tenta uma reabilitação com a comunidade científica. Isto deve-se ao grande desenvolvimento que vem alcançando esta abordagem para o processamento da informação.

Um exemplo disso, são as contribuições significativas dos modelos desenvolvidos por HOPFIELD [38]. Para tanto, toma as redes intratáveis pelo modelo de ROSENBLATT e tenta superar algumas de suas limitações, como neurônios síncronos e o não interesse pelas propriedades coletivas computacionais que podem emergir da evolução da rede.

O modelo que introduz é então, semelhante ao PERCEPTRON, a diferença está nos novos resultados que são obtidos a partir do uso de neurônios assíncronos e da preocupação com propriedades computacionais emergentes, tais como, capacidade de generalização, reconhecimento de familiaridade, categorização, correção de erro e retenção de seqüência de tempo. Estas produzem uma memória associativa.

O modelo original utiliza neurônios de estados discretos e binários, caracterizados pelos valores de saída s_i iguais a 0 ou 1. Há duas fontes de entrada para cada neurônio i , sendo I_t a entrada total dada por:

$$I_t = \sum_{i \neq j} w_{ij}s_j + I_i, \quad (\text{III.3})$$

onde, I_i é a entrada externa e w_{ij} , o peso da ligação entre i e j .

O algoritmo escolhe aleatoriamente uma unidade para atualizar-se, e para esta atualização aplica a regra do limiar dada pela equação (III.4), onde o limiar é dado por θ_i . Num dado instante, apenas uma única unidade se atualiza.

$$\begin{aligned} s_i &\rightarrow 1 & \text{se } I_t > \theta_i \\ s_i &\rightarrow 0 & \text{se } I_t < \theta_i. \end{aligned} \quad (\text{III.4})$$

Com este modelo, HOPFIELD mostra que se $w_{ij} = w_{ji}$, ou seja, se as ligações são simétricas, o sistema converge para estados estáveis. Isto fica garantido pela construção

de uma função de energia que decresce para qualquer mudança de estado.

$$E = -\frac{1}{2} \sum_i \sum_j w_{ij} s_i s_j - \sum_i I_i s_i + \sum_i \theta_i s_i. \quad (\text{III.5})$$

Assim, de acordo com este modelo, atingir um estado estável ou estabilizar a rede, significa minimizar a sua energia inicial, dada pela equação acima. O objetivo então, é apresentar uma rede que tenha evolução similar a uma memória associativa, ou seja, dada uma entrada encontrar o estado estável mais próximo, que é um mínimo local da equação de energia.

Apresenta ainda um outro modelo que preserva os comportamentos significantes do modelo original, que é determinístico, e tem os valores de saída dos neurônios, s_i para cada neurônio i , variando dentro do intervalo $[0, 1]$. E por isso é conhecido como modelo contínuo. Entre outras aplicações, os modelos de HOPFIELD são utilizados em problemas de Otimização Combinatória [37]. A função de energia introduzida pode ser comparada a uma função objetivo, utilizada nestes problemas.

Porém, em certos problemas um mínimo local não representará a solução do problema que se pretende tratar, embora este seja um estado estável da rede. Esta é uma limitação deste modelo. Através de aplicações da técnica do "Simulated Annealing" [39], HINTON et. alli. [40] descobriram como superar esta limitação, atingindo o mínimo global de energia.

O "Simulated Annealing" é uma técnica de busca estocástica que foi utilizada em problemas de Otimização e que pode ser vista como uma forma do Método de Monte Carlo. Este método, primeiro introduzido por Von Neumann e Ullam durante a Segunda Guerra Mundial, foi modificado por METROPOLIS et. alli. [41]. O objetivo era encontrar soluções aproximadas para problemas que implicariam em sistemas, cujo tratamento analítico seria complexo e dispendioso.

O método descrito em [41], pode ser resumido como o processo para alcançar o equilíbrio de uma substância atingindo o mínimo de energia numa temperatura fixa T . Isso é feito considerando alterações no estado ou configuração inicial do sistema (recipiente com a substância). Assim, calcula-se a energia do estado atual, caso seja menor que a do anterior aceita-se este novo estado, que é agora o inicial. Em caso contrário toma-se a probabilidade dada por $p(\Delta E) = e^{(-\Delta E/KT)}$, onde ΔE é a diferença de energia entre

os dois estados e K , a constante de Boltzmann. Se esta probabilidade for maior que um número aleatório o novo estado é também aceito, de outra feita não. Até atingir o mínimo, o sistema passa por um grande número de estados, sendo portanto, bastante lento [41].

Através deste Método de Monte Carlo é possível atingir um mínimo de energia pois ao permitir 'saltos' de energia evita-se a estabilização num mínimo local desta. Porém, alcançar o equilíbrio (estado estável) em baixas temperaturas é um processo lento, devido à probabilidade de aceitar novas configurações ser muito pequena apenas as que minimizam a energia são aceitas. Mas, em altas temperaturas o processo é rápido por não serem consideradas pequenas diferenças de energia, sendo aceita um grande número de configurações.

Assim, a técnica [39] "Simulated Annealing", resolve este problema por utilizar uma seqüência decrescente de temperaturas. O processo começa em alta temperatura e prossegue com decréscimos graduais desta. Quando o equilíbrio do sistema é alcançado numa dada temperatura esta é decrementada e reinicia-se o processo, uma nova fase. O processo termina ao atingir um estado em que nenhuma modificação é aceita e a temperatura neste caso é mínima (ponto de congelamento). Assim, em altas temperaturas é feita uma pesquisa grosseira no espaço de estados do sistema, diminuindo-a, refina-se a pesquisa. Assim, esta técnica de busca que utiliza refinamentos sucessivos permite atingir o mínimo global de energia do sistema.

Tomando esta técnica como base, HINTON et. alli. [40] introduziram a Máquina de Boltzmann [31,42]. Esta é uma rede de HOPFIELD que implementa o Método de Monte Carlo [31], cujo o processo de evolução permite que se atinja o mínimo global e para isso utiliza a mesma equação de energia (III.6), porém não serão consideradas as entradas externas, isto é, elimina-se I_i para $i = 1, \dots, n$ e assim tem-se a equação abaixo:

$$E = -\frac{1}{2} \sum_i \sum_j w_{ij} s_i s_j + \sum_i \theta_i s_i. \quad (\text{III.6})$$

A rede é constituída de unidades que têm valores binários para os estados de ativação, 0 inativo e +1 ativo, as ligações são simétricas e o processo é assíncrono tomando uma unidade por vez, como no modelo anterior. A equação de energia está associada a cada estado do sistema. Para atingir a energia mínima utiliza um esquema de atualização probabilístico segundo a regra de decisão que é dada por:

$$s_j = \begin{cases} 1 & \text{com probabilidade } p_j \\ 0 & \text{caso contrário,} \end{cases} \quad (\text{III.7})$$

onde, a probabilidade da j -ésima unidade (p_j) se ativar é como abaixo:

$$p_j = \frac{1}{1 + e^{(-\Delta E_j/T)}}.$$

A expressão $\Delta E_j = \sum_i w_{ji}s_i - \theta_j$, significa a diferença de energia entre o estado inativo e ativo da unidade j . Considerando a equação de energia (III.6), ΔE_j expressa a diferença de energia entre dois estados do sistema, ou seja, sendo j a unidade escolhida aleatoriamente, esta expressão é a diferença de energia do sistema todo. Quando T se aproxima de zero a Máquina de Boltzmann procede deterministicamente. Observe que neste caso, se $\Delta E_j \leq 0$ então $s_j = 0$ e, caso contrário, $\Delta E_j > 0$, $s_j = +1$. Na verdade, é como na equação (III.4), com as devidas mudanças de nomenclatura.

O modelo que será descrito (Capítulo IV) faz uso da Máquina de Boltzmann, que é uma combinação do modelo original de HOPFIELD e o "Simulated Annealing". Apenas uma modificação é introduzida, os valores dos estados de ativação das unidades, ou seja, ao invés de 0 e 1, foram assumidos -1 para inativo e 1 ativo, contudo há uma correspondência biunívoca entre as duas versões.

Assim, para estes valores tem-se que $\Delta E_j = 2(\sum_i w_{ji}s_i - \theta_j)$, a diferença entre o estado inativo (-1) e ativo (+1) da unidade j . Então a probabilidade de se ativar é dada por $p(\Delta E_j)$, e a de se tornar inativa por $p(-\Delta E_j)$. Um algoritmo (figura III.1) que reflete a utilização destas duas técnicas é como se segue, onde C é uma configuração, T temperatura e k indica a fase do processo. Porém, é muito lento devido ao "Simulated Annealing", que exige várias temperaturas, e também o Método de Monte Carlo, que como visto, precisa de um grande número de passos, que é determinado de forma experimental.

Os modelos de HOPFIELD e a Máquina de Boltzmann são duas das contribuições mais significativas em termos de modelagem e processamento gerais da linha conexionista. Mas, o aspecto mais interessante desta última é sua capacidade de aprendizagem, que ocorre através de exemplos que são dados a rede para que mude os pesos das ligações. Assim, quando frente a qualquer exemplo particular, a rede pode interpretá-lo. Além disso, exemplos parciais podem ser completados.

É um método não supervisionado, pois a rede pode criar uma representação in-

```

k := 0; C := Ck; T := Tk;
enquanto sistema não congelado faça
  início
    enquanto não equilíbrio em T faça
      início
        escolha elemento j;
         $\Delta E_j := 2(\sum_i w_{ji}s_i - \theta_j)$ ;
         $p_j := \frac{1}{1 + e^{(-\Delta E_j/T)}}$ ;
        com probabilidade  $p_j$ ;
          faça  $s_j := +1$ 
        caso contrário
          faça  $s_j := -1$ 
      fim;
      k := k + 1; T := Tk;
  fim;

```

Figura III.1: Algoritmo para a Máquina de Boltzmann.

terna que relaciona os exemplos utilizados na aprendizagem [31]. Faz uso de unidades escondidas ou invisíveis, isto é, unidades que não representam qualquer conceito no domínio do problema, apenas existem computacionalmente. Apesar de se obter bons resultados, é muito lento.

Um esquema de aprendizagem que tenta superar a utilização de ligações simétricas da Máquina de Boltzmann é o “Back-propagation”. É um procedimento supervisionado de dois estágios. No primeiro, sentido progressivo, (“forward”), a rede é estimulada e as respostas anotadas e, no regressivo (“backward”), os pesos das ligações existentes com as unidades escondidas são ajustados [31].

Há uma vasta quantidade de modelos, para as mais diferentes aplicações, visão, audição, linguagem natural, percepção entre outras. Mas, como pode ser observado nestes que foram descritos, todos possuem várias similaridades, sendo as diferenças uma questão de adaptação de um modelo ao problema.

Em [30] são apresentadas algumas características de um modelo geral. Uma destas, que pode inclusive ser observada em vários modelos relacionados a PLN, é a formação de redes WTA (“Winner-Take-All”), que permite fazer distinções entre unidades que tenham funções similares. Nestas redes há ligações inibitórias entre todas as unidades,

ou seja, todas se inibem mutuamente. É também denominada “inibição lateral”, encontrando plausibilidade biológica para seu estabelecimento, de acordo com o que foi dito a este respeito.

McCLELLAND et. alli. [43], que desenvolvem pesquisas junto ao grupo do Processamento Distribuído e Paralelo, explorando o paralelismo dos modelos conexionistas, também apresentam uma tentativa de estabelecer um modelo geral e completo para esta área.

Este modelo possui oito componentes: um conjunto de unidades de processamento, um estado de ativação, uma função de saída, um padrão de conectividade, ou seja, a matriz que representa todas as ligações entre as unidades; uma regra de propagação que combina as saídas das unidades e os pesos das ligações para produzir entrada para uma unidade; uma regra de ativação para combinar as entradas incidentes numa unidade através da regra de propagação e do estado de ativação da unidade; uma regra de aprendizagem, cujas mudanças ocorrem por experiência, e um ambiente dentro do qual o sistema opera.

De acordo com McCLELLAND [32], na visão do Processamento Distribuído e Paralelo, as unidades representam hipóteses e as ligações são as restrições. Por isso, os sistemas ou redes são ditas redes de satisfação de restrições. Então, se determinadas hipóteses ocorrem devem existir ligações positivas entre as unidades que as representam, caso contrário não.

Deste grupo, que é um dos maiores incentivadores de procedimentos baseados no paradigma biológico, têm surgido vários modelos para o processamento da informação. Entretanto, o objetivo de torná-lo paralelo, ainda não foi alcançado. Um limitador para esta perspectiva é o fato de ter que se valer de simulações, pois máquinas que tenham a arquitetura ideal para esta teoria ainda não estão totalmente disponíveis.

Sendo PLN uma das aplicações mais desenvolvidas, a maioria dos modelos têm focalizado o problema de análise sintática e atribuição de papel semântico, utilizando a teoria de casos de FILLMORE [10] ou redes semânticas e do lado conexionista, a inibição lateral e a “Spreading Activation”.

III.3 A Nova Perspectiva para PLN

Dois fatores psicológicos conduzem à idéia de tratar PLN através de modelos conexionistas, além da já mencionada inspiração no modelo biológico.

Um deles, é que mesmo antes da expansão conexionista, a maioria dos lingüistas computacionais e psicolingüistas já acreditavam que os conhecimentos lingüísticos não se apresentavam separadamente num processo de comunicação, como representado no processamento seqüencial. Conexionismo, naturalmente, abandona esta forma de processamento, pretendendo que todos os processos relacionados aos diferentes tipos de conhecimento, ocorram simultaneamente como na mente humana.

A teoria psicológica de QUILLIAN ("Spreading Activation") resumida e discutida em [17], já é uma tentativa de introduzir paralelismo em PLN, e que foi introduzido antes deste renovado interesse pelo conexionismo. Atualmente, têm sido empregada em vários modelos desta linha. Esta teoria baseia-se em algumas hipóteses sobre a estrutura e funcionamento da memória semântica humana. Mas, a principal hipótese diz que seres humanos ativam todas as interpretações possíveis para uma dada sentença em paralelo e a mais comum prevaleceria sobre as outras.

O outro, mencionado em [44], é que a abordagem tradicional tratou a representação dos conceitos com estruturas de alto nível, como por exemplo, os "scripts" e "plans" de SCHANK [6,45]. Aqui, é possível utilizar características muito simples através da representação distribuída.

Entretanto, apesar desta crítica, um exemplo deste tipo de representação pode ser observada no sistema MARGIE de SCHANK et. alia. [6], um modelo clássico de Inteligência Artificial, realizado antes da expansão conexionista. Como pode ser notado da seção II.2, houve uma tentativa de representar substantivos através de características simples.

Estes fatos demonstram que há plausibilidade psicológica nas propostas conexionistas, embora as teorias não sejam novas e partam de conclusões também antigas. Mas, a principal questão que divide as duas linhas é a utilização de regras, não só para PLN,

como para toda área da cognição.

Alguns seguidores da abordagem conexionista acreditam que regras não são necessárias e que poderiam ser excluídas desta teoria. Argumentos como, o emprego de regras foi muito influenciado pelo tipo de processamento cabível nas máquinas disponíveis, isto é, o processamento simbólico, alimentam tal posição. Outros, entretanto, defendem a possibilidade de introduzi-las em modelos conexionistas, como é o caso de SELMAN [46].

Em [47] tenta-se demonstrar que modelos conexionistas para linguagem natural não precisam utilizar regras. E que apesar de desmotivarem o seu uso, os pesquisadores desta linha não conseguiram explicar os processos envolvidos em linguagem sem usá-las de alguma maneira. Acrescentam ainda, a força do conexionismo está na capacidade de associação das unidades de processamento.

Como citado por GASSER et. alia. [48], um outro problema pouco explorado em modelos conexionistas é a questão de geração de linguagem. Isto deve-se a dificuldade de tratar esta parte de PLN com paralelismo, já que este destrói a ordenação das sentenças, propriedade essencial para esta atividade. Não apenas para esta, como também para análise, mas neste caso tem sido tratada de alguma maneira através da adição de uma subrede que garante informações de seqüencialidade.

O impasse permanece, pois as teorias lingüísticas fazem crer que linguagem é fundamentalmente dirigida por regras. Porém, alguns adeptos do conexionismo defendem a idéia de serem as regras uma conseqüência dos processos de aprendizagem, que devem estar incluídos nestes sistemas. Esta parece a melhor postura a adotar neste caso.

III.4 Primeiros Modelos

Um modelo que serviu de inspiração para alguns trabalhos em linguagem natural, foi o desenvolvido por McCLELLAND et. alia. [32]. Neste tem-se uma rede para processamento de palavras, formada de três camadas: na primeira, as unidades representam as

características de cada letra, a segunda tem-se as letras e na terceira as palavras. As características que não fazem parte de determinada letra, inibem a unidade que a representa. Assim, desta maneira também as letras inibem as palavras a que não pertencem. E nestes dois níveis, de letras e de palavras, há inibição lateral, as unidades se inibem mutuamente. Com este modelo foi possível constatar que o contexto facilita a identificação de uma letra, isto é, uma letra é mais facilmente reconhecida quando dentro de uma palavra, do que isoladamente.

Outro modelo de McCLELLAND [49], que visa a aquisição de passados de verbos do Inglês e também aprendizagem das regras para este procedimento, foi analisado por PINKER et. alli. [19]. Entre outras, apontaram as seguintes falhas: não pode representar certas palavras, aprender muitas regras, não exibe diferenças entre formas regulares e irregulares e pode aprender regras que não são encontradas em linguagem humana. A conclusão a que chegaram é que a crença de que regras possam ser dispensadas com explicações na Psicologia da Aprendizagem, como quer a abordagem conexionista, deve ser abandonada, e que ao contrário, os fatos lingüísticos fornecem boa evidência para tais regras.

O modelo mais significativo de McCLELLAND [50], no entanto, é um para atribuição de papel semântico a constituintes de uma sentença, que tem muitas propriedades em comum com o anterior. É um modelo distribuído, que contém dois conjuntos de unidades, um representa a estrutura de superfície da sentença e o outro, a sua estrutura de caso. O modelo aprende através de pares de estrutura de superfície/estrutura de caso.

Pretende-se mostrar que é possível selecionar significado adequado para palavras ambíguas, através de contexto; selecionar "frame" de verbo apropriada utilizando suas características semânticas; completar sentenças com valores plausíveis e generalizar o conhecimento sobre atribuição de papel semântico correto.

As sentenças são constituídas de um verbo e três sintagmas, onde um é sempre um sujeito. O formato de entrada das sentenças é de acordo com uma lista de características, que pode em geral, ser produzida por um analisador sintático e um componente léxico. Mas, estes não são partes do modelo.

A ativação das unidades de uma estrutura de sentença pode ser através de uma

regra de ativação determinística ou também por um método que se baseia em probabilidades.

A principal diferença deste modelo com os demais que surgiram para esta mesma tarefa é a utilização da representação distribuída. Como não inclui um procedimento para análise sintática e nem um léxico, a entrada das sentenças preserva a ordem dos constituintes.

O mesmo não ocorre no modelo de COTTRELL [36]. Um primeiro modelo conexionista completo para PLN foi desenvolvido com motivações no modelo de reconhecimento de palavras discutido acima. A sua forma é mostrada na figura III.2. O nível LÉXICO consiste de uma unidade para cada palavra na linguagem. A sentença de entrada ativa as unidades do LÉXICO de forma seqüencial.

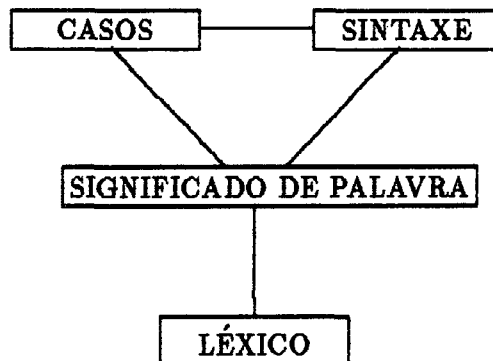


Figura III.2: Um modelo para um sistema conexionista.

As ligações existentes do LÉXICO para o nível de SIGNIFICADO DE PALAVRA, que tem unidades para representar todas as definições ou significados das palavras, são unidirecionais. As unidades deste segundo nível se ligam simetricamente a unidades dos níveis CASOS e SINTAXE. O nível SIGNIFICADO DE PALAVRA é um conjunto de redes, onde cada uma possui unidades para representar cada palavra naquela posição da sentença. Nestas redes há inibição lateral, que também é observada no nível CASOS.

É utilizado a teoria de casos de FILLMORE [10], porém, isso retira qualquer

necessidade de um componente sintático dentro deste contexto. É o que na verdade se faz, pois teoria de casos fornece informação bastante, para o contexto deste modelo, sobre as relações funcionais entre os constituintes da sentença, desde que estabelece o 'sujeito' (caso agente), e os vários tipos diferentes de 'objeto' (caso objetivo). Portanto, o que existe na parte intitulada SINTAXE deste modelo é apenas uma subrede que garante a seqüencialidade das sentenças de entrada, pois as posições em que ítems como 'preposição' ocorrem, é significativa para a retirada de ambigüidade. A entrada seqüencial não foi evitada pois tendo-se apenas uma entrada léxica para cada palavra no componente LÉXICO não seria possível paralelismo. No caso da sentença 'O lenhador matou o lobo', a palavra 'o' aparece duas vezes, se isto ocorresse em paralelo, as duas ativariam a mesma unidade no LÉXICO e não seria estabelecido suas posições na sentença. Assim, a entrada seqüencial deixa intacta a ordenação da sentença. E o procedimento que se tem em SINTAXE, não poderia ser comparado com o que é conhecido, tradicional ou intuitivamente, como análise sintática.

No entanto, COTTRELL apresenta uma versão deste trabalho em [51], onde inclui um procedimento melhor para análise sintática. Este, não é poderoso, embora manuseie construções recursivas, que são fixas por um comprimento máximo. Na verdade, não se tenta fazer com que este tenha realidade psicolingüística ou neurolingüística, e nem abrange grande quantidade de fenômenos.

Este analisador ainda utiliza a rede para seqüenciamento da entrada. Apenas inclui alguns níveis nesta para distinguir as categorias sintáticas. Esta maneira simples de tratar sintaxe deve-se ao reconhecimento de que apenas informações de posição são necessárias para retirar ambigüidades e que algo mais complexo, como uma árvore de derivação sintática seria dispensável.

No modelo que propõem WALTZ et. alli. [52] também tentam produzir um sistema completo para PLN. As fontes de conhecimento se interligam de forma modular, mas o processamento é interativo. Neste, são utilizadas a ativação espalhada ("Spreading Activation") e inibição lateral. A ativação espalhada se dá através de um algoritmo de busca em largura, visto em [16] e também por distribuir energia de ativação pela rede. Porém, o primeiro encontrará caminhos desinteressantes e o segundo provocará ativação uniforme da rede, por isso a inibição lateral é utilizada.

São observados quatro níveis na rede: o nível de análise sintática, a entrada, o léxico, que é uma rede com inibição lateral entre as unidades que representam significado, que por sua vez se ligam a unidades para categorias sintáticas e o nível contextual, que consta apenas de unidades representando os diferentes significados dos verbos.

Este último se liga ao léxico que se liga aos demais. É formada uma árvore de derivação sintática no primeiro nível citado acima. No final tem-se uma única interpretação para a sentença de entrada, cujas as unidades são ativadas de forma seqüencial através de uma rede que controla este processo, como também ocorre em [36,51].

Um modelo para análise sintática baseada em regras dado por FANTY [53], está inspirado num esquema determinístico. É construído um algoritmo para transformar uma gramática livre de contexto dentro de uma rede conexionista, que foi executado em um simulador síncrono. Analisa cadeias de tamanho arbitrário, mas fixo, explora a retirada de ambigüidade e aprendizagem de novas produções. Garante que a rede trabalha para todas as entradas sendo conceitualmente simples e rápida. Análises conflitantes não se inibem, todas são processadas em paralelo. Isto exige um grande número de unidades.

A rede contém unidades representando os terminais e não-terminais da gramática (várias para cada) e também unidades de instanciação, que representam as várias instâncias para as produções. Pode-se conjecturar que as unidades estão organizadas numa tabela, onde colunas representam as posições de entrada na sentença e as linhas os comprimentos. Há uma unidade para cada símbolo terminal em cada posição da coluna e uma unidade para cada não-terminal em toda posição da tabela.

Unidades terminais são ativadas por alguma fonte externa e representam a entrada para o analisador. Uma unidade não-terminal se tornará ativa se outras unidades representando o lado direito de uma de suas produções, e tendo posições e comprimentos iniciais apropriados estão ativas. Primeiro, a análise ocorre de forma ascendente, dando uma pequena ativação para as unidades ("priming"), depois de forma descendente, as unidades pertencentes à árvore de derivação correta se tornam hiperativadas. No início do processo ativam-se os símbolos terminais que representam a entrada e uma unidade que determina o fim desta, e esta o símbolo inicial.

FANTY [53] afirma que algumas ambigüidades serão retiradas naturalmente,

desde que nem todas as unidades se ativarão durante o processo. Mas, quando há evidência de ambigüidade, mais de uma árvore de derivação ficará ativa. Para evitar que isto ocorra, utiliza inibição lateral entre as unidades de instanciação.

A aprendizagem neste modelo se baseia em entradas incompletas, que podem ser completadas com um comprimento máximo fixo. A partir disso é possível que surjam novas produções compostas de constituintes conhecidos. Contudo, não é objetivo ter um mecanismo adequado para aquisição de gramática.

O procedimento de SELMAN[†] [46] é uma proposta mais ambiciosa que a apresentada acima. Tenta construir as árvores de derivações corretas a partir de regras, sem que se tenha todas representadas separadamente como é o caso deste modelo. Contudo, não tem nenhum processo de aprendizagem, embora se proponha a resolver algumas ambigüidades.

Uma teoria interessante é a passagem de marcadores ("marker passing"), que embora não seja propriamente conexionista, tem sido bastante usada nestes modelos. De acordo com esta, para processar informação numa rede, é necessário ativá-la várias vezes até que todas as unidades relacionadas à entrada estejam ativas, ou seja, a cada passo do processo ativa-se a rede marcando as unidades que se referem à entrada, isto prossegue até que não se possa marcar mais nenhuma unidade corretamente com a entrada. É um processo muito lento. Esta é uma forma de "Spreading Activation" [6], sobre a qual já se falou anteriormente.

Em geral, pode-se ver este processo implementado em redes semânticas, como mostra o modelo de SUMIDA et. alli. [54]. Na verdade, tomam a teoria da Dependência Conceitual proposta por SCHANK [6,45] e organizam conhecimento sintático e conceitual dentro de uma rede semântica. O processamento segue aquele dos modelos conexionistas. Isto é, aos nós que representam alguma daquelas ações, segundo a classificação de SCHANK, são associadas informações sintáticas como sujeito, objeto, além de preposições. Os rótulos das ligações seriam aqueles de uma rede semântica comum, tomando a direção de forma a obedecer estes tipos de relação.

Empregando também uma rede semântica, GASSER [48] apresenta uma proposta para geração de linguagem, onde a rede construída é formada de subredes que representam

[†] Será apresentado e analisado detalhadamente no próximo capítulo.

esquemas como aqueles de SCHANK. Também a ativação espalhada é empregada, junto com inibição lateral para tomar decisões entre os esquemas. Cada um destes, contém um nó representando a frase toda, um ou mais nós para os constituintes e um ou mais para aspectos semânticos e pragmáticos. A entrada para a geração consiste de um conjunto de nós ativados representando um evento, quando a ativação espalhada termina a rede converge para um padrão geral representando este evento. O principal objetivo é mostrar como fazer geração sem utilizar informação de sequenciamento, como mencionado a ordem dos constituintes é de extrema importância para este processo. Aqui, é tratado apenas em termos das características básicas dos modelos conexionistas. É garantido que funciona bem para pequenos exemplos.

Pode-se notar, a partir destes poucos exemplos, que o grande problema dos modelos conexionistas é justamente o procedimento que tenta evitar os modelos seqüenciais de PLN. Pois, ao tentar o paralelismo, tem-se a conseqüente perda da ordenação das palavras, que é fundamental para geração e análise, e em especial para a retirada de ambigüidades. Por isso, a maioria dos modelos ainda se valem de procedimentos que garantam a seqüencialidade. Além disso, a integração entre todos os componentes de um tal sistema assim feito, também torna-se complexo. Estas considerações podem justificar a descrença de alguns quanto às contribuições significativas dos modelos conexionistas para PLN e Ciência Cognitiva.

Capítulo IV

O Modelo

IV.1 Projeto da Rede

A especificação de uma rede neuronal para análise sintática baseada em regras, a partir do modelo exposto por SELMAN em [46] será apresentada em duas partes: *topologia* e *síntese*[†].

O objetivo pretendido com este modelo ou sistema é que, ao final do processamento ou simulação da rede construída segundo as regras de uma gramática livre de contexto, se obtenha a árvore de derivação sintática da sentença de entrada.

Além deste, existe um objetivo de caráter bastante geral. Pretende-se mostrar que é possível processamento de regras num modelo conexionista, por isso tenta-se ao máximo preservar os méritos de tal processamento. Isto impõe uma grande controvérsia, desde que as bases conexionistas se contrapõem às simbolistas, e estas enfatizam a utilização de regras.

Embora tenha sido usada uma gramática livre de contexto, SELMAN afirma que seria possível utilizar Gramática Transformacional, e que isto apenas aumentaria a complexidade da topologia da rede. Mas, de tentativas de construir tal topologia com

[†] Esta é a denominação para a tarefa de atribuição de pesos e limiares, que será usada na tentativa de formalizar a terminologia

gramática livre de contexto, é visto que seria altamente impraticável a construção manual de uma rede com Gramática Transformacional.

A representação localizada, diz o autor foi a escolhida por permitir expressar as regras de forma mais direta do que a representação distribuída.

Da forma como foi apresentado é um modelo bem empírico, por não tratar formalizações explícitas das regras para determinar a topologia e a síntese da rede. Por isso, pretende-se acrescentar algumas modificações e sugestões para facilitar a compreensão e a utilização do modelo.

SELMAN em [46] não discute a possibilidade de integração de um analisador sintático assim feito com os demais componentes de um Sistema de PLN. Mas, em [55] comenta a intenção de introduzir atribuição de casos visando tratamento de ambigüidade a nível semântico. O tratamento de ambigüidades é um de seus objetivos mais específicos, aliás, como em todo modelo conexionista para análise sintática de Linguagem Natural, porém aqui será tratada apenas as existentes entre classes gramaticais. Além disso, também será discutida uma possível conexão entre os outros componentes mais comuns.

Assim, a seguir é apresentado o modelo original, uma rede especificada dessa maneira, as sugestões e modificações introduzidas e por fim o processamento e os resultados obtidos através desta e de outras redes, inclusive a rede dada por SELMAN [46].

IV.1.1 Topologia

De acordo com o exposto acima, na especificação de uma rede segundo a proposta de SELMAN [46] são considerados dois fatos básicos: uma gramática livre de contexto e a representação localizada. Assim, cada símbolo da gramática é representado por uma unidade e as regras são expressas nas ligações entre estas unidades.

Os símbolos terminais e não terminais ou variáveis serão representados por unidades denominadas *unidades principais*. Na rede se distingue duas camadas, a *camada de*

entrada e a camada de análise sintática, ligadas entre si.

A camada de entrada é composta por unidades principais que representam os terminais da gramática. A disposição destas unidades se dá segundo uma divisão em grupos, onde o *i*-ésimo grupo está associado com a *i*-ésima palavra da sentença de entrada. Assim, o número de grupos fica determinado pelo comprimento da maior sentença reconhecida pela gramática escolhida. Para cada grupo há unidades representando todos os terminais da gramática. Isto permite que a entrada seja feita completamente em paralelo.

A gramática que usa para exemplificar é retirada de WINOGRAD [7] e supõe recursão, mas que na verdade não pode ocorrer em tal modelo, dado que o número de grupos é fixo pelo comprimento da maior sentença da gramática. Recursão não é, no entanto impossível, mas deve ser limitada.

A camada de entrada de uma rede construída a partir da gramática da figura II.2, deve conter seis grupos, pois sua maior sentença, que tem a forma geral 'det' 'sub' 'verbo' 'prep' 'det' 'sub', é formada por seis palavras. E como há quatro tipos de terminais diferentes, 'sub', 'det', 'prep' e 'verbo', esta camada consta de vinte e quatro unidades principais.

A camada de análise sintática, onde se encontra a rede propriamente dita, é constituída de pequenas subredes representando as regras, chamadas *primitivas conexionistas*. Na figura IV.1, as primitivas para as regras, $SN \rightarrow SN1$ e $SN \rightarrow det\ SN1$.

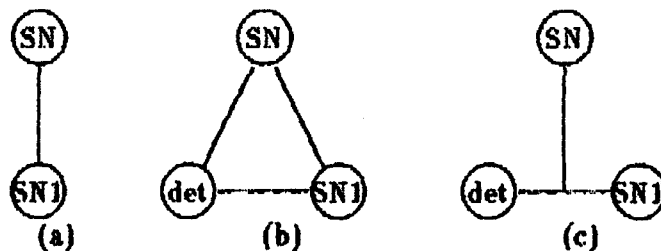


Figura IV.1: Primitivas conexionistas com duas (a) e três (b) unidades.(c) simplificação de (b).

Destas primitivas constrói-se a rede para todas as regras da gramática. Contudo,

apenas o uso de primitivas conduziria à enumeração de todas as árvores de derivação sintática, isto é, a rede seria um conjunto de subredes, que se inibiriam mutuamente em algum nível da rede e seriam unidas talvez pelo símbolo inicial. Mas, esta enumeração não permitiria a retirada de ambigüidade, pois a existência de estruturas comuns facilitaria a ativação de várias subredes simultaneamente. Além disso, o número de unidades seria muito grande. Um compartilhamento das unidades de uma primitiva reduziria estas desvantagens.

Este compartilhamento é possível com a introdução de outro tipo de unidade, *unidades intermediárias* ou *de ligação*, que funcionam como uma *ponte* entre as unidades da primitiva e permitem a escolha entre regras. E isto se dá através de uma ligação inibitória existente *apenas* entre estas unidades. Na figura IV.2 um exemplo, o compartilhamento para as primitivas da figura IV.1.

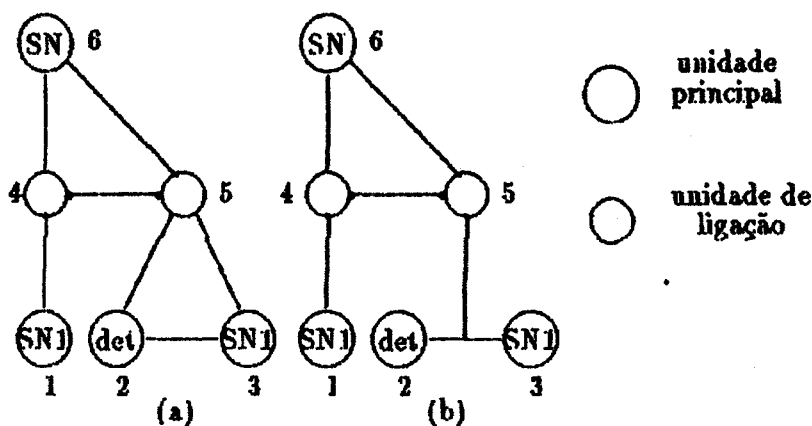


Figura IV.2: (a) Compartilhamento para as primitivas da figura IV.1 e (b) sua simplificação.

No entanto, apenas tomar as regras da gramática não é suficiente. Há uma ligeira diferença entre analisar as regras da gramática e suas derivações ao se tentar construir uma rede. Pois, o objetivo é ter todas as árvores de derivação representadas na rede, então é necessário considerar não só as regras, mas principalmente as próprias derivações. Deve-se, a princípio, verificar os pontos comuns entre todas as derivações possíveis da gramática e traçar o compartilhamento a partir destas regularidades. Por exemplo, tomando todas as derivações da gramática da figura 11.2, um compartilhamento seria como na figura IV.3,

Nota-se que uma única unidade pode representar o SN no início da sentença sendo compartilhada pelas dez primeiras derivações. E há uma unidade de ligação para os SN's iniciais das derivações a), b), c), d) e e) e outra para f), g), h), i) e j). Para os SN finais e os SNP

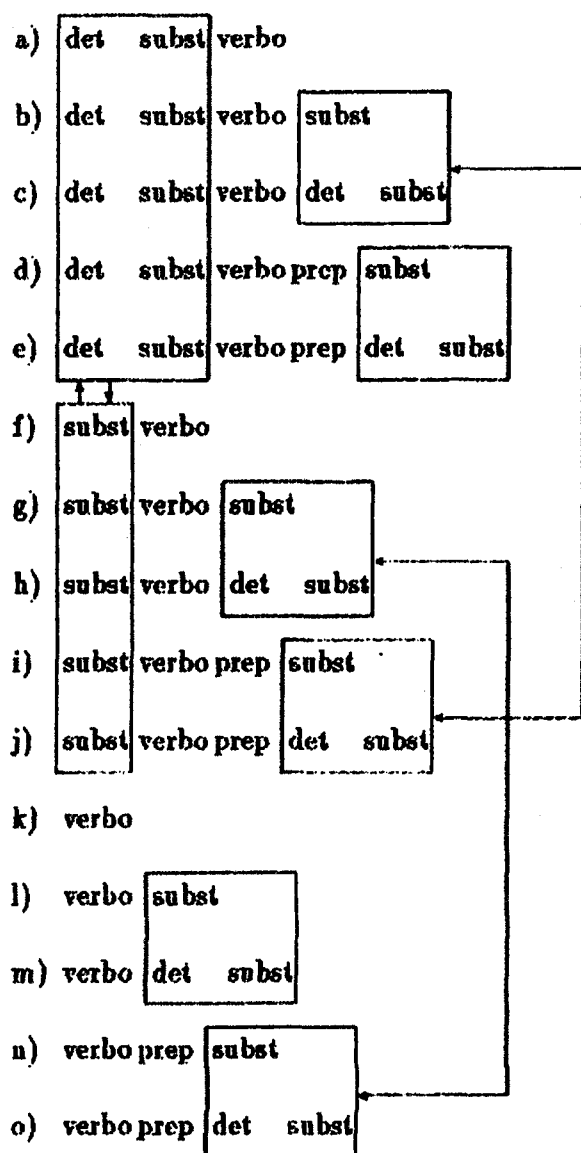


Figura IV.3: Um compartilhamento para as derivações da gramática da figura II.2.

observe que b) e c) compartilham com i) e j) o mesmo SN. As duas primeiras se ligam a 'verbo' do terceiro grupo de entrada e as outras a 'prep' do mesmo grupo. Isto também ocorre com g) e h) & n) e o), só que as ligações são com 'verbo' e 'prep' do segundo grupo. Apenas d) e e) não compartilham o SN, a não ser entre elas mesmas. Além disso, cada unidade para os terminais 'verbo' nos três primeiros grupos de entrada se ligará de duas formas: a um SN e a um SNP (sintagma preposicional) e, também haverá a opção de não

se ligar a nenhum complemento no final da sentença.

SELMAN diz que uma análise das possíveis combinações das categorias sintáticas deve ser feita. Mas, não apresenta nenhuma regra ou método para tal tarefa, assim deve-se tentar vários compartilhamentos, não sendo possível um conhecimento prévio de qual seja o melhor. A maneira sugerida aqui é tomar todas as derivações, como anteriormente mostrado, e através do processamento alguns compartilhamentos se mostrarão inviáveis. Isto desmotiva a utilização deste modelo, desde que a construção de uma rede se dá pelo método da tentativa e erro.

A construção de uma rede através de tal modelo com esta topologia pode não ser imediata, pois além de depender da complexidade da gramática, a diversidade de compartilhamentos possíveis introduz um alto grau de liberdade, podendo gerar várias redes diferentes para a mesma gramática. Algumas considerações de ordem prática podem ser seguidas tanto para facilitar a compreensão e legibilidade da rede quanto para seu processamento.

SELMAN em [46] sugere que se deve tentar diminuir o número de SV's, apoiando isto nas plausibilidades biológica e psicológica, que diz que são poucos os conceitos de mais alto nível no cérebro e mente humanos. Assim, os conceitos mais simples seriam em maior quantidade e, portanto poderia haver várias unidades para representar os outros símbolos, como SN, SN1 e SNP. Além disso, SELMAN acrescenta que o comprimento da cadeia de não terminais nas árvores de derivação deve ser menor que o número de unidades na camada de entrada. Mas, isso deve ocorrer normalmente.

Assim, de acordo com as observações da figura IV.3, foi construída aqui, a rede da figura IV.4. Algumas considerações sobre esta rede se fazem necessárias. A camada de entrada não está completa, como já mencionado deveria ter vinte e quatro unidades, mas estão representadas apenas as que admitem ligações (dezessete), as demais estariam presentes sem que se ligassem a qualquer unidade. Primeiro foram enumeradas as unidades principais e por fim as de ligação. A enumeração da camada de entrada está obedecendo à ordem utilizada no processamento, cujos os grupos estão separados por um pequeno traço, sendo da esquerda para a direita, a enumeração destes grupos.

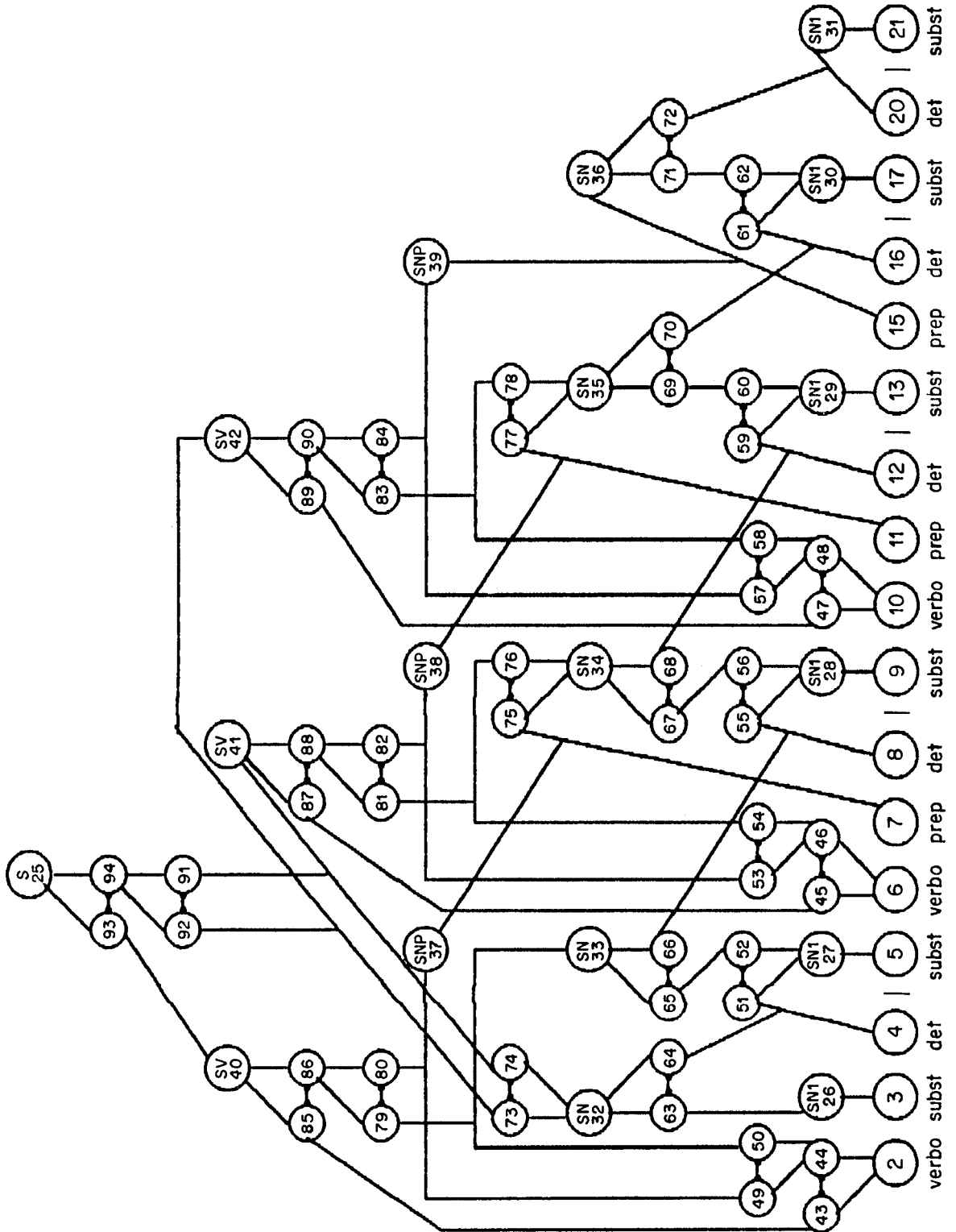


Figura IV.4: A rede construída para gramática da figura II.2

Antes de chegar a tal topologia, algumas outras foram testadas, principalmente,

tentando minimizar o número de unidades SV's, porém não foi possível reduzi-las a um número menor do que o apresentado e ter um bom processamento. Intuitivamente, parece razoável uma tal estrutura, que é bastante homogênea devido à alta frequência de determinados grupos de unidades. Mas, devido às dificuldades de se obter um conjunto de treinamento para redes que tratem Linguagem Natural, métodos e regras de aprendizagem ainda não surgiram. Portanto, para finalizar o projeto da rede resta atribuir limiares e pesos às unidades e ligações da rede, respectivamente. Através da análise da estrutura da rede é possível estabelecer relações entre pesos e limiares que fornecem regras para determinar seus valores.

IV.1.2 Síntese

Dado o esquema computacional da Máquina de Boltzmann (seção III.2), que é utilizado no modelo de SELMAN, tem-se que o objetivo principal é a minimização global da energia inicial. Isto foi utilizado para tratar esta parte do projeto da rede.

A análise ascendente e descendente não são evidenciadas, sendo argumentado que se realizam de forma integrada. A distinção entre os dois processos pode ser vista no procedimento apresentado por FANTY [52], que utilizou um esquema determinístico.

Assim, destas considerações e da figura IV.2(a) é possível determinar os pesos das ligações excitatórias. A unidade 4 está ligada apenas à unidade 1, que representa o lado esquerdo da regra, enquanto a unidade 5 está ligada a duas outras 2 e 3, ligações estas que devem ser iguais, isto é, $w_{25} = w_{35}$, pois constituem uma mesma regra. Também a ligação entre 2 e 3, w_{23} , deve ser igual às duas outras, pois isso garante a formação da regra, dado que todos os símbolos devem estar presentes. Como as duas regras têm probabilidades iguais de ocorrer w_{14} deve ser igual à soma de w_{25} e w_{35} . Como não se sabe qual a técnica, ascendente ou descendente, que será utilizada num dado instante do processamento, as ligações w_{46} e w_{56} são iguais à soma das demais ligações das unidades 4 e 5. Além disso, $w_{46} = w_{56}$, como já discutido. Depois destas considerações a figura IV.2(a) se transformaria na figura IV.5, sendo α , um valor positivo.

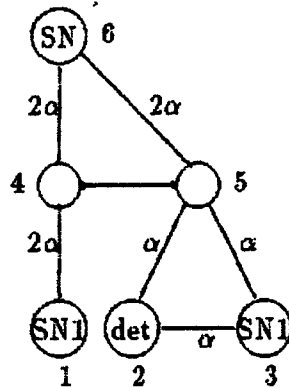


Figura IV.5: Atribuições de pesos às ligações excitatórias.

Dado que deve haver repetições de várias categorias sintáticas em posições diferentes da rede, a sua estrutura é bastante homogênea. Desta homogeneidade extraí-se alguns grupos sobre os quais é feita toda a análise para síntese da rede.

É argumentado que analisar pequenos grupos de unidades é no mínimo melhor que considerar as unidades individualmente para atribuição de pesos e limiares. Esta última gera muitos conflitos, pois a contribuição de energia de uma única unidade para minimizar a energia global pode não ser, e não será em geral, o mínimo de outras unidades. Então, o que se conseguirá com tal procedimento é um mínimo bastante grosseiro, distante do real. Portanto, a utilização de pequenos grupos garante uma aproximação menos grosseira, mas que contudo ainda não é o mínimo real da rede.

Além deste, um outro motivo discutido é a inviabilidade de fazer uma análise individual de cada unidade para redes muito grandes, que pode ser o caso deste modelo, desde que depende da gramática escolhida.

A princípio, os limiares das unidades foram considerados iguais a zero, mas tendo o modelo estados de ativação -1 e $+1$, tem-se que a contribuição de energia de uma unidade, segundo a equação (IV.1), com estado de ativação $+1$ é oposta, mas de mesma magnitude, à desta unidade com estado de ativação -1 .

$$E_{loc,j} = \left(-\frac{1}{2} \sum_i w_{ji} s_i + \theta_j \right) s_j \quad (\text{IV.1})$$

Desta simetria observa-se que a energia de uma primitiva connexionista com todas unidades ativas é a mesma quando todas unidades estão inativas. Isto significa que o sistema não

tem conhecimento prévio se a regra expressa nesta primitiva deve ou não fazer parte da árvore de derivação. Este conhecimento antecipado é verificado no modelo de FANTY [53], que primeiro utiliza a análise ascendente e depois a descendente.

Mas, quando se trata de pares de unidades de ligação, tal fato não deve ocorrer, pois duas regras que se opõem não podem estar presentes na árvore de derivação. Portanto, a energia de um grupo de unidades contendo unidades de ligação não pode ser a mesma estando as unidades ativas ou inativas, e isso ocorreria caso os limiares destas unidades permanecessem iguais a zero. Segue-se daí que é necessário reformulações na atribuição de limiares às unidades de ligação e determinação do peso da ligação inibitória, que originará regras gerais para desempenhar tal tarefa.

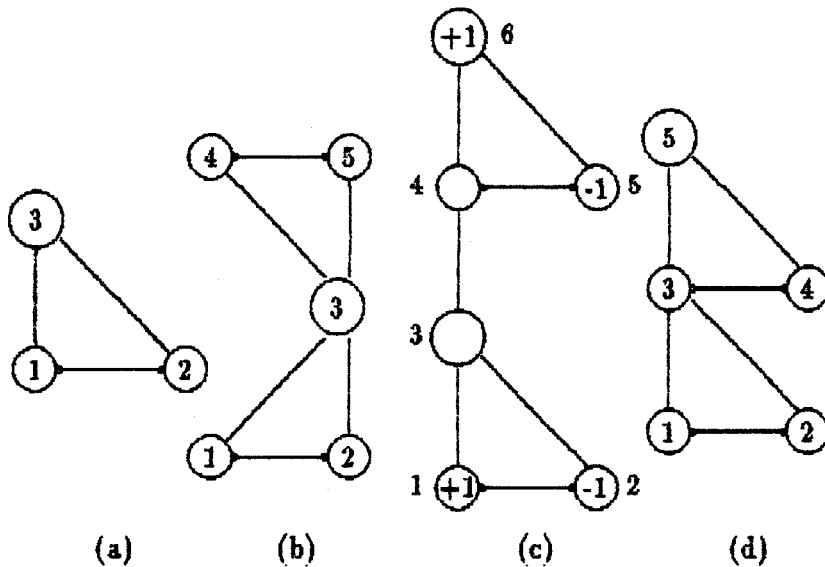


Figura IV.6: Quatro grupos básicos.

Na figura IV.6, estão representados quatro grupos básicos, sendo que apenas os grupos (a), (b) e (c) foram analisados por SELMAN [46], mas o grupo (d) foi acrescentado aqui por permitir observações interessantes a cerca do processamento da rede.

A análise tem como base dois fatos fundamentais e gerais. Primeiro, a energia de uma configuração com todas unidades ativas deve ser maior que a energia da configuração com todas unidades inativas. Isto porque este tipo de configuração (com todas as unidades ativas) não deve ser aceita, portanto impor que sua energia seja a maior evita que tal fato

ocorra. Não são feitas restrições quanto as demais configurações. Segundo, a energia da configuração com todas unidades inativas deve sempre ser menor que a energia de uma configuração admissível, pois uma configuração assim diz que nenhuma regra de um dado grupo pertence à árvore de derivação sintática da sentença de entrada, portanto deve permanecer inativo. Na verdade, este segundo fato não está explícito nas argumentações de SELMAN. No entanto, deve ter sido utilizado para concluir alguns valores que serão ora apresentados.

Para a atribuição do valor da ligação inibitória, considerou-se a figura IV.6(a), da qual se retirou quatro configurações conforme figura IV.7. É dito por SELMAN que através da energia de um par de unidades de ligação é possível deduzir que o peso da ligação inibitória é $-\beta$. De fato, tomando a energia da configuração B, nota-se que esta depende apenas da ligação inibitória. Sendo o valor desta $-\beta$, o da energia também o é

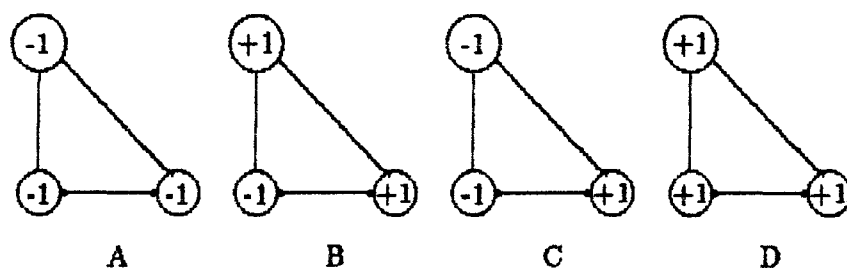


Figura IV.7: Configurações do grupo (a) da figura IV.6.

e vice-versa. Isto é uma tentativa de estabelecer as relações propostas. Mas, neste caso, poder-se-ia apenas dizer que o peso desta ligação é $-\beta$, desde que é sabido que esta ligação deve ter peso negativo, no princípio seria atribuído valor $-\beta$ e através das relações entre pesos e limiares sua magnitude ficaria estabelecida.

SELMAN afirma que β deve ser maior que 2α para impedir que as unidades de ligação de um par estejam ativas no mesmo instante. Assim, tem-se que $\beta > 2\alpha$ e está estabelecido o peso da ligação inibitória.

Permanecendo o limiar da unidade principal (θ_P) deste grupo igual a zero, resta saber qual o valor de θ_L , limiar das unidades de ligação. De acordo com as considerações acima, é necessário encontrar limiares tal que $E_A < E_D$, e também $E_A < E_B$, já que B é

uma configuração admissível. Para tanto tomou-se as equações de energia para A, B, C e D dadas abaixo:

$$\begin{aligned} E_A &= \beta - 4\alpha - 2\theta_L \\ E_B &= -\beta \\ E_C &= -\beta \\ E_D &= \beta - 4\alpha + 2\theta_L, \end{aligned} \tag{IV.2}$$

para que $E_A < E_D$, $\theta_L > 0$. Mas, $E_A < E_B$ então $\theta_L > \beta - 2\alpha$. Como $\beta > 2\alpha$, qualquer valor maior que $\beta - 2\alpha$ tem-se $\theta_L > 0$. Assim, considerou-se $\theta_L = \beta - \alpha$.

A configuração C foi acrescentada nesta análise, pois observou-se que embora seja inadmissível, tem energia igual ao de uma configuração admissível. Isto decorre do fato de que no trabalho original não se analisa todas as configurações do grupo para a síntese. Assim, para que C não ocorra seria necessário que θ_P (o limiar da unidade principal) fosse negativo, por exemplo $-\alpha$. No entanto, isto implicaria em $E_B = E_A$, as únicas duas configurações que devem ocorrer com este grupo. Claro que também a configuração similar a B, com a outra unidade de ligação ativa ocorrerá, contudo tem energia igual a de B.

Na figura IV.6(b), há uma unidade principal em meio altamente simétrico, se seu limiar permanecer igual a zero, a energia desta configuração com esta unidade ativa ou inativa será a mesma. Para evitar isso, SELMAN atribuiu a seu limiar um valor negativo, neste caso -2α . Mas, pode ser concluído utilizando as considerações acima, que deveria ser maior que -2α .

No modelo original, à unidade 3 da figura IV.6(c) foi atribuído valor -2α . Ao considerar aquela configuração, SELMAN afirma que as unidades 3 e 4 não se tornariam ativas quando limiar de 3 igual a 0 (zero). Mas considerando três configurações deste grupo (figura IV.8) observou-se que este valor pode ser alterado.

As equações de energia para estas configurações sendo o limiar da unidade 3 igual a zero são:

$$\begin{aligned} E_A &= -(2\beta + 6\alpha) \\ E_B &= -(2\beta + 2\alpha) \end{aligned}$$

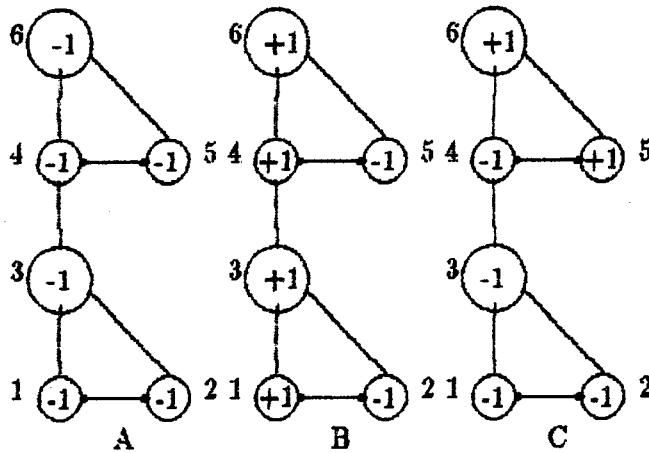


Figura IV.8: Configurações do grupo (c) da figura IV.6.

$$E_C = -(2\beta + 4\alpha) \quad (\text{IV.3})$$

Nota-se que $E_A < E_B$ e $E_A < E_C$ e isto deve ocorrer já que a configuração *A* deve ser a de menor energia dentre todas possíveis, pois este grupo está inativo. Além disso, $E_C < E_B$ significando que a configuração *C* tem mais chance de ocorrer se unidade 6 está ativa.

Tomando as energias destas configurações com limiar da unidade 3 igual -2α , as equações de energia para as mesmas configurações são:

$$\begin{aligned} E_A &= -(2\beta + 4\alpha) \\ E_B &= -(2\beta + 4\alpha) \\ E_C &= -(2\beta + 2\alpha) \end{aligned} \quad (\text{IV.4})$$

Assim, $E_A = E_B$ e $E_B < E_C$. Isto diz que a configuração *B* pode ocorrer sempre, pois sua energia é a menor energia, desde que é igual a de *A*.

Baseado nas energias destas configurações quando o limiar da unidade 3 é simplesmente θ_P , nas afirmações iniciais desta análise e nas equações de energia abaixo:

$$\begin{aligned} E_A &= -2\beta - 6\alpha - \theta_P \\ E_B &= -2\beta - 2\alpha + \theta_P \\ E_C &= -2\beta - 4\alpha - \theta_P, \end{aligned}$$

tem-se então, que $\theta_p > -2\alpha$. Logo, a atribuição de zero para o valor do limiar da unidade 3 neste grupo é válida.

Sob o ponto de vista prático, algumas vezes seria necessário ter o limiar igual a zero, em outras -2α . Isso na verdade depende em parte das outras ligações da unidade 5. O risco de usar -2α é que a energia deste grupo tenderá na maioria das vezes para esta configuração visto que é a de menor energia. É possível que valor zero não perturbe a estabilização do rede, pois também na prática resultados que contrariam a afirmação inicial foram obtidos.

Assim, generalizando, foram estabelecidas as seguintes regras para síntese. No entanto pelo que foi visto acima, estas não estão devidamente formuladas.

- peso da ligação inibitória é $-\beta$
- peso da ligação excitatória é α (em primitivas com três unidades)
- peso da ligação excitatória é 2α (em primitivas com duas unidades)
- limiar da unidade principal é -2α (em meio altamente simétrico, ligada a apenas pares de unidades de ligação, podendo ter ainda uma ligação com uma outra unidade de ligação que não pertença aos pares)
- limiar das demais unidades principais é 0 (zero)
- limiar da unidade de ligação é $\beta - \alpha$

Basta atribuir valores a α e a β , aqui foram utilizados os valores $\alpha = 1$ e $\beta = 3$, respeitando o fato de que $\beta > 2\alpha$. Assim, utilizando as regras acima tem-se todos os pesos e limiares para uma rede segundo a topologia descrita.

Sobre a figura IV.6(d) SELMAN diz que sua utilização seria útil para esquemas connexionistas gerais. Este grupo é uma junção de dois grupos da figura IV.6(a), sendo que, a unidade pela qual se dá a junção é uma unidade de ligação e não principal. No entanto, de algumas observações foi constatado que isto pode promover comportamentos indesejáveis da rede. É fácil ver que se quaisquer das duas unidades de cada um dos dois pares se ativar, as energias destas configurações serão iguais. Isto permitirá resultados conflitantes.

Para retirar este conflito, notou-se aqui, que o limiar da unidade 3, deveria ser menor que das outras unidades de ligação. Por isso, considerou-se seu limiar igual a 0 (zero), como se esta fosse uma unidade principal da figura IV.6(a), pois se esta unidade ficar inativa quando houver evidências para que 1 ou 2 se ativem, não implicará em problemas, contudo tal fato não deveria ocorrer. Isto pode ser visto a partir das configurações *B* e *C* da figura IV.6(a). Assim, ao considerar seu limiar igual a zero, permite-se que esta unidade torne-se ativa mais facilmente, já que uma configuração deste tipo terá energia menor. Observe as duas configurações da figura IV.9.

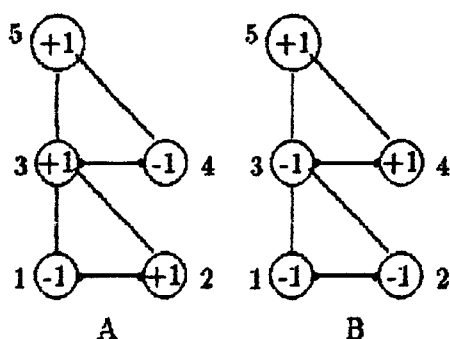


Figura IV.9: Configurações do grupo (d) da figura IV.6.

A energia da configuração *A* será menor que a de *B*, se limiar da unidade 3 for igual a zero. Mas, se este limiar for igual a $\beta - \alpha$, isto é, limiar de uma unidade de ligação, ocorrerá o contrário. Mais uma vez, nota-se que dependerá da conveniência de qual é melhor ocorrer na maioria das vezes. Aqui, utilizou-se zero, pois além de ser mais eficiente, permitiu atingir bons resultados com menos passos de simulação. Contudo, este grupo será utilizado para mostrar que tal síntese não conduz ao bom funcionamento da Máquina de Boltzmann e alguns resultados desejados só foi possível com a introdução desta modificação.

IV.2 Processamento da Rede

Dada estrutura da rede com todas as atribuições de pesos para as ligações e limiares para as unidades e o algoritmo da figura III.1, que foi implementado na linguagem PASCAL e executado em um IBM/4381, inicia-se o processo de simulação por ativar na camada de entrada os grupos correspondentes à sentença de entrada e também o símbolo inicial. Cada palavra da sentença ativa em seu grupo a unidade que possa representar sua classe gramatical, as demais unidades estão inativas. Todas as unidades da camada de entrada e o símbolo inicial não mudam seus estados durante todo o processo de relaxação.

Para o processo de *Simulated Annealing* foram tomadas as doze temperaturas de acordo com SELMAN [46], 10.000, 4.0, 2.0 até 0.2, com variação de 0.2. Da primeira para segunda fase há um resfriamento rápido e a partir daí torna-se lento. Isto restringe logo nas primeiras fases os possíveis mínimos entre os quais o sistema oscilará.

Uma fase, no algoritmo aqui implementado, se dá a cada γn^2 passos de simulação, onde γ é um inteiro maior que zero, estabelecido através de simulações. Este método foi adaptado de uma proposta para um algoritmo geral para o "Simulated Annealing" em ARAGON et. ali. [56]. De acordo com este algoritmo, é sugerido que em cada fase se utilizasse n^2 passos de simulação. Por isso, as tentativas em descobrir os passos de simulação que seriam suficientes para estabilização da rede foram realizadas com acréscimos de n^2 , pois nem sempre apenas este número é o suficiente. Em suas simulações, SELMAN utilizou 2.000 passos de simulação para cada unidade em cada fase do "annealing", sendo portanto utilizado 90.000 passos em cada fase. Não houve tentativas no sentido de minimizá-los e nem oferece um critério formal para estabelecer estes valores.

Em cada fase, SELMAN tomou a saída média de cada unidade, esta expressa a frequência de ativação da unidade, isto é, a unidade que permaneceu mais vezes ativa tem esta média positiva, caso contrário não. Nas últimas fases esta média se aproxima de +1 para as unidades que devem ficar ativas e -1 para as inativas. Assim, tomando as unidades que têm esta média próxima a +1, teria a árvore de derivação correta. O processo termina quando temperatura igual à 0.2, embora seja possível que em duas ou três fases anteriores o resultado da saída média represente a árvore de derivação correta. Por isso, em uma

revisão de sua proposta [55], SELMAN et. alia. diz que o ponto de congelamento do sistema poderia ser considerado em $T = 0.6$. Isto decorre de experimentos que mostram a possibilidade de se encontrar a saída média desejada nesta temperatura.

O que faz de fato, é não considerar a configuração final da rede como o resultado, pois esta poderá não representar a árvore de derivação desejada. Mas, em algumas fases anteriores poderá haver uma tendência para este resultado, por isso toma as médias das saídas. Aqui, entretanto, a configuração final foi considerada por acreditar ser este o objetivo de modelos conexionistas, verificar o estado da rede que constrói. E além disso, no caso de interligação entre os vários componentes de um sistema, seria necessário levar em conta a configuração da rede.

Foram simuladas três redes, com 27, 45 e 94 (figura IV.4) unidades. Sendo que a primeira e a terceira representam gramáticas com 3 e 15 árvores de derivação, respectivamente. A segunda é a rede dada em SELMAN [46,55], que é apenas uma pequena parte do que realmente deveria ser uma rede para a gramática lá considerada. Nesta tem-se quatro sentenças de entrada, onde apenas uma é ambígua.

Os passos de simulação exigido para cada rede ou entrada foram obtidos através de várias simulações, assumindo valores diferentes para γ . Para a primeira rede foram necessários $2n^2$ (1.458) passos, para a segunda $8n^2$ (16.200) e para a terceira $3n^2$ (26.508), onde n é o número de unidades, sendo que estes foram duplicados nas três primeiras fases. Com estes números conseguiu-se a simulação de entradas sem ambigüidades, ou ainda, entradas corretas.

As duas primeiras redes se comportaram conforme o proposto para todas as entradas. No entanto, das quinze entradas corretas para a terceira, apenas dez foram bem sucedidas. Nem mesmo aumentando o número de passos se garantiu bons resultados em todas as entradas, pois em geral, o aumento dos passos de simulação garante melhores resultados.

As sentenças que têm forma geral de acordo com: *det subst verbo prep subst*, *det subst verbo det subst*, *det subst verbo subst*, *det subst verbo e verbo*, representadas respectivamente, pelas entradas: 4, 5, 10, 15, 17 e 25; 4, 5, 10, 16, 17 e 25; 4, 5, 10, 13 e 25; 4, 5, 10 e 25 & 2 e 25, não foram bem sucedidas (vide Apêndice A.1.1).

Nas quatro primeiras, ao invés de se ativar a unidade 42 que representa o SV destas sentenças, tornou-se ativa a unidade 41, que representa um outro SV. Assim, também na prática, demonstra-se o que foi discutido em relação a figura IV.6(a), a unidade 42 não se ativa estando 90 ativa. Como uma das unidades SV deve estar ativa para minimizar a energia global, ativa-se 41. Mas para minimizar a energia deste grupo com esta unidade ativa, outras unidades pertencentes a ele devem se ativar. É o que ocorre, as unidades 81 e 88 se ativam. Também no grupo formado pelas unidades 32, 63, 64, 73 e 74, esta última se ativa, sendo que a 73 é que deveria estar neste estado, mas é visto que quaisquer das unidades dos pares que se ativarem a energia será a mesma. Isto não deveria ocorrer pois gera situações como esta. Para última também pode ser aplicada observação semelhante. As unidades 44, 50, 79 e 80 se ativam enquanto as corretas 43 e 85 não.

Um resultado que confirma as observações da seção IV.1.2 sobre o grupo (c) da figura IV.6, foi obtido simulando a segunda rede (exemplo de SELMAN) que contém um grupo deste, utilizando limiar igual a 0 (zero). Todas as árvores de derivações corretas ocorreram no final do processamento.

As ambigüidades que podem ser tratadas neste modelo, são as existentes entre classes gramaticais. Em especial, aquelas entre verbo e substantivo. Para retirar este tipo de ambigüidade, começa por ativar no grupo de entrada, correspondente a palavra ambígua, as unidades que representam estes itens gramaticais. A árvore de derivação sintática deve refletir a escolha correta. Foram testadas todas as possíveis ambigüidades da primeira rede, a sugerida por SELMAN em seu exemplo e também algumas da terceira.

As entradas ambíguas consideradas para a terceira rede foram: *det subst verbo verbo prep det subst*, *det subst subst verbo prep det subst*, *subst subst verbo prep det subst*, *verbo subst verbo prep det subst* e *verbo subst prep det subst*, cujas as entradas seriam como no Apêndice A.1.2.

Utilizando os mesmos passos de simulação anteriores para esta rede, os resultados são como segue. Para a primeira, o fato da unidade 6 estar ativa promoveu a ativação de parte de uma árvore de derivação relacionada a ela, precisamente 34, 46, 56, 67, 76, 81 e 88. Nota-se aqui, que o grupo formado pelas unidades 28, 55 e 56, ocorre o fato discutido na seção III.1.1 com relação às configurações *B* e *C* da figura IV.7. A unidade 56 se ativa enquanto 28 não.

Para a segunda, houve menos evidência para a árvore de derivação correta, se comparado com a primeira. Ocorreu uma árvore derivação completa para entrada relacionada à ambigüidade, mas que não era a esperada. Na terceira, valem observações similares àquelas da primeira. Para a quarta, uma derivação indesejada relacionada à unidade 2 quase se completa e mais uma vez nota-se as últimas observações feitas para a primeira. Na quinta e última, a derivação correta e uma errada ficam igualmente na rede, mas próximo ao símbolo inicial, há evidências para a configuração errada.

Observou-se que a rede pode estabilizar na árvore de derivação desejada, porém outras unidades, não pertencentes a esta, também estarão ativas no final do processamento. Especificamente, aquelas ligadas à unidade que representa a ambigüidade. Como já dito este fato justifica considerar as saídas médias das unidades. Contudo, saídas médias erradas também estarão próximas a +1 (testes feitos utilizando este processo confirmam isto, pelo menos para entradas sem ambigüidade). Sendo assim, o resultado dependerá de estipular o limite para esta aproximação.

Ao se pensar numa possível integração deste analisador com as outras partes de um sistema de Processamento de Linguagem Natural, isto seria um problema. As informações necessárias que seriam extraídas desta rede poderiam ser apenas as unidades principais da árvore de derivação que estão ativas, caso fosse utilizado o sistema de casos de FILLMORE [10] como o fez COTTRELL [36,51]. Assim, seria construída uma rede para a parte semântica do sistema completo e a rede proposta por SELMAN indicaria as posições de itens como preposição, que facilita a retirada de ambigüidade. Por exemplo, o SNP representado pela unidade 39 diz que a preposição pertence ao quinto grupo de entrada, isto é, é o quinto item da sentença, mas outras unidades conflitando com esta, poderão estar ativas igualmente, assim não seria possível a retirada de ambigüidade. Também a entrada paralela dificultaria a ligação com um componente léxico, que deveria ser sequencial para evitar a repetição de vários itens para as diferentes posições.

Embora resultados que confirmam a proposta de SELMAN tenham sido obtidos, é possível construir redes segundo este modelo e encontrar um mínimo menor que o proposto, e que não representa a árvore de derivação sintática desejada. Por outro lado, pode-se estabilizar num mínimo acima do pretendido e também não ter a solução esperada.

Então o fato de que para algumas entradas a rede não ter sido bem sucedida,

deve-se à síntese que não permite que a Máquina de Boltzmann funcione bem, ou seja, como afirma o próprio SELMAN em [55], pode haver um estado da rede em que uma configuração não admissível tenha energia menor que a proposta por seu modelo.

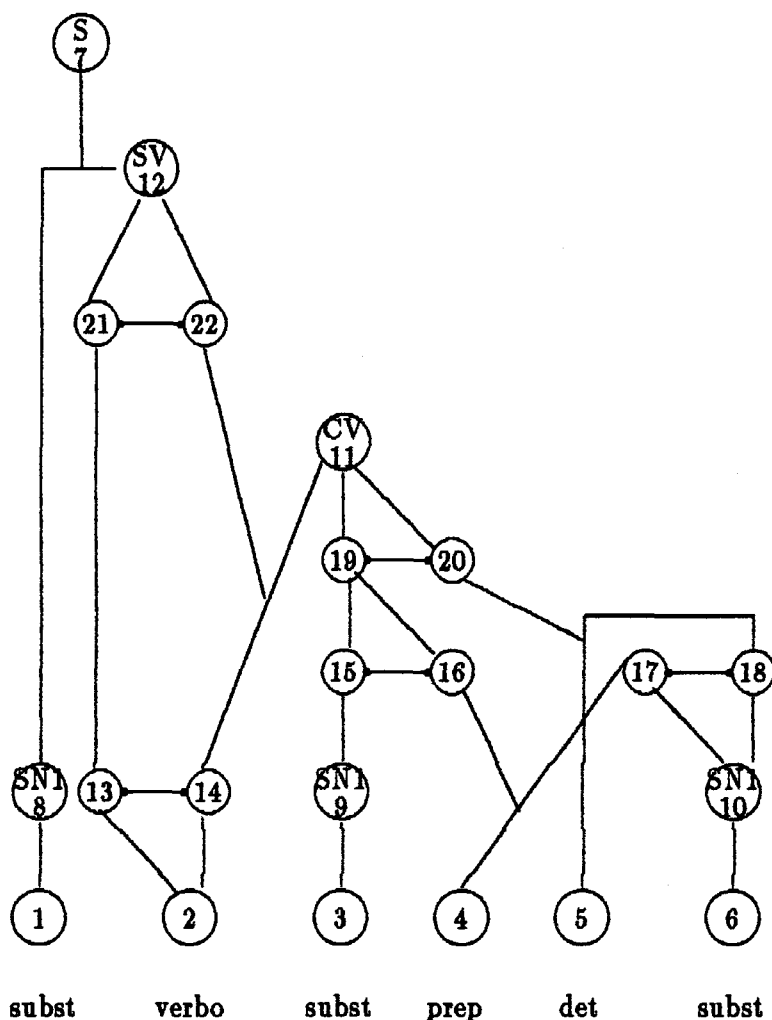


Figura IV.10: Contra-exemplo para o modelo.

Um exemplo de uma rede[†] em que o mínimo de energia não representa a configuração desejada, está na figura IV.10. Na verdade, o bom funcionamento para as algumas entradas nas outras redes e até mesmo nesta, deve-se à homogeneidade da rede. Como já dito várias redes poderiam ser construídas para uma mesma gramática, esta que é utilizada

[†] Estão representadas na camada de entrada apenas as unidades que admitem ligações.

neste contra-exemplo é apenas uma das possíveis redes para a gramática da figura IV.11.

S	→	SN1 SV
SV	→	verbo CV
		verbo
CV	→	prep SN1
		det SN1
		SN1
SN1	→	subst

Figura IV.11: Gramática para o contra-exemplo.

Assim, uma sentença de entrada que ativasse as unidades 1, 2, 4, 6 e 7, atingiria um mínimo de energia igual a -39, sendo as unidades ativas na camada de análise sintática 8, 10, 12, 13, 16, 17 e 21.

Contudo, não é a árvore de derivação correta para esta sentença de entrada. Esta deveria ser 8, 10, 11, 12, 14, 16, 17, 19 e 22, cuja energia é -37, maior que o mínimo atingido, pois este último é o mínimo global da rede com esta entrada. O mesmo ocorre para uma sentença dada pelas unidades 1, 2, 3 e 7. A energia da derivação correta, 8, 9, 11, 12, 14, 15, 19 e 22 é -39, mas estabiliza na configuração 8, 9, 12, 13, 15 e 21, que tem energia -41 (vide Apêndice A.2).

O que se pretende com tal modelo é que durante o processo de relaxação, ao tentar atingir o mínimo de cada grupo (figura IV.6), tenha-se no final todos estes mínimos em conjunto representando a árvore de derivação sintática correta de acordo com a entrada. Mas, isto pode ser interpretado como a soma dos mínimos ser o mínimo proposto para a rede. Pois a síntese foi feita isoladamente, considerando cada grupo em particular. Contudo, a soma dos mínimos não representa o mínimo global de energia.

Tomando os grupos formados pelas unidades 2, 13 e 14, e o outro por 12, 21 e 22 da figura IV.10, a soma de suas energias individuais, com uma das unidades do par ativa, é -2β . Mas, considerando este grupo como um todo, a energia seria $-2(\beta + \alpha)$. É portanto, menor que quando utilizando o procedimento proposto por SELMAN.

É proposto ainda em [46,55], uma alternativa para equação de energia, mas que

tornaria o processo mais lento, pois esta exigiria o cálculo da energia (equação IV.1) de cada uma das unidades em todo passo de simulação. Contudo, nenhum teste foi realizado aqui, utilizando esta outra opção.

Capítulo V

Conclusões

Em termos gerais, é ainda necessário esperar pela confirmação ou não das hipóteses biológicas que permitiram o desenvolvimento desta linha de pesquisa, e como uma consequência a efetivação da tecnologia que tornará viável a evolução das teorias cognitivistas compatíveis com esta abordagem. A partir destes pontos é que se poderá verificar a legitimidade e a validade do procedimento conexionista.

Especificamente, o que se tem atualmente, como proposta para o Processamento de Linguagem Natural, é que não pode resolver subproblemas que envolvem estes processos, nem mesmo no nível altamente questionável em que abordagem simbolista os tratou. E por isso outros problemas mais complexos como: contexto, representação do conhecimento do mundo e aprendizagem estão em aberto para qualquer uma das abordagens.

O modelo enquanto testa uma possível teoria para processamento de linguagem natural, em especial análise sintática, sob o ponto de vista do conexionismo mostra que há modelos simbolistas como a Gramática de Cláusulas Definidas e Redes de Transição Aumentadas, que tratam esta parte do conhecimento lingüístico de forma mais eficiente.

Contudo foi visto que processamento de regras é viável nesta teoria, apesar de algumas falhas do modelo e do elevado tempo de processamento. Este último aspecto não motiva a prática deste modelo, mas deve-se considerar que esta teoria foi bastante inspirada no crescimento da área de Processamento Paralelo, e que, portanto poderia ter melhor funcionamento se utilizando máquinas compatíveis para este processamento,

desde que suprimidas as falhas teóricas. Embora, o funcionamento ideal da Máquina de Boltzmann, ainda assim, não seria garantido. A máquina seqüencial utilizada permitiu alguns bons resultados.

A principal falha a se notar, é a síntese proposta como uma alternativa para a questão da aprendizagem em modelos conexionistas relacionados à linguagem. Esta não permite o bom funcionamento do esquema computacional utilizado. Como é afirmado pelo próprio autor [54], os resultados desejados a que se chega devem-se à homogeneidade da topologia da rede e não a sua síntese. Contudo, a topologia da rede não é muito clara, dado o grau de liberdade para sua construção. Também está claro que faltam formalizações a estas regras para síntese da rede, embora possam mostrar algum caráter geral.

Não se garantindo resultados exatos nem em entradas não ambíguas, é fácil concluir que ambigüidades em todos os grupos de entrada não possam ser resolvidas.

Porém, uma conclusão antiga é que análise sintática tão complexa, quanto a apresentada neste modelo, não seria necessária. Seres humanos utilizam mais as informações semânticas que as relacionadas às estruturas das sentenças. Os modelos conexionistas estão de alguma forma comprovando este fato, dada às dificuldades de fazer tal análise baseada em regras com a construção de árvores de derivação sintáticas, segundo esta abordagem.

Os procedimentos desta parte de um sistema para PLN deveriam estar mais direcionados a fornecer informações posicionais, que são fundamentais em alguns casos de retirada de ambigüidade, um dos grandes problemas nestes trabalhos.

Além destes procedimentos deveriam também ser pesquisados outros que evitassem a seqüenciamento ainda necessário em modelos conexionistas para PLN. Parece haver evidência psicológica e biológica o suficiente para se acreditar no paralelismo dos processos cognitivos relacionados à linguagem. É observado no modelo analisado uma tentativa de paralelismo, mas que não teria progresso ao ser unido aos demais componentes de um sistema de PLN.

E por último, aquele que talvez seja o de mais difícil tratamento, é o problema da aprendizagem em modelos conexionistas para PLN. A alternativa aqui proposta não surte bons efeitos, nem mesmo em alguns casos pequenos. Então, os trabalhos futuros em PLN deveriam, antes mais nada, se preocuparem com questões como aprendizagem e

sequenciamento, do que com análise sintática baseada em regras.

Referências Bibliográficas

- [1] STILLINGS, N. et alli., *Cognitive Science : An Introduction*, Cambridge, Massachusetts, MIT Press, 1988.
- [2] DREYFUS, H. L. e DREYFUS, S. E., "Making a Mind Versus Modeling the Brain: Artificial Intelligence Bock at a Branchpoint", in *DÆDALUS*, Vol. 117, No. 1, *Proceedings of the American Academy of the Arts and Science*, pp. 15-43, 1988.
- [3] MINSKY, M., *Semantic Information Processing*, MIT Press, pp. 1-29, 1968.
- [4] SLOBIN, D. I., *Psicolingüística*, Ed. da Universidade de São Paulo (EDUSP), 1980.
- [5] ALLEN, J., *Natural Language Understanding*, The Benjamin/ Cummings Publishing Company, Inc., Menlo Park, California, 1987.
- [6] SCHANK, R. C., *Conceptual Information Processing*, North-Holland Publishing Company, Vol. 3, pp. 1-21, 1975.
- [7] WINOGRAD, T., *Language as a Cognitive Process, Volum 1: Sintax*, Addison-Wesley Publishing Company, 1983.
- [8] CHOMSKY, N., *Aspects of Theory of Sintax*, Cambridge, Massachusetts, MIT Press, 1965.
- [9] PEREIRA, M. A. B., *Gênero e Número em Português*, Tese Msc., Ed. UFRJ, 1987.
- [10] FILLMORE, J., "The Case for Case", In *Universals in Linguistic Theory*, E. Back e R. T. Harms (Eds.), Chicago: Halt, 1968.
- [11] PITRAT, J., "Using Declarative Knowledge for Understanding Natural Language", in *Natural Language Parsing Systems*, L. Bolc (Ed.), Spring-Verlag, pp. 33-48, 1987.
- [12] CHARNIAK, E., "Passing Markers : A Theory of Contextual Influence in Language Comprehension". *Cognitive Science*. No. 7. pp. 171-190. 1983.

- [13] FREDERKING, R. E., *Syntax and Semantics in Natural Language Parsers*, Computer Science Department, Carnegie-Mellon University, TR CMU-CS-85113, May/1985.
- [14] WEIZENBAUM, J., "ELIZA: A Computer Program for the Study of the Natural Language Communications between Man and Machine", *Communications of the ACM*, Vol. 9, No. 1, pp. 36-45, 1966.
- [15] WOODS, W. A., "Transition Networks Grammars for Natural Language Analysis", D. G. Bobrow (Ed.), *Communications of the ACM*, Vol. 13, No. 10, pp. 591-606, Outubro/1970.
- [16] QUILLIAN, M. R., "The Teachable Language Comprehender: A Simulation Program and Theory of Language", D. G. Bobrow (Ed.), *Communications of the ACM*, Vol. 12, No. 8, pp. 459-476, Agosto/1969.
- [17] COLLINS, A. M. e LOFTUS, E. F., "A Spreading-Activation Theory of Semantic Processing", *Psychological Review*, Vol. 82, No. 6, pp. 407-428, 1975.
- [18] PINKER, S. e PRINCE, A., "On Language and connectionism: Analysis of a Parallel Distributed Processing Model of Language Acquisition", *Cognition*, No. 28, pp.73-193, 1988.
- [19] PEREIRA, F. C. N. e WARREN, D. H. D., "Definite Clausula Grammars for Language Analysis - A Survey of the Formalism and a Comparison with Augmented Transition Networks", *Artificial Intelligence*, No. 13, pp. 231-278, North-Holland Publishing Company, 1980.
- [20] FININ, T. W. e PALMER, M. S., "Parsing with Logical Variables", in *Natural Language Parsing Systems*, Leonard Bolc (Ed.), pp. 33-48, Springer-Verlag, 1987.
- [21] BRESNAN, J. e KAPLAN, R. M., "Introduction: Grammars as Mental Representations of Language", In: *The Mental Representation of Grammatical Relations*, J. Bresnan (Ed.), Cambridge, Mass., MIT Press.
- [22] MARCUS, M. P., "A Theory of Syntactic Recognition for Natural Language", in *Artificial Intelligence: An MIT Perspective, Vol. 1: Expert Problem Solving Natural Language Understanding Intelligent Computer Couches Representation and Learning*, P. H. Winston e R. H. Brown (Eds.), Cambridge, Massachusetts, MIT Press, pp. 191-229, 1979.

- [23] CARBONELL, J. G., e HAYES, P. F., "Robust Parsing Using Multiple Construction-Specific Strategies", in *Natural Language Parsing Systems*, L. Bolc (Ed.), pp. 1-32, Springer-Verlag, 1987.
- [24] HANKAMER, J. e AISSSEN, "Ambigüidade", In *Enciclopédia Einaudi*, Vol. 2, Linguagem-Enunciação, Lisboa, Imprensa Nacional-Casa da Moeda, pp.251-283, 1984.
- [25] DAHLGREN, K., "Using Commonsense Knowledge to Disambiguate Word Sense", in *Natural Language Understanding and Logic Programming II*, V. Dahl e P. Saint-Dizier (Eds.), North-Holland, pp.255-275, 1988.
- [26] SCHMIDT, R. F., *Neurofisiologia*, Ed. da Universidade de São Paulo, 1979.
- [27] RUMELHART, D. E. e ZIPSER, D., "Feature Discovery by Competitive Learning". In: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations*, J. A. Feldman, P. J. Hayes e D. E. Rumelhart (Eds.), pp. 151-194, MIT Press, 3a. ed., 1987.
- [28] PAPERT, S., "One AI or Many", In: *DÆDALUS*, Vol. 117, No. 1, *Proceedings of the American Academy of the Arts and Science*, pp. 1-14, 1988.
- [29] VEMURI, V., "Artificial Neural Networks: An Introduction", in *Artificial Neural Networks: Theoretical Concepts*, V. Vemuri (Ed.), *The Computer Society of the IEEE*, No. 855, Vol. 1, Computer Society Press, 1988.
- [30] FELDMAN, J. A. e BALLARD, D. H., "Connectionist Models and Their Properties", *Cognitive Science*, No. 6, pp. 205-254, 1982.
- [31] COWAN, J. D. e SHARP, D. H., "Neural Nets and Artificial Intelligence", In: *DÆDALUS*, Vol. 117, No. 1, *Proceedings of the American Academy of the Arts and Science*, pp. 85-141, 1988.
- [32] McCLELLAND, J. L., RUMELHART, D. E. e HINTON, G. E., "The Appeal of Parallel Distributed Processing", In: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations*, J. A. Feldman, P. J. Hayes e D. E. Rumelhart (Eds.), pp. 3-44, MIT Press, 3a. ed., 1987.

- [33] HURLBERT, A. e POGGIO, T., "Making Machines (and Artificial Intelligence) See". In: DÆDALUS, Vol. 117, No. 1, *Proceedings of the American Academy of the Arts and Science*, pp. 213-239, 1988.
- [34] REEKE, G. N. e EDELMAN, G. M., "Real Brains and Artificial Intelligence", In: DÆDALUS, Vol. 117, No. 1, *Proceedings of the American Academy of the Arts and Science*, pp. 143-173, 1988.
- [35] RECCE, M. L., TRELEAVEN, P. C., VELLASCO, M. e PACHECO, M., "Computação Neural: Uma Introdução", VIII Congresso da SBC, 1988.
- [36] COTTRELL, G. W. e SMALL, S. L., "Viewing Parsing as Word Sense Discrimination: A Connectionist Approach", In: *Computational Models of Natural Language Processing*, B. G. Bara e G. Guida (Eds.), North-Holland, pp. 91-119, 1984.
- [37] CARVALHO, L. A. V., "Síntese de Redes Neurais com Aplicações à Representação do Conhecimento e à Otimização", Tese D. Sc., COPPE/UFRJ, Programa de Engenharia de Sistemas e Computação, 1989.
- [38] HOPFIELD, J. J., "Neurons with Graded Response have Collective Computational Properties like those of two-state neurons", *Proceedings of National Academy of Science of USA*, pp. 3088-3092, Fev./1984.
- [39] KIRKPATRICK, S., GELLAT, C. D. e VECCHI, M. P., "Optimization by Simulated Annealing", *Science*, Vol. 220, No. 4598, pp. 671-680, Maio/1983.
- [40] HINTON, G. E., ACKLEY, D. H. e SEJNOWSKI, T. J., "A Learning Algorithm for Boltzmann Machines", *Cognitive Science*, n. 9, pp. 147-169, 1985.
- [41] METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. "Equation of State Calculations by Fast Computing Machines", *The Journal of Chemical Physics*, Vol. 21, No. 6, pp. 1087-1092, Junho/1953.
- [42] BARBOSA, V. C., "Redes Neurais e "Simulated Annealing" como Ferramentas para Otimização Combinatória", I Escola Brasileira de Otimização, 1989.
- [43] RUMELHART, D. E., HINTON, G. E. e McCLELLAND, J. L., "A General Framework for Parallel Distributed Processing", In: *Parallel Distributed Processing: Ex-*

- plorations in the Microstructure of Cognition, Vol. 1: Foundations*, J. A. Feldman, P. J. Hayes e D. E. Rumelhart (Eds.), pp. 45-194, MIT Press, 3a. ed., 1987.
- [44] RUMELHART, D. E., SMOLENSKY, P. e McCLELLAND, J. L., "Schemata and Sequential Thought Processes in PDP Models", In: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations*, J. A. Feldman, P. J. Hayes e D. E. Rumelhart (Eds.), pp. 7-57, MIT Press, 3a. ed., 1987.
- [45] SCHANK, R., *Scripts, Plans, Goals and Understanding*, LEA Eds., 1977.
- [46] SELMAN, B., *Rule-based Processing in a Connectionist System for Natural Language Understanding*, TR CSRI-168, Department of Computer Science, University of Toronto, Abril/1985.
- [47] LACHTER, J., BEVER, T. G., "The Relation between Linguistic Structure and Associative Theories of Language Learning - A Constructive Critique of some Connectionist Learning Models", *Cognition*, No. 28, pp. 195-247, 1988.
- [48] GASSER, M. e DYER, M. G., "Sequencing in a Connectionist Model of Language Processing", Tech. Rep. CSD - 880037, Computer Science Dept., University of California, Los Angeles, Maio/1988.
- [49] McCLELLAND, J. e RUMELHART, D. E., "On Learning the Past Tenses of English Verbs", In: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations*, J. A. Feldman, P. J. Hayes e D. E. Rumelhart (Eds.), pp. 216-271, MIT Press, 3a. ed., 1987.
- [50] McCLELLAND, J. L. e KAWAMOTO, A. H., "Mechanisms of Sentence Processing: Assigning Roles to Constituents of Sentences", In: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations*, J. A. Feldman, P. J. Hayes e D. E. Rumelhart (Eds.), pp. 272-331, MIT Press, 3a. ed., 1987.
- [51] COTTRELL, G. W., "A Connectionist Approach to Word Sense Disambiguation", Tech. Rep. 154, Computer Science Dept., University of Rochester, Rochester, N. Y., Maio/1985.
- [52] WALTZ, D. L. e POLLACK, J. B., "Massively Parallel Parsing: A Strongly Interactive Models of Natural Language Interpretation", *Cognitive Science*, No. 9, pp. 51-74, 1985.

- [53] FANTY, M., "Context-Free Parsing in Connectionist Networks". Tech. Rep. 174, Computer Science Dept., University of Rochester, Rochester, N. Y., Nov./1985.
- [54] SUMIDA, R. A., DYER, M. G. e FLOWERS, M., "Integrating Marker Passing and Connectionism for Handling Conceptual and Structural Ambiguities", Tech. Rep. CSD - 880044, Computer Science Dept., University of California, Los Angeles, Junho/1988.
- [55] SELMAN, B. e HIRST, G., "Parsing as an Energy Minimization Problem", In: *Genetic Algorithms and Simulated Annealing*, L. Davis (Ed.), pp. 141-154, Cambridge, 1987.
- [56] ARAGON, C. R., JAHNSON, D. S., MeGEOCH, L. A. e SCHEVON, C., "Optimization by Simulated Annealing: An Experimental Evaluation". *Workshop on Statistical Physics in Engineering and Biology*, 1984.

Apêndice A

A.1 Simulações da Rede da figura IV.4

A.1.1 Entradas sem ambigüidades

1. 4, 5, 10, 15, 20, 21 e 25
2. 4, 5, 10, 15, 17 e 25
3. 4, 5, 10, 16, 17 e 25
4. 4, 5, 10, 13 e 25
5. 4, 5, 10 e 25
6. 3, 6, 11, 16, 17 e 25
7. 3, 6, 11, 13 e 25
8. 3, 6, 12, 13 e 25
9. 3, 6, 9 e 25
10. 3, 6 e 25
11. 2, 7, 12, 13 e 25
12. 2, 7, 9 e 25
13. 2, 8, 9 e 25
14. 2, 5 e 25
15. 2 e 25

Unidades	Entradas														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
71	-1	+1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
72	+1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
73	+1	-1	-1	-1	+1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
74	-1	+1	+1	+1	-1	+1	+1	+1	+1	+1	-1	-1	-1	-1	-1
75	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	+1	+1	-1	-1	-1
76	-1	+1	+1	+1	-1	-1	-1	+1	+1	-1	-1	-1	-1	-1	-1
77	-1	-1	-1	-1	-1	+1	+1	-1	-1	-1	-1	-1	-1	-1	-1
78	-1	-1	+1	+1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
79	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	+1	+1	+1
80	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	+1	+1	-1	-1	-1
81	-1	+1	+1	+1	-1	-1	-1	+1	+1	-1	-1	-1	-1	-1	-1
82	-1	-1	-1	-1	-1	+1	+1	-1	-1	-1	-1	-1	-1	-1	-1
83	-1	-1	+1	+1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
84	+1	+1	-1	-1	+1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
85	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
86	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	+1	+1	+1	+1	+1
87	-1	-1	-1	-1	-1	-1	-1	-1	-1	+1	-1	-1	-1	-1	-1
88	-1	+1	+1	+1	-1	+1	+1	+1	+1	-1	-1	-1	-1	-1	-1
89	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
90	+1	+1	+1	+1	+1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
91	-1	-1	-1	-1	-1	+1	+1	+1	+1	+1	-1	-1	-1	-1	-1
92	+1	+1	+1	+1	+1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
93	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	+1	+1	+1	+1	+1
94	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	-1	-1	-1	-1	-1

Observações:

- As entradas 2, 3, 4, 5 e 15 não foram bem sucedidas.
- Estão representadas apenas a partir da unidade 26, pois até a unidade 25 (inclusive) todas as unidades têm seus estados fixos. As que compreendem a entrada têm seus estados iguais a +1.
- Estas seqüências de unidades representam as sentenças, que correspondem às derivações da figura IV.3, onde a unidade 25 representa o símbolo inicial.

A.1.2 Entradas com ambigüidades

1. 4, 5, 6, 10, 15, 20, 21 e 25
2. 4, 5, 9, 10, 15, 20, 21 e 25
3. 3, 5, 6, 11, 16, 17 e 25

4. 2, 3, 6, 11, 16, 17 e 25

5. 2, 3, 7, 12, 13 e 25

Unidades	Entradas				
	1	2	3	4	5
26	-1	-1	+1	+1	+1
27	+1	+1	-1	-1	-1
28	-1	+1	-1	-1	-1
29	-1	-1	-1	-1	+1
30	-1	-1	+1	+1	-1
31	+1	+1	-1	-1	-1
32	+1	+1	+1	+1	+1
33	-1	-1	-1	+1	-1
34	+1	+1	-1	-1	+1
35	-1	-1	+1	+1	-1
36	+1	+1	-1	-1	-1
37	-1	-1	-1	-1	+1
38	-1	-1	+1	+1	-1
39	+1	+1	-1	-1	-1
40	-1	-1	-1	-1	-1
41	-1	+1	+1	+1	+1
42	+1	-1	-1	-1	-1
43	-1	-1	-1	-1	-1
44	-1	-1	-1	+1	+1
45	-1	-1	-1	-1	+1
46	+1	+1	+1	+1	-1
47	-1	-1	-1	-1	-1
48	+1	+1	-1	-1	-1
49	-1	-1	-1	-1	+1
50	-1	-1	-1	+1	-1
51	+1	+1	-1	-1	-1
52	-1	-1	-1	+1	-1
53	-1	-1	+1	+1	-1
54	+1	+1	-1	-1	-1
55	-1	-1	-1	-1	-1
56	+1	+1	-1	-1	-1
57	+1	+1	-1	-1	-1
58	-1	-1	-1	-1	-1
59	-1	-1	-1	-1	+1
60	-1	-1	-1	-1	-1
61	-1	-1	+1	+1	-1
62	-1	-1	-1	-1	-1
63	-1	-1	+1	+1	+1
64	+1	+1	-1	-1	-1
65	-1	-1	-1	+1	-1
66	-1	-1	-1	-1	-1

Unidades	Entradas				
	1	2	3	4	5
67	+1	+1	-1	-1	-1
68	-1	-1	-1	-1	+1
69	-1	-1	-1	-1	-1
70	-1	-1	+1	+1	-1
71	-1	-1	-1	-1	-1
72	+1	+1	-1	-1	-1
73	+1	-1	-1	-1	-1
74	-1	+1	+1	+1	+1
75	-1	-1	-1	-1	+1
76	+1	+1	-1	-1	-1
77	-1	-1	+1	+1	-1
78	-1	-1	-1	-1	-1
79	-1	-1	-1	+1	-1
80	-1	-1	-1	-1	+1
81	+1	+1	-1	-1	-1
82	-1	-1	+1	+1	-1
83	-1	-1	-1	-1	-1
84	+1	+1	-1	-1	-1
85	-1	-1	-1	-1	-1
86	-1	-1	-1	+1	+1
87	-1	-1	-1	-1	+1
88	+1	+1	+1	+1	-1
89	-1	-1	-1	-1	-1
90	+1	+1	-1	-1	-1
91	-1	+1	+1	+1	+1
92	+1	-1	-1	-1	-1
93	-1	-1	-1	-1	-1
94	+1	+1	+1	+1	+1

A.2 Simulações do Contra-exemplo

1. 1, 2 e 7

2. 1, 2, 3 e 7

3. 1, 2, 5, 6 e 7

4. 1, 2, 4, 6 e 7

Unidades	Entradas			
	1	2	3	4
8	+1	+1	+1	+1
9	-1	+1	-1	-1
10	-1	-1	+1	+1
11	-1	-1	-1	-1
12	+1	+1	+1	+1
13	+1	+1	+1	+1
14	-1	-1	-1	-1
15	-1	+1	-1	-1
16	-1	-1	-1	+1
17	-1	-1	-1	+1
18	-1	-1	+1	-1
19	-1	-1	-1	-1
20	-1	-1	+1	-1
21	+1	+1	+1	+1
22	-1	-1	-1	-1