UMA ABORDAGEM BASEADA EM REDES BAYESIANAS PARA A SOLUÇÃO DA AMBIGÜIDADE LÉXICA

Leila Maria Ripoll Eizirik

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 A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Aprovada por:

Profa. Sueli Bandeira Teixeira Mendes, Ph.D.
(Presidente)

Prof. Valmir Carneiro Barbosa, Ph.D.

Profa. Dóris Ferraz de Aragon, D.Sc.

Prof. Jayme Luiz Szwarcfiter, Ph.D.

Prof. Ruy Luiz Milidiú, Ph.D.

RIO DE JANEIRO, RJ – BRASIL
maio de 1990

EIZIRIK, LEILA MARIA RIPOLL

Uma Abordagem Baseada em Redes de Bayes para a Solução da Ambigüidade Léxica [Rio de Janeiro] 1990

xii , 105 p. 29,7 cm (COPPE/UFRJ, D.Sc., Engenharia de Sistemas e Computação, 1990)

Tese – Universidade Federal do Rio de Janeiro, COPPE.

1. Ambigüidade Léxica, Linguagem Natural, Redes Bayesianas.

I. COPPE/UFRJ II. Título (Série)

À memória de meu pai.

À minha mãe pelo apoio constante.

Ao Nelson que soube me amar e acompanhar meu crescimento profissional.

Às minhas filhas Julia, Alice e Cecilia.

AGRADECIMENTOS

À minha orientadora Sueli B. T. Mendes pelo incentivo constante ao longo do meu doutoramento.

Ao meu orientador Valmir C. Barbosa pelo apoio, paciência, e dedicação durante todas as fases desta tese.

Aos colegas Inês de C. Dutra e Luís Adauto Pessoa pela eficiente implementação dos simuladores.

Ao Projeto de Processamento Paralelo pelo apoio e incentivo.

A todos os meus colegas do Programa de Sistemas, em particular, às minhas colegas da linha de Inteligência Artificial, por me liberarem das atividades acadêmicas.

Ao Nelson pela paciência na correção do português.

Às minhas filhas pelo carinho e compreensão.

Ao Aloysio pelo apoio nos momentos difíceis.

À Daisy pelo cuidadoso trabalho de datilografia.

RESUMO DA TESE APRESENTADA À COPPE/UFRJ COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS (D.Sc.)

UMA ABORDAGEM BASEADA EM REDES BAYESIANAS PARA A SOLUÇÃO DA AMBIGÜIDADE LÉXICA

Leila Maria Ripoll Eizirik
maio de 1990

ORIENTADORES: Valmir Carneiro Barbosa
Sueli Bandeira Teixeira Mendes
PROGRAMA: Engenharia de Sistemas e Computação

Neste trabalho apresentamos uma proposta de solução da ambigüidade léxica das linguagens naturais utilizando um modelo Bayesiano. A rede proposta está dividida em duas sub–redes, a sintática e a semântica. Desenvolvemos algoritmos para a construção automática da sub–rede sintática a partir de uma gramática livre de contexto e apresentamos a organização da sub–rede semântica baseada em uma gramática de casos.

As sub–redes sintática e semântica interagem de forma que na solução da ambigüidade o diagnóstico sintático é alterado pelo semântico e vice–versa. A análise das sentenças é feita de forma inteiramente paralela e a interpretação semântica emerge da interação entre a análise sintática e a análise semântica.

Implementamos um simulador para redes Bayesianas e fizemos testes com sentenças contendo ambigüidades de categoria gramatical, de estrutura sintática e de significados das palavras. Concluímos que as redes Bayesianas são adequadas para expressar e compatibilizar as relações semânticas com as informações sintáticas viabilizando um método de análise das sentenças completamente paralelas.

ABSTRACT OF THESIS PRESENTED TO COPPE/UFRJ AS PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF SCIENCES (D.Sc.)

A BAYESIAN–NETWORK APPROACH TO LEXICAL DISAMBIGUATION

Leila Maria Ripoll Eizirik
May, 1990

THESIS SUPERVISORS: Valmir Carneiro Barbosa
Sueli Bandeira Teixeira Mendes

DEPARTMENT: Systems Engineering and Computer Science

In this thesis we present a Bayesian–network approach to the lexical disambiguation of natural language sentences. The network is subdivided into a semantic subnetwork and a syntactic subnetwork. The syntactic subnetwork is obtained algorithmically for a given context–free grammar, and the semantic subnetwork is based on a case grammar.

The sentence analysis is carried out completely in parallel and during the process the two subnetworks interact continually. The semantic interpretation of the sentences emerges from the global interaction of the two subnetworks.

We have implemented a Bayesian–network simulator and performed tests with sentences where both syntactic and semantic ambiguities were present. We found that semantic relations and syntactic information are expressed in an efficient and precise way by using a Bayesian model, which in addition is very amenable to a parallel implementation.

ÍNDICE

Pág.

ÍNDICE DE ALGORITMOS

ÍNDICE DE FIGURAS

Pág.

Pág.

CAPÍTULO I

INTRODUÇÃO

A ambigüidade é uma das características centrais das linguagens naturais. Uma sentença é ambígua quando possibilita a construção de mais de uma interpretação semântica.

Quando afirmamos a existência de um ou mais significados de uma sentença e buscamos solucionar a ambigüidade, estamos presumindo que existe um contexto no qual a sentença está inserida e no qual a ambigüidade é eventualmente resolvida.

Entretanto, a compreensão de qualquer sentença pressupõe um interlocutor/leitor. Então, não há como assegurar uma compreensão única e completa do que está sendo dito/escrito. Acreditamos que isto acontece porque cada pessoa tem seus conceitos (significados de palavras/sentenças) suficientemente gerais para permitir a comunicação com os outros e suficientemente particulares para gerar a ambigüidade de interpretações.

Estas observações levam à conclusão sobre a existência de uma continuidade entre a lingüística e outras formas de conhecimento do mundo. A constatação de que a interpretação semântica de uma sentença não está completa em representações do tipo estrutura profunda*, proposta por CHOMSKY [1965] está presente em seu trabalho posterior [CHOMSKY, 1970].

CHOMSKY [1970] afirma ser natural (embora parcialmente correto) supor que a interpretação semântica de uma sentença é determinada pelo conteúdo semântico intrínseco dos itens léxicos e a maneira como eles estão relacionados ao nível da estrutura profunda. A razão pela qual esta afirmação é apenas parcialmente correta reside, segundo Chomsky, no fato de que fatores não–lingüísticos, tais como crenças e intenções, interferem na interpretação semântica.

Além disso, se analisarmos as afirmações de Chomsky, veremos que ele se

---

* Para uma definição do conceito de estrutura profunda veja–se CRYSTAL [1988].

refere a "uma semântica intrínseca dos itens léxicos", ou seja, os itens léxicos possuem significados claros, pré–definidos e independentes. Acreditamos que significados assim definidos de forma estanque constituem uma descrição falha dos significados que são consensuais e que viabilizam a comunicação entre as pessoas.

Esta questão da existência de significados "discretos" das palavras reflete–se nos sistemas computacionais mediante uma representação mais ou menos localizada dos significados. Ou seja, um determinado significado de uma palavra pode ser representado pelo preenchimento de determinadas características básicas (micro–características) ou mediante uma única entidade que representa o significado como um todo.

Segundo WILKS [1988], a existência de um ou mais significados distintos para cada palavra é fundamental para a discussão da ambigüidade: "... the issue at the heart of lexical ambiguity resolution – namely, are there, in any real sense, discrete word senses, or is it all just a matter of fuzzy, numerical–boundary classifications of individual examples of use ?".

Do ponto de vista computacional, essa questão está diretamente ligada à forma de representação dos conceitos. Nesse sentido, as abordagens fundamentadas na lógica são demasiado rígidas para modelar os significados das palavras.

Além disso, os sistemas computacionais em que a análise sintática e semântica estão implementadas como processos autonômos não possuem mecanismos que viabilizem a utilização de informações da análise semântica e contextual para alterar a análise sintática e vice–versa. Esse modo de funcionamento dificulta a solução da ambigüidade, que exige em geral informações sintáticas, semânticas, contextuais e até mesmo de conhecimento do mundo.

Esta tese apresenta uma proposta para o tratamento da ambigüidade léxica, onde os processos de análise sintática e semântica são implementados de forma independente do ponto de vista de estrutura interna. Porém, a comunicação entre os processos sintático e semântico é bilateral e a interpretação da sentença emerge da interação dos processos.

Foi feita uma implementação completa da rede sintática considerando sentenças de comprimento máximo 5. A rede sintática foi testada para diversas entradas, incluindo entradas ambíguas. Além disso, foram feitos alguns testes com

parte da sub–rede semântica. A rede funcionou corretamente no caso de ambigüidades de categoria gramatical e de significado.

Foram feitos, ainda, alguns testes com sentenças cujas ambigüidades de categorias gramaticais resultam em mais de uma estrutura sintática correta. Nesses casos, a solução da ambigüidade deu–se no nível semântico, a sub–rede sintática adaptou–se à solução semântica e optou pelo diagnóstico sintático correto.

Concluindo os testes realizados, apresentamos à rede uma entrada completamente agramatical do ponto de vista sintático, porém com itens semanticamente relacionados. Obtivemos então diagnósticos sintáticos que localizavam corretamente o verbo principal. Além disso, a inexistência de uma estrutura sintática correta não impediu uma interpretação parcial da semântica associando corretamente os itens relacionados.

No Capítulo II apresentamos uma classificação das ambigüidades de acordo com as suas origens: ambigüidade estrutural, caso em que a ambigüidade provém das diversas formas possíveis de agrupar os constituintes de uma sentença, ou ambigüidade léxica, caso em que provém dos significados distintos de uma mesma palavra. Ilustramos com alguns exemplos os tipos de ambigüidade existentes e os seus contextos de solução. Apresentamos também uma breve descrição dos principais sistemas computacionais que tratam do problema e estabelecemos algumas comparações com o nosso trabalho.

No Capítulo III descrevemos o modelo de redes Bayesianas, enunciando os teoremas principais que estabelecem o poder de expressão do modelo e que provêm um método seguro de especificação de redes consistentes com um modelo probabilístico.

Inicialmente apresentamos um critério de separação em DAGs (Directed Acyclic Graphs) que expressa mais acuradamente as dependências do modelo probabilístico. A seguir enunciamos os teoremas que estabelecem a vizinhança mínima que deve ser consultada para atualização do valor de um nó da rede e descrevemos as restrições que devem ser atendidas para a especificação das probabilidades condicionais associadas a um DAG. Finalmente, discutimos os problemas de atualização de valores em redes que contêm ciclos, apresentando o

método de simulação estocástica que foi por nós utilizado.

Os Capítulos IV, V e VI constituem a parte central desta tese, apresentando respectivamente a descrição geral da abordagem adotada, a descrição detalhada da parte sintática, e a descrição detalhada da parte semântica.

O Capítulo IV apresenta uma visão global da rede. Descrevemos em linhas gerais como estão organizados os módulos que compõem a rede e como interagem para obter a solução da ambigüidade. A rede está dividida em duas sub–redes: a sub–rede sintática e a sub–rede semântica. Apresentamos mediante exemplos, o funcionamento geral da rede e a forma como interagem as sub–redes sintática e semântica. Além disso, descrevemos duas formas de implementar testes de concordância gramatical para verificar a correção da entrada e também para utilizar estas informações na solução da ambigüidade.

O Capítulo V trata especificamente da análise sintática. O método de análise sintática está baseado em uma gramática livre de contexto. Esta gramática define a estrutura sintática das sentenças analisáveis pelo sistema.

Apresentamos um algoritmo que tem como entrada uma gramática livre de contexto G e fornece como saída uma rede Bayesiana que reconhece exatamente L(G). O resultado da análise sintática é sumarizado por alguns nós especiais da rede que simbolizam estruturas sintáticas corretas. Se existir ambigüidade sintática todos os possíveis diagnósticos sintáticos resultarão igualmente prováveis.

A rede construída pelo algoritmo contém alguns nós inúteis, já que não integram nenhum diagnóstico sintático. Apresentamos então um algoritmo que elimina todos os nós inúteis.

Além disso, se a gramática de entrada estiver fatorada de uma forma especial a rede obtida é mais eficiente. A eficiência a que nos referimos diz respeito ao grau de paralelismo existente na rede. Apresentamos finalmente um algoritmo para obter a forma fatorada de uma gramática livre de contexto.

No Capítulo VI descrevemos a sub–rede responsável pela análise semântica e detalhamos a forma como se dá a interligação das sub–redes sintática e semântica. Mostramos como foram construídos os diversos níveis da sub–rede semântica e como devem ser especificadas as probabilidades condicionais

associadas aos nós. Apresentamos também um algoritmo para construção da interligação das sub–redes sintática e semântica. Finalmente, descrevemos detalhadamente a análise de algumas sentenças ambíguas.

No Capítulo VII apresentamos os resultados da implementação, extraímos algumas conclusões, mencionamos algumas possíveis extensões do sistema e alguns tópicos que permanecem em aberto.

No Apêndice I apresentamos os algoritmos de simulação utilizados, bem como algumas informações sobre o número de simulações necessárias para a convergência da rede.

No Apêndice II apresentamos uma análise da complexidade dos algoritmos apresentados.

CAPÍTULO II

O PROBLEMA E SOLUÇÕES ANTERIORES

II.1 – INTRODUÇÃO

No Capítulo I, discutimos a importância da solução da ambigüidade na questão da compreensão das linguagens naturais e destacamos as dificuldades de abordar esse problema que está presente em todos os níveis de análise da sentença.

Neste capítulo, apresentaremos uma classificacão das ambigüidades tendo em vista as suas origens e os seus foros de solução. Além disso, descreveremos os principais sistemas computacionais de compreensão das linguagens naturais que abordam a questão da ambigüidade, bem como estabeleceremos algumas comparações com o nosso trabalho.

II.2 – CLASSIFICAÇÃO DAS AMBIGÜIDADES

As diversas interpretações semânticas de uma sentença podem ter origem na existência de vários significados para uma palavra ou nas formas de agrupar os constituintes que compõem a estrutura da sentença.

Quando a ambigüidade refere–se a significados distintos dos itens léxicos, é chamada ambigüidade léxica. A ambigüidade léxica pode ser classificada de acordo com o tipo de distinção entre os diversos significados de um item léxico. Uma palavra tem significados polissêmicos quando estes significados, embora distintos, estão relacionados. Por exemplo, a frase "O menino descobriu a menina" é ambígua pois o verbo "descobrir" tem, segundo HOLANDA [1986], vinte e quatro significados. Entre estes, descobrir (retirar a coberta) e descobrir (encontrar) são significados polissêmicos do verbo "descobrir", que resultam em interpretações semânticas distintas.

Um item léxico tem significados homônimos quando os significados são distintos e não estão relacionados. Por exemplo, na frase "O menino correu para o banco", o substantivo "banco" tem significados homônimos: banco (objeto para sentar) ou banco (instituição financeira). Ambos são plausíveis, resultando na ambigüidade da sentença.

A ambigüidade que provém das associações entre constituintes que compõem a sentença é chamada ambigüidade estrutural. Exemplificando, a frase "A menina viu o menino com o binóculo" é estruturalmente ambígua pois o sintagma preposicional "com o binóculo" pode estar associado ao sintagma nominal "A menima" ou não, resultando em interpretações semânticas distintas.

Poderíamos mencionar, ainda, a ambigüidade às vezes citada como ambigüidade referencial. Este tipo de ambigüidade provém principalmente das associações de substantivos e pronomes nas sentenças. Por exemplo, na frase "Os soldados atiraram nos bandidos e eles caíram", o pronome "eles" pode referir–se tanto ao substantivo "soldados" quanto ao substantivo "bandidos".

Observamos que muitas ambigüidades presentes nas sentenças passam completamente desapercebidas pelas pessoas, que utilizam inconscientemente as informações sintáticas, semânticas e pragmáticas para resolver a ambigüidade. Então, no último exemplo apresentado, a maioria das pessoas associaria com certeza o pronome "eles" ao substantivo "bandidos". Assim como, se fosse perguntado a alguém se a frase "O menino bateu na menina" é ambígua, provavelmente a resposta seria "não". No entanto, como o dicionário HOLANDA [1986] lista quarenta e dois significados para o verbo "bater" e apenas um é adequado para a frase acima, evidentemente houve a solução da ambigüidade utilizando o contexto da sentença.

As ambigüidades podem ser classificadas, ainda, de acordo com o tipo de informação necessária para solucioná–las. Assim, dependendo de cada caso, para solucionar a ambigüidade podem ser necessárias informações de diversos tipos, a saber:

1) informações sintáticas;

2) informações semânticas locais (contexto da sentença);

3) informações semânticas globais (contexto no qual a sentença está inserida); e

4) informações pragmáticas.

A seguir, apresentamos alguns exemplos de sentenças ambíguas e as informações necessárias para solucionar a ambigüidade.

Exemplo 1

Consideremos a frase "A casa é azul". A palavra "casa" pode representar o substantivo "casa" ou o verbo "casar" flexionado. A solução desta ambigüidade dá–se com informações sintáticas, já que a estrutura sintática viável para a sentença não permite a interpretação de "casa" como verbo.

Exemplo 2

Consideremos a frase "O menino bateu o recorde". A distinção entre os vários significados do verbo "bater" ocorre devido às características semânticas de outros componentes da sentença. No caso, o que determina o sentido do verbo "bater" como "ultrapassar, vencer" é a palavra "recorde".

Exemplo 3

Consideremos a frase "O menino correu para o banco". A distinção entre os significados da palavra "banco" só pode ser obtida mediante informações sobre o contexto no qual a sentença está inserida. Assim, se a sentença "O menino viu a mãe na porta do banco" precedesse a sentença em questão, a ambigüidade estaria solucionada.

Exemplo 4

Retomemos o exemplo já apresentado "Os soldados atiraram nos bandidos e eles cairam". A associação do pronome "eles" ao substantivo "bandidos" ocorre pelo conhecimento comum às pessoas sobre os fatos e suas inter–relações, a saber: quem é alvejado, em geral, cai.

Concluindo esta seção, observamos que o objetivo central do nosso trabalho é o tratamento da ambigüidade léxica utilizando informações semânticas locais, não sendo abordados os tópicos de ambigüidade estrutural ou referencial.

II.3 – PRINCIPAIS TRABALHOS SOBRE A AMBIGÜIDADE

Os sistemas computacionais para compreensão das linguagens naturais podem ser classificados em dois grandes grupos:

1) Orientados pela sintaxe: a análise sintática dirige o processo de análise da sentença e paralelamente são feitos alguns testes semânticos para excluir alternativas que não são semanticamente viáveis.

   A maioria dos sistemas de compreensão automática das linguagens naturais encontra–se neste grupo. Estes sistemas foram desenvolvidos sob a influência de modelos lingüísticos de interpretação semântica bastante influenciados pela lógica formal.

2) Orientados pela semântica: o processo de análise da sentença é dirigido prioritariamente pelas informações semânticas, usando eventualmente informações parciais da sintaxe.

   Os trabalhos que tratam da questão da ambigüidade são na sua maioria orientados pela semântica.

Um dos primeiros a abordar a questão da ambigüidade nas linguagens naturais foi WINOGRAD [1972] que desenvolveu o sistema SHRLDU. Este sistema foi projetado para um universo restrito (micromundo), eliminando desta forma a questão da ambigüidade léxica. As sentenças analisadas por SHRLDU dizem respeito ao movimento de blocos dentro de uma sala. A frase "put the block in the box on the table" é um exemplo clássico de ambigüidade estrutural resolvido por SHRLDU.

Os processos de análise sintática e semântica estão fortemente interligados em SHRLDU. O sistema foi desenvolvido usando um modelo do tipo ATN [WOODS, 1970] onde a aplicação das regras de transição está condicionada a testes semânticos. Embora SHRLDU tenha obtido bons resultados na solução da ambigüidade estrutural, a extensão do modelo para um universo mais rico e abrangente é bastante complicada, pois muitas soluções adotadas em SHRLDU só são viáveis em micromundos. Por exemplo, para identificar corretamente um objeto que corresponde a um determinado sintagma nominal, SHRLDU guarda

uma lista dos objetos já mencionados nas sentenças anteriores.

HAYES [1977] foi um dos primeiros a considerar especificamente o problema da ambigüidade léxica. Hayes discute as formas de representação dos significados das palavras e questiona a proposta de associar a uma palavra vários significados distintos sem nenhuma inter–relação.

Segundo Hayes, esta representação só é válida se os significados forem homônimos. Para os significados polissêmicos é proposta uma representação parametrizada ("parametrical senses"). Então, uma palavra tal como "head" tem um sentido paramétrico e toma como parâmetro um objeto físico com orientação vertical. Resultam assim os significados de "head of page", "head of woman" etc.

Além disso, são apresentadas várias técnicas para a solução da ambigüidade léxica. As técnicas envolvem regras que estabelecem relações entre os significados dos constituintes da sentença e destacam as informações mais importantes para solucionar a ambigüidade. As regras expressam restrições de seleção, associações dirigidas pelos verbos, e restrições de preferência.

O sistema foi implementado em LISP, usando ATNs. As técnicas que expressam as restrições e associações semânticas foram implementadas como testes e ações associadas aos arcos da rede de transição.

O trabalho de Hayes é interessante pelo estudo de casos e pela tentativa de estabelecer regras que orientem o processo de solução da ambigüidade, no entanto o sistema obtido é bastante complexo pois as regras não são muito gerais.

Mais recentemente, LYTINEN [1988] rediscutiu a questão da classificação e do tratamento da ambigüidade. No seu trabalho é proposta uma nova classificação da ambigüidade léxica, quanto aos tipos de significados (polissêmicos x homônimos). Segundo Lytinen, existem palavras vagas ("vague words") e palavras genuinamente ambíguas, sendo um equívoco tratar os diversos significados de uma palavra "vaga" de forma isolada. Por exemplo, "went" seria uma palavra "vaga" mas não ambígua pois haveria um núcleo de significado (mudança de lugar) comum a todos os seus significados.

Esta proposta de distinção entre os significados é acompanhada de uma

forma de solução chamada "concept refinement". Então, para tornar preciso o significado de uma palavra, é apresentada uma hierarquia de "frames" e um conjunto de regras de inferência que possibilitam situar precisamente uma palavra em um "frame".

MARCUS [1980] propôs um método determinístico para compreensão das linguagens naturais. A sua abordagem foi bastante inovadora e apesar de ser um método totalmente orientado pela sintaxe e determinístico, apresenta um bom desempenho na análise de um amplo conjunto de sentenças.

Embora Marcus não tenha abordado especificamente o problema da ambigüidade, o seu método foi utilizado por outros autores no tratamento da questão.

Marcus utiliza uma gramática livre de contexto para dirigir o processo de análise da sentença e realizar testes de "lookahead" para a tomada de decisões. No entanto, não são utilizadas as operações de "empilha/reduz" comuns aos métodos de análise "bottom–up". As associações das categorias sintáticas aos itens de entrada são registradas em um "buffer" que é alterado na medida em que as regras da gramática vão sendo utilizadas. Além disso, os testes de "lookahead" não são feitos com os símbolos da entrada mas com os dois primeiros constituintes do "buffer".

O trabalho de MILNE [1988] é uma extensão do método de MARCUS [1980] para incorporar o tratamento da ambigüidade léxica. Milne propõe que a ambigüidade léxica seja resolvida de forma determinística utilizando restrições válidas para o inglês sobre a ordem das palavras na sentença e regras de concordância gramatical.

As informações que são agregadas ao "parser" para tomada de decisões são:

- informações morfológicas (verbos podem ser facilmente identificados se estiverem flexionados);

- informações que definem a categoria gramatical, tais como: quando um artigo antecede uma palavra, esta não pode ser verbo;

- informações sobre gênero/número definem associações de palavras e como

conseqüência restringem as estruturas viáveis.

O trabalho de Milne aborda somente a ambigüidade de categorias sintáticas e alguns tipos de ambigüidade estrutural. As regras utilizadas para restringir a ambigüidade sintática são bastante semelhantes às que implementamos. Porém, a análise sintática que realizamos não é conclusiva e é sempre sensível às informações da análise semântica.

O trabalho de HIRST [1988] para solução da ambigüidade léxica integra um projeto geral de compreensão das linguagens naturais. O projeto é composto de: um módulo sintático Paragram, [CHARNIAK, 1983a] baseado no parser proposto por [MARCUS, 1980]; um interpretador semântico chamado Absity, [HIRST, 1987]; um módulo para resolver a ambigüidade estrutural chamado "Semantic Enquiry Desk", [HIRST, 1987]; e um sistema de representação do conhecimento chamado Frail [CHARNIAK, GAVIN and HENDLER, 1983].

O trabalho de HIRST [1988] apresenta uma abordagem que poderíamos chamar de "dirigida pela palavra". As "polaroid words" constituem representações intermediárias desse significados das palavras que vão sendo construídos durante o processo de análise da sentença. Esta abordagem, sob certos aspectos, poderia ser classificada como orientada pela sintaxe. De fato, para cada estrutura sintática parcial fornecida pelo módulo sintático Paragram, o interpretador semântico Absity associa uma interpretação semântica.

Porém, quando ocorre uma palavra ambígua Absity não consegue obter esta interpretação e então Absity trabalha com um objeto semântico falso. Este objeto corresponde a uma "polaroid photograph" do objeto semântico real e será progressivamente completado. As "polaroid words" são processos ativados em paralelo, cada uma correspondendo a um significado de uma palavra. Os processos comunicam–se entre si e com os outros componentes do sistema para solução da ambigüidade.

As alternativas de significados possíveis obtidas por um processo são colocadas em uma área que pode ser lida pelos outros processos. Com essa informação, os processos conseguem eliminar significados que não são adequados ao contexto.

As "polaroid words" são executadas paralelamente com Paragram e Absity. Assim que Paragram identifica uma palavra com uma ou mais categorias sintáticas associadas, é criado um processo para cada significado/categoria sintática. Os processos que obtiverem resultados inadequados ao contexto serão eliminados. No caso de ambigüidade (sobrevivência de mais de um processo), um sistema de "marker passing" tenta eliminar significados mediante associações obtidas pelo módulo de representação de conhecimento Frail.

Apesar de apresentar uma descrição semântica bastante elaborada e flexível, a comunicação entre as "polaroid words" e o módulo sintático é unidirecional, ou seja, as alternativas semânticas são ativadas por uma classificação sintática. Segundo HIRST [1988], embora o sistema seja proposto para funcionar em paralelo, a implementação foi feita com apenas um processo ativo de cada vez.

O trabalho de SMALL [1980] propõe uma abordagem semelhante à de Hirst, no sentido de ser uma análise dirigida pela palvra. No entanto, distingue–se do sistema discutido acima por não estar vinculado a uma análise sintática prévia da sentença; portanto, pode ser classificado como um sistema orientado pela semântica.

No sistema WEP ("Word Expert Parsing"), a palavra concentra e orienta todos os processos de análise da sentença, agindo interativamente com outras palavras e fontes de conhecimento. Cada palavra (processo) interage com as palavras que compõem o contexto local da sentença (conceitos já processados ou localmente esperados).

Cada processo representando uma palavra ("word expert") pode ser visto como uma rede de discriminação de significados, consistindo de nós de sondagem de contexto e arcos ligando o conceito aos prováveis domínios de resposta.

O principal problema desta abordagem é que a informação sintática disponível é apenas local, vinculada à palavra. Então, muitas frases não são interpretadas corretamente pela ausência de informações sintáticas que seriam trivialmente obtidas por um analisador sintático.

ADRIENS & SMALL [1988] reapresentam o sistema WEP e discutem as idéias que nortearam o projeto do ponto de vista da plausibilidade

psico–lingüística e biológica do modelo, e concluem sobre a validade da proposta para testar hipóteses lingüísticas e psicológicas do processo de compreensão.

Os processos de WEP foram implementados como co–rotinas LISP. No entanto, não foram bem sucedidas as tentativas de implementação dos processos de forma completamente distribuída. Como alternativa está sendo estudada uma redefinição de WEP com restrições que viabilizem a sua implementação usando uma versão paralela de PROLOG.

Finalmente, descreveremos alguns trabalhos que estão mais diretamente relacionados ao nosso. Nestes trabalhos, a análise semântica tem um papel fundamental; porém, as informações sintáticas, morfológicas e contextuais também contribuem interativamente para a solução da ambigüidade.

Os sistemas que apresentaremos a seguir distinguem–se dos trabalhos de HIRST [1988] e SMALL [1980], principalmente pelo modelo computacional utilizado. De fato, tanto Hirst como Small propõem um modelo paralelo e distribuído de computação, mas dentro de uma abordagem simbolista, já que os processos comunicam–se entre si mediante a troca de mensagens simbólicas. Além disso, embora as propostas de Hirst e Small sejam de um modelo paralelo, nenhum dos dois obteve de fato uma implementação em paralelo.

Os trabalhos de WALTZ & POLLACK [1985], COTTRELL [1985] e MCCLELLAND & KAWAMOTO [1986] utilizam modelos conexionistas para tratar o problema da ambigüidade. Nestes modelos a computação é feita por um grande número de unidades simples que se comunicam mediante valores numéricos. Estas unidades podem estar ativas ou inativas e o seu valor de ativação é dado por uma função dos valores de entrada e um limiar que regula a ativação. O sistema consiste em uma rede de unidades e a entrada na rede é feita mediante a fixação (grampeamento) do valor de determinadas unidades. Para uma dada entrada, são processados e atualizados os valores das outras unidades e a rede chega eventualmente a um estado estável (ponto mínimo de uma função de energia). Esta configuração estável é o resultado do processamento associado à entrada.

Os sistemas conexionistas de tratamento das linguagens naturais exploram os mecanismos de "spreading activation" e "lateral inibihition" para obter a interpretação semântica de uma sentença.

O processo de "spreading activation" é importante para a propagação das características semânticas associadas a uma palavra em uma rede de representação do contexto ou conhecimento. Obtém–se dessa forma o efeito de "semantic priming". Este efeito explica a dificuldade das pessoas de associar o significado correto de certas palavras em frases do tipo "garden path". Por exemplo, na frase "The astronomer married the star" o efeito de "semantic priming" explica a tendência inicial de associar "star" com um astro celeste e não como uma pessoa.

O processo de "lateral inibihition" é essencial na decisão de um significado para uma palavra. De fato, ativados em paralelo todos os possíveis significados para uma palavra, estabelece–se uma competição entre os significados. A decisão ocorre quando um dos significados torna–se mais ativo devido às influências das informações morfológicas, sintáticas e contextuais. Nesse caso, o que torna os outros significados inativos é o efeito da inibição lateral.

WALTZ & POLLACK [1985] propõem uma rede neuronal organizada em níveis para solucionar a ambigüidade estrutural e léxica. As unidades de entrada que representam as palavras estão ligadas somente ao nível léxico. Neste nível as palavras ambíguas possuem várias representações, cada uma associada à respectiva categoria gramatical. As unidades correspondentes às categorias gramaticais (verbo, substantivo, adjetivo, etc.) estão ligadas bilateralmente a certas unidades do nível sintático referentes aos constituintes sintáticos da sentença (NP, VP, PP).

As unidades referentes aos diversos significados de uma palavra estão ligadas bilateralmente a determinadas unidades que definem contextos. Os significados que correspondem a verbos estão associados a um grupo de unidades que representam estruturas de casos verbais (FILLMORE [1968], SAMLOWSKI [1978]).

Waltz & Pollack propõem uma representação distribuída dos conceitos, mediante microcaracterísticas associadas aos conceitos, pois a representação distribuída reforça o efeito de "semantic priming" via "spreading activation".

Embora o trabalho de Waltz & Pollack seja ambicioso em termos de proposta de interação entre os processos léxico, sintático e contextual, não fica claro como foram escolhidos os pesos e limiares que resultam num comportamento adequado da rede. Além disso, a análise sintática parece ser feita de forma bastante vinculada aos exemplos, ou seja, aparentemente não há um processamento sintático que reconheça uma classe de sentenças definidas por uma

gramática. Finalmente, não é mencionado o problema da ordem das palavras na frase. Então, como o sistema é apresentado como um modelo completamente paralelo, as frases "João ama Maria" e "Maria ama João" obteriam do sistema interpretações semânticas idênticas.

O sistema proposto por COTTRELL [1985] para solução da ambigüidade léxica está organizado em três níveis: o nível léxico; o nível de significados das palavras; e o nível de casos verbais e sintaxe.

Cottrell utiliza uma representação localizada dos significados; cada item léxico está ligado a uma ou mais unidades que representam os seus significados. No nível de significados das palavras existem três sub–redes: sub–rede dos sintagmas nominais; sub–rede dos verbos; e sub–rede das palavras funcionais (preposições). As sub–redes referentes aos casos verbais e à sintaxe estão no mesmo nível e atuam interativamente para a solução da ambigüidade.

A rede sintática é construída a partir de uma gramática denominada por Cottrell "role–constituent grammar"; semelhante a uma gramática livre de contexto onde os não–terminais referentes ao verbo e sintagmas nominais estão associados a constituintes estruturais da sentença (objeto direto, objeto indireto, sujeito, etc.). Então, por exemplo, quando uma unidade correspondente a um sintagma nominal é ativada, imediatamente são ativados os seus possíveis papéis dentro das produções da gramática (objeto direto, objeto indireto, sujeito), que competem entre si. Simultaneamente, na rede que executa a análise semântica os casos verbais associados a esta palavra são ativados e também competem entre si. Cottrell utiliza uma organização hierárquica dos casos verbais, os quais estão conectados com uma rede semântica. As unidades de ligação entre a rede sintática e a semântica estabelecem as restrições da sintaxe à semântica e vice–versa.

Comparando o trabalho de Cottrell com o de Waltz & Pollack vemos que o primeiro apresenta os processos de análise sintática e semântica mais modularizados e estruturados. Além disso, a troca de informações entre os processos é estabelecida de forma mais geral e fundamentada em certas regras. O trabalho de Cottrell apresenta, ainda, uma forma mais sistemática e geral de construção da rede sintática, embora não seja apresentado nenhum algoritmo para a construção da rede sintática a partir da gramática.

Cottrell afirma que o seu sistema funciona de forma inteiramente paralela;

no entanto, de fato, é necessário um certo intervalo de tempo entre as ativações das unidades, as quais devem estar de acordo com a ordem das palavras na frase. Existem unidades de "feed–back" que habilitam ou não as ativações. Este recurso é necessário para que a rede consiga resolver as associações das unidades de significado com a sua posição na entrada.

Não fica claro no trabalho de Cottrell como são encontradas as configurações estáveis da rede. No modelo conexionista. [FELDMAN & BALLARD, 1982] utilizado por Cottrell não são apresentadas formas de obter as configurações estáveis de uma rede neuronal. Observamos que a estabilização efetiva de uma rede neuronal é, em geral, obtida mediante a minimização de uma função, de energia. Além disso, Cottrell não apresenta os critérios usados para o estabelecimento dos pesos associados às conexões e dos limiares das unidades. Estes problemas constituem uma limitação para viabilização da proposta de Cottrell, já que com o aumento do número de unidades torna–se impossível estabelecer os pesos e sintonizar a rede de forma "ad–hoc".

O trabalho de MCCLELLAND e KAWAMOTO [1986] aborda a questão da ambigüidade focalizando especificamente a forma de representação dos conceitos. Diferentemente dos trabalhos descritos anteriormente, é proposta uma representação completamente distribuída dos significados das palavras. Além disso, o principal objetivo é o preenchimento dos casos verbais. Mediante a solução dos casos verbais adequados é obtida a solução da ambigüidade. Não é feito nenhum tipo de análise sintática e as sentenças devem ter um formato fixo. Também não é feita nenhuma análise de contexto global, sendo que o contexto usado para dirimir as ambigüidades é o da sentença.

Os principais resultados obtidos são relacionados com os efeitos de "semantic priming" e a representação distribuída dos significados. O sistema apresenta bons resultados na obtenção de significados cuja especificação é incompleta ou imprecisa. Além disso, é abordada a questão do aprendizado da rede: mediante a apresentação de diversas entradas são feitos os ajustes nos pesos das conexões da rede.

O modelo conexionista utilizado está descrito em HINTON & SEJNOWSKI [1986]. Neste modelo as unidades são atualizadas probabilisticamente e as configurações estáveis são obtidas mediante a minimização de uma função de energia global. Portanto, utilizando este modelo McClelland & Kawamoto

obtiveram resultados sistemáticos sobre o comportamento da rede.

Nosso trabalho, da mesma forma que o de Cottrell, utiliza uma forma localizada de representação dos significados e uma gramática de casos para distinguir entre os significados verbais.

No entanto, nossa abordagem é inovadora em vários aspectos. Do ponto de vista do modelo utilizado obtivemos as vantagens de um modelo efetivamente paralelo, sem controle centralizado, no qual não perdemos o poder de expressar as relações causais inerentes ao problema.

Sob o aspecto de construção da rede, apresentamos um método automático de geração da rede sintática e de sua interligação com a rede semântica. Desta maneira obtivemos uma forma precisa de controlar os impactos da análise sintática na semântica e vice–versa.

Por fim, os resultados de nossos testes são confiáveis pois a estabilização da rede foi obtida mediante critérios claros e bem fundamentados.

As principais contribuições desta tese podem ser assim resumidas:

1) O processamento das sentenças é completamente paralelo e independente da ordem;

2) A rede sintática é obtida algoritmicamente a partir de uma gramática livre de contexto. Além disso, conforme mostramos no Capítulo V, testes de concordância gramatical podem ser facilmente incorporados à rede;

3) A interligação da rede sintática com a rede semântica também é obtida mediante um algoritmo;

4) O modelo Bayesiano enquanto modelo probabilístico assegura a convergência e a estabilidade da rede. Portanto os nossos resultados estão mais bem fundamentados.

CAPÍTULO III

REDES BAYESIANAS

III.1 – INTRODUÇÃO

Os sistemas automáticos de compreensão das linguagens naturais podem ser classificados simplificadamente em dois grupos: aqueles orientados pela sintaxe e aqueles orientados pela semântica. Em qualquer dos casos, a maioria dos sistemas existentes foi especificada sob uma ótica simbolista, que pressupõe um módulo de controle centralizado e a aplicação de regras sobre um conjunto de dados para a obtenção de resultados.

Os modelos conexionistas constituem uma alternativa para abordar uma série de problemas que não são passíveis de tratamento adequado mediante os modelos simbolistas. Esses problemas não são satisfatoriamente resolvidos nos modelos simbolistas devido a certas características inerentes ao paradigma simbolista.

No tocante ao processamento das linguagens naturais, os modelos conexionistas são adequados para a implementação de certos processos da análise semântica, tal como o da utilização de informações contextuais para a solução da ambigüidade. Além disso, esses sistemas têm um desempenho semelhante ao do ser humano quando a entrada é incompleta ou com algumas incorreções, ou seja, são tolerantes a falhas, resultando em respostas aproximadas. No entanto, justamente por causa desta flexibilidade, a implementação dos processos onde a existência de regras é evidente fica bastante complicada.

Parece–nos, então, que para a especificação de sistemas de compreensão das linguagens naturais, os modelos simbolistas são excessivamente restritivos e não respondem satisfatoriamente às situações de exceção. Por outro lado, a implementação dos processos de análise sintática e dos que exigem a realização de inferências (relações de causa/efeito) nos modelos conexionistas é difícil.

As unidades do tipo OR–of–AND ("disjunctive normal form") definidas por FELDMAN & BALLARD [1982] e a estruturação das redes neuronais em níveis, proposta por FELDMAN et al. [1988], constituem uma tentativa evidente de inserir alguma estrutura lógica no modelo de redes neuronais. No entanto, de qualquer maneira, as soluções são obtidas de forma "ad–hoc" (ajuste de pesos,

definição dos limiares das unidades), sem nenhuma garantia de que a rede responda corretamente às entradas.

As redes Bayesianas constituem uma solução intermediária, na qual é possível especificar as relações de causa/efeito (regras) e além disso é prevista no próprio modelo a existência de exceções e evidências que alteram globalmente o comportamento da rede.

Acrescente–se às vantagens acima as assinaladas por PEARL [1988] em relação ao chamado "abductive reasoning": se A implica B então o fato de B ser verdadeiro torna A mais provável. As redes Bayesianas apresentam este comportamento naturalmente. Ademais, como conseqüência, apresentam o comportamento descrito pela regra de "explaining away": se A implica B e C implica B e B é verdadeiro então o fato de C ser verdadeiro torna A menos provável. Estas características são extremamente importantes para a exclusão de hipóteses no processo de solução da ambigüidade.

Finalmente, cabe ser ressaltado que o fato das redes Bayesianas estarem fundamentadas na teoria das probabilidades constitui outra vantagem. Conforme PEARL [1988] são muitos os méritos dos modelos bem fundamentados teoricamente, entre os quais:

1) A teoria pode ser utilizada para garantir que as restrições (particularizações) foram feitas estritamente quando necessárias; e

2) Quando o comportamento do sistema não corresponde às expectativas é mais fácil identificar as causas e fazer os ajustes.

As redes Bayesianas enquanto grafos são adequadas para expressar de forma natural as relações de dependência presentes no problema a ser tratado. Por outro lado, enquanto modelos probabilísticos, apresentam a vantagem de regular o impacto das relações de causa/efeito mediante os valores das probabilidades condicionais associadas ao grafo.

Neste capítulo, descreveremos formalmente o modelo Bayesiano, apresentando o critério de separabilidade definido por Pearl, a construção de uma rede Bayesiana a partir de um grafo, a definição das vizinhanças mínimas que devem ser consultadas para atualização da rede, e os métodos de propagação das

evidências na rede. Para uma descrição completa do modelo Bayesiano, com resultados e provas dos teoremas apresentados deve–se consultar PEARL [1988] Capítulos III e IV.

III.2 – DEFINIÇÕES PRELIMINARES

As redes Bayesianas são DAGs (grafos dirigidos acíclicos) cujos nós representam variáveis aleatórias e cujos arcos representam a existência de influências diretas entre as variáveis conectadas. O grau de influência entre as variáveis é estipulado mediante probabilidades condicionais associadas a um nó dados os valores de seus pais. A questão crucial é definir uma perfeita associação do grafo que representa a rede com o modelo probabilístico definido pelas variáveis associadas aos nós e suas respectivas probabilidades condicionais. Nesse sentido, é necessário definir um critério de separação, tal que as variáveis independentes no modelo probabilístico correspondam a nós separados no grafo e vice–versa.

Definição III.1

Seja $U = \{\alpha, \beta, ...\}$ um conjunto finito de variáveis aleatórias com valores discretos. Seja P uma distribuição sobre as variáveis em U. Sejam X, Y e Z subconjuntos de U. X e Y são ditos **condicionalmente independentes dado Z** se e só se $P(y|z) > 0 \Rightarrow P(x|y,z) = P(x|z)$ para quaisquer configurações x, y, z das variáveis de X, Y, Z, respectivamente.

Usamos a notação $I(X, Z, Y)_P$ para representar que X e Y são condicionalmente independentes dado Z.

Consideremos o critério de separabilidade usual para grafos não dirigidos, isto é, dado um grafo G e três subconjuntos de nós do grafo X, Y, Z dizemos que **Z separa X de Y** (e usamos a notação $<X|Z|Y>_G$) se Z intercepta todas as trajetórias de X a Y.

Seja G = (V,E) um grafo não–dirigido caracterizado por um conjunto V de vértices e um conjunto E de arestas. Se identificarmos os vértices de V com um conjunto de variáveis U sobre o qual definimos uma distribuição P, dizemos que G é uma **representação gráfica** do modelo probabilístico. Estabelecida esta relação desejamos estudar em que situações $<X|Z|Y>_G$ implica $I(X,Z,Y)_P$ e vice–versa para quaisquer $X, Y, Z \subseteq U$.

Existem muitos modelos probabilísticos que não satisfazem esta dupla implicação, ou seja, modelos para os quais o critério de separação de grafos não é capaz de representar corretamente as relações de dependência e independência.

No nosso caso, desejamos que o modelo expresse dependências induzidas, cujo significado explicamos a seguir, mediante um exemplo.

Consideremos, os significados homônimos de uma mesma palavra, digamos "banco". Os sentidos associados à banco como objeto para sentar ou como instituição financeira não estão, em princípio, relacionados. Suponhamos que a variável binária B represente a ocorrência da palavra "banco". Suponhamos que as variáveis binárias B1 e B2 representem a ocorrência dos conceitos associados aos significados da palavra "banco". Especificamos uma função de probabilidade conjunta para as variáveis B, B1 e B2 espelhando as seguintes relações de dependência:

1) As variáveis B1 e B estão relacionadas, pois a palavra "banco" está associada a cada um dos seus significados. Da mesma forma, B2 e B estão relacionados.

2) As variáveis B1 e B2 não estão relacionadas.

No entanto, a ocorrência da variável B cria uma dependência entre B1 e B2, uma vez que em um dado contexto apenas um dos significados da palavra "banco" é o adequado. Então temos:

$I(B1,\phi,B2)_P$, já que $P(B1=x|B2=y) = P(B1=x)$ para $x \in \{0,1\}$ e

$\neg I(B1,B,B2)_P$, pois temos por exemplo: $P(B1=1|B2=0, B=1) \neq P(B1=1|B=1)$

Estes resultados inviabilizam uma representação gráfica para a qual valha a equivalência dos critérios de independência, pois o critério de separabilidade em grafos não–dirigidos satisfaz sempre a seguinte propriedade:

$$<X|Z_1|Y>_G \Rightarrow <X|Z_1 \cup Z_2|Y>_G \text{ para quaisquer } Z_1, Z_2 \subseteq V$$

Faz–se então necessária a construção de uma representação gráfica menos restritiva em relação às independências no modelo probabilístico.

Definição III.2

Um grafo não–dirigido G é um **mapa–I** de uma distribuição P se e só se para quaisquer conjuntos disjuntos de variáveis X, Y, Z temos:

$$<X|Z|Y>_G \Rightarrow I(X, Z, Y)_P.$$

Pela Definição III.2 vemos que, se um grafo G é um mapa–I de P então dois vértices separados em G correspondem a variáveis independentes em P; mas não assegura que vértices ligados em G correspondam, obrigatoriamente, a variáveis dependentes em P.

Para o exemplo discutido acima, constatamos que o grafo da Figura III.1 não é um mapa–I de P porque temos $<B1|B|B2>_G$ e $\neg I(B1, B, B2)_P$.

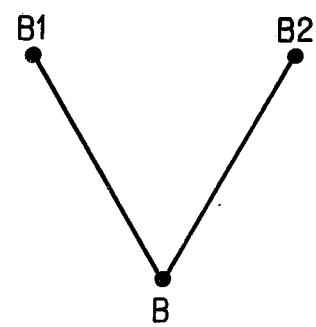

Fig. III.1 – Exemplo de um grafo não–dirigido que não expressa completamente as dependências do modelo probabilístico associado

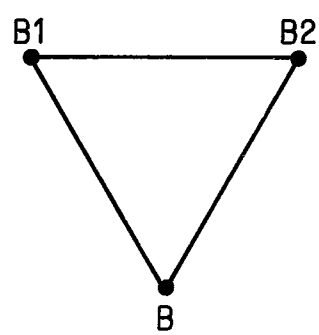

Fig. III.2 – Grafo não–dirigido completo com três vértices

Por outro lado, o mapa–I da Figura III.2 é um grafo completo e não espelha o fato de que se B não está instanciado B1 e B2 são independentes.

Para superar este problema e obter uma representação gráfica mais acurada, com capacidade de expressar dependências induzidas e não–transitivas, utilizam–se grafos dirigidos acíclicos (DAGs) com um novo critério de separabilidade.

Definição III.3

Seja $D = (V, \vec{E})$ um grafo dirigido acíclico. Sejam X, Y, Z subconjuntos disjuntos de V. Dizemos que **Z d–separa X de Y** (e usamos a notação $<X|Z|Y>_D$) se e somente se:

a) Z intercepta todas as trajetórias orientadas de X a Y ou de Y a X; e

b) Se existe uma trajetória não–orientada entre X e Y, então seja K o conjunto dos vértices da trajetória onde as setas são convergentes. Se Z intercepta K ou um descendente de um vértice de K, então o número de vértices da trajetória é maior do que 1 e a trajetória está inteiramente contida em Z.

Definição III.4

Seja $D = (V, \vec{E})$ um DAG, dizemos que D é um **mapa–I** de uma distribuição P se e só se para quaisquer subconjuntos disjuntos de variáveis X, Y, Z temos:

$$<X|Z|Y>_D \Rightarrow I(X, Z, Y)_P.$$

Um DAG D é um **mapa–I minimal** de P se a retirada de qualquer uma de suas arestas faz com que não seja mais um mapa–I de P.

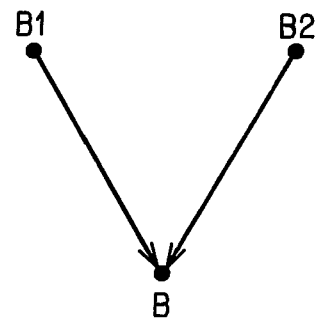

Fig. III.3– Exemplo de um mapa–I associado à uma distribuição

Retomando o exemplo apresentado temos que o DAG da Figura III.3 é um mapa–I de P pois $<B1|\phi|B2>_D \Rightarrow I(B1, \phi, B2)_P$ e $<B1|\phi|B2>_D$ é a única d–separação em D.

Em particular, B não separa B1 de B2 pois a trajetória não–orientada {B} não satisfaz a condição b) da definição III.3, refletindo o fato de que B1 e B2 são independentes exceto quando B é instanciado. A instanciação de B torna B1 e B2 dependentes, já que o conhecimento de B torna o conhecimento de B1 relevante para B2 e vice–versa.

Definição III.5

Dada uma distribuição P sobre um conjunto de variáveis U, um DAG $D = (U, \vec{E})$ é dito uma **rede Bayesiana** de P se e só se D é um mapa–I minimal de P.

III.3 – CONSTRUÇÃO DE REDES BAYESIANAS

Apresentaremos a seguir os resultados mais importantes para a obtenção de uma rede Bayesiana dada uma distribuição P.

A idéia consiste em construir, para uma determinada ordem no conjunto de variáveis U, o menor conjunto que torna cada variável independente das demais. Esta construção é feita de forma que para cada variável X, o conjunto de seus pais, Π_X, d–separa X de todos os seus não–descendentes.

Definição III.6

Dada uma distribuição P sobre um conjunto de variáveis U. Dado $\alpha \in U$, dizemos que $S \subseteq U$ é um **cobertor de Markov de α** com respeito a U se vale:

$I(\alpha, S, U - S - \alpha)$ e $\alpha \notin S$.

Usamos a notação $S = BL_I(\alpha)$.

Um conjunto é dito **limite de Markov de α** com respeito da U denotado $B_I(\alpha)$ se é um cobertor de Markov minimal, ou seja, nenhum subconjunto próprio de $B_I(\alpha)$ é um cobertor de Markov de α.

Exemplo 1

Seja $U = \{X_1, X_2, X_3, X_4\}$ um conjunto de variáveis para o qual vale a seguinte relação de independências condicionais:

$I(X_2, X_1, X_3)$, $I(X_2, X_3, X_4)$, $I(X_1, X_3, X_4)$, $I(\{X_1, X_2\}, X_3, X_4)$,

$I(X_2, \{X_1, X_3\}, X_4)$,

$I(X_1, \{X_2, X_3\}, X_4)$, $I(X_3, X_1, X_2)$, $I(X_4, X_3, X_2)$, $I(X_4, X_3, X_1)$,

$I(X_4, X_3, \{X_1, X_2\})$,

$I(X_4, \{X_1, X_3\}, X_2)$, $I(X_4, \{X_2, X_3\}, X_1)$

Os cobertores de Markov das variáveis são os seguintes:

$BL_I(X_1)$: $\{X_2, X_3, X_4\}$, $\{X_2, X_3\}$

$BL_I(X_2)$: $\{X_1, X_3, X_4\}$, $\{X_1, X_3\}$

$BL_I(X_3)$: $\{X_1, X_2, X_4\}$

$BL_I(X_4)$: $\{X_1, X_2, X_3\}$, $\{X_2, X_3\}$, $\{X_1, X_3\}$, $\{X_3\}$

e os limites de Markov são:

$B_I(X_1) = \{X_2, X_3\}$, $B_I(X_2) = \{X_1, X_3\}$, $B_I(X_3) = \{X_1, X_2, X_4\}$, $B_I(X_4) = \{X_3\}$

O limite de Markov de α é o menor conjunto que separa α das outras variáveis. De fato, por exemplo, para a variável X_4 temos $(X_4, X_3, \{X_1, X_2\}) \in I$, isto é, vale $I(X_4, X_3, U - \{X_4\} - \{X_3\})$ e portanto $B_I(X_4) = \{X_3\}$.

Definição III.7

Seja P uma distribuição sobre um conjunto de variáveis $U = \{X_1,X_2,...,X_n\}$, seja $\delta = (X_1, X_2, ..., X_i, ...)$ uma ordem qualquer em U. A **camada limite** de P em relação a δ é um conjunto ordenado de subconjuntos de U, $(B_1, B_2, ..., B_i, ...)$ tal que cada B_i é um limite de Markov de X_i com respeito a $U_i = \{X_1, X_2, ..., X_i\}$. O DAG criado designando cada B_i como conjunto de pais de X_i é chamado **DAG limite** de P em relação à δ.

Consideremos o modelo probabilístico do exemplo 1, com a ordem $\delta = (X_1, X_2, X_3, X_4)$. Então temos

$U_1 = \{X_1\}$ $U_2 = \{X_1,X_2\}$, $U_3 = \{X_1, X_2, X_3\}$ e $U_4 = \{X_1, X_2, X_3, X_4\}$

B_1 é tal que $I(X_1, B_1, U_1 - \{X_1\} - B_1)$ e B_1 é o menor conjunto que satisfaz esta condição logo $B_1 = \phi$.

$B_2 = \{X_1\}$ já que $I(X_2, X_1, U_2 - \{X_2\} - \{X_1\}) = I(X_2, X_1, \phi)$

$B_3 = \{X_1\}$ já que $I(X_3, X_1, U_3 - \{X_3\} - \{X_1\}) = I(X_3, X_1, X_2)$

$B_4 = \{X_3\}$ já que $I(X_4, X_3, U_4 - \{X_4\} - \{X_3\}) = I(X_4, X_3, \{X_1, X_2\})$

Então, conforme a Definição III.7 temos que o DAG da Figura III.4 é o DAG limite de P em relação à δ.

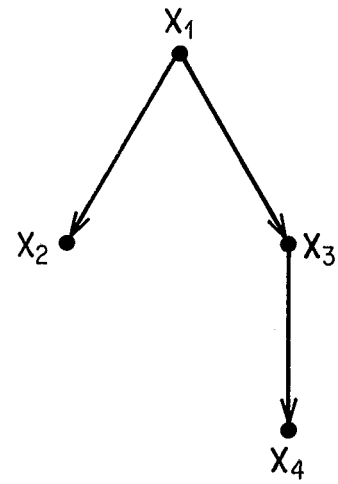

Fig. III.4 – Exemplo de DAG limite de uma distribuição P em relação à uma ordem δ

Temos, então, uma forma de construir o DAG limite de uma distribuição P, dada uma ordem δ. O teorema que enunciamos abaixo é um dos principais resultados sobre redes Bayesianas.

Teorema III.1

Seja P uma distribuição sobre um conjunto de variáveis U. Se D é um DAG limite de P em relação a uma ordem δ então D é um mapa–I minimal de P. Ou seja, D é uma rede Bayesiana de P.

A rede obtida conforme o teorema é dependente da ordem δ e diferentes ordens resultam em redes com topologias distintas. Então, de acordo com a ordem escolhida, a topologia da rede pode variar e a fidelidade da rede ao espelhar as relações causais também. No entanto, as independências do modelo probabilístico permanecem representadas, embora como recurso gráfico a rede obtida possa não ser expressiva.

Na prática, de um modo geral, é percorrido o caminho inverso, que é quase sempre, mais intuitivo. Ou seja, criamos um DAG a partir das relações de causa/efeito identificadas no problema e a seguir atribuímos valores às probabilidades condicionais associadas ao DAG. Com isso obtivemos uma distribuição P e um DAG que por construção é uma rede Bayesiana de P.

Dado um DAG $D = (U, \vec{E})$, seja Π_{X_i} o conjunto de pais de um nó $X_i \in U$.

Para definir o impacto das relações causais expressas pelo DAG, devemos definir um valor para as probabilidades condicionais associadas a cada nó X_i dados os seus pais Π_{X_i} e mediante estes valores obter uma distribuição conjunta sobre as variáveis em U. Ou seja, devemos definir funções $F_i(x_i, \Pi_{X_i})$ onde x_i são os possíveis valores de X_i e as variáveis de Π_{X_i} assumem todas as configurações possíveis.

A única restrição que devemos atender é que $\sum_{x_i} F_i(x_i, \Pi_{X_i}) = 1$.

De fato, a especificação assim feita é consistente e a distribuição conjunta obtida é dada pela fórmula:

$$P_a(x_1, ..., x_n) = \prod_i F_i(x_i, \Pi_{X_i})$$

e o cálculo da probabilidade condicional $P_a(x_i | \Pi_{X_i})$ fornece $F_i(x_i, \Pi_{X_i})$.

O corolário que enunciamos a seguir garante que o DAG especificado é uma rede Bayesiana da distribuição P_a definida acima, e além disso é único a menos de uma determinada ordem no conjunto das variáveis.

Corolário III.1

Sejam $P(x_1, ..., x_n)$ uma distribuição qualquer e δ uma ordem qualquer sobre o conjunto de variáveis. Seja X_i uma variável e seja Π_{X_i} um conjunto minimal qualquer de predecessores de X_i que satisfaça:

$$P(x_i | \Pi_{X_i}) = P(x_i | x_1, ..., x_{i-1})$$

Então, o DAG obtido definindo Π_{X_i} como o conjunto de pais de X_i é uma rede Bayesiana de P. Se P é estritamente positiva então os conjuntos Π_{X_i} são únicos para uma dada ordem δ.

A rede Bayesiana descrita nos Capítulo IV, V e VI foi projetada seguindo os passos abaixo:

1) Especificação de um DAG que expressa as relações causais definidas pelo problema;

2) Atribuição de valores para as probabilidades condicionais de cada vértice X_i dados os seus pais Π_{X_i}. Os valores atribuídos são positivos e dado que as variáveis usadas são binárias, definimos $P(X_i = 1 | \Pi_{X_i})$ e obrigatoriamente temos $P(X_i = 0 | \Pi_{X_i}) = 1 - P(X_i = 1 | \Pi_{X_i})$;

Agora considere a ordem δ no conjunto de vértices (variáveis) do DAG satisfazendo a seguinte condição: O número de ordem de um filho é sempre maior do que o número de ordem de seus pais. Por construção da rede a ordem δ é tal que $P(x_i | x_1, ..., x_{i-1}) = P(x_i | \Pi_{X_i})$. Portanto, pelo Corolário III.1, temos que o DAG especificado é uma rede Bayesiana de P. Além disso, como P é por construção estritamente positiva temos que a rede é única (dada a ordem δ).

Na implementação do modelo, principalmente no módulo de análise sintática, usamos dois tipos de unidade que definimos a seguir com o objetivo de facilitar a notação nos capítulos seguintes. Dada uma unidade X e o conjunto de seus pais $\Pi_X = \{U_1, ..., U_n\}$ dizemos que:

1) X é uma unidade do tipo **filha–e** se $P(X = 1 | U_1 = 1,...,U_n=1)$ é alta e $P(X = 1 |$ qualquer outra configuração das variáveis $U_i)$ é baixa.

2) X é uma unidade do tipo **filha–ou** se para todo $i = 1,...,n$, $P(X = 1 | U_i = 1, U_k = 0$ para $k \in J_i)$ é alta e $P(X=1 |$ qualquer outra configuração das variáveis $U_i)$ é baixa, onde $J_i = \{1 \leq j \leq n | j \neq i\}$.

III.4 – ATUALIZAÇÃO DA REDE E PROPAGAÇÃO DE EVIDÊNCIAS

Nesta seção, apresentaremos os resultados que possibilitam a definição de uma vizinhança associada a cada nó X_1 da rede. Esta vizinhança é o conjunto mínimo de nós que devem ser consultados para a atualização do valor de X. Além disso, descreveremos as formas de propagação de evidências (nós com valor grampeado) e os problemas decorrentes da existência de ciclos na rede.

O Teorema III.1, apresentado na seção anterior, tem como conseqüência o corolário que enunciamos a seguir. Este corolário define a vizinhança que d–separa um nó X do restante da rede.

Corolário III.2

Seja D uma rede Bayesiana, seja X um nó qualquer da rede. Então a união

dos seguintes conjuntos é um cobertor de Markov de X:

1) Os pais diretos de X.

2) Os sucessores diretos de X.

3) Os pais dos sucessores diretos de X.

Este corolário do Teorema III.1 é uma generalização da propriedade das cadeias de Markov que estabelece o seguinte:

Se em uma seqüência de n tentativas $X_1, X_2, ..., X_n$ o resultado de qualquer tentativa X_k $(2 \leq k \leq n)$ depende apenas da tentativa anterior X_{k-1} então o resultado de X_k depende apenas de X_{k-1} e X_{k+1}.

A construção da rede Bayesiana como um DAG limite garante que cada nó X_i é independente dos anteriores $\{X_1, ..., X_{i-1}\}$ dado Π_{X_i}.Daí resulta o Corolário III.2. A equação que estabelece a dependência de X com a sua vizinhança é dada pelo teorema que enunciamos a seguir:

Teorema III.2

Sejam D uma rede Bayesiana e X uma variável. Seja W o conjunto de todas as variáveis. Seja $U_X = \{U_1,...,U_n\}$ o conjunto de pais de X. Seja $Y_X = \{Y_1, ..., Y_m\}$ o conjunto de filhos de X. Seja F_j o conjunto de pais de Y_j. Seja $W_X = W - X$ o conjunto de todas as variáveis exceto X. A probabilidade da variável X condicionada ao estado de todas as outras variáveis é dada por:

$$P(x|w_x) = \alpha\, P(x|u_x) \prod_{j \in Y_j} P(y_j|f_j(x))$$

onde α é uma constante normalizadora independente de x e além disso, x, w_x, u_x, y_j e $f_j(x)$ correspondem a instanciações de X, W_X, U_X, Y_j e F_j, respectivamente.

Concluímos, então, pelo Corolário III.2, que $V_X = U_X \cup Y_X \bigcup_{j=1}^{m} F_j$ é um cobertor de Markov de X, e portanto vale $I(X, V_X, W - V_X - X)$. Além disso, a equação que define a dependência entre X e V_X é dada pelo Teorema III.2.

Uma vez definida a relação entre um nó da rede e sua vizinhança, apresentaremos as formas de propagação de evidências na rede. As evidências são um conjunto E de unidades que servem como entrada da rede e devem estar sempre grampeadas, ou seja, não têm o seu valor atualizado no processo de propagação.

A rede Bayesiana, enquanto modelo probabilístico, provê uma distribuição conjunta sobre as variáveis (nós da rede); portanto, se instanciarmos algumas variáveis, é possível calcular o impacto das instanciações sobre as demais variáveis de forma puramente algébrica.

Dada uma distribuição P sobre um conjunto de variáveis W, o impacto de um conjunto de evidências $E \subset W$ sobre o valor de uma variável $X_i \in W$ pode ser calculado mediante a fórmula:

$$P(x_i|x_j;\ X_j \in E) = \frac{P(x_i,x_j;\ X_j \in E)}{P(x_j;\ X_j \in E)} = \frac{\sum_{x_k;X_k \neq X_i,\ X_k \notin E} P(x_1,\ldots,x_n)}{\sum_{x_k;X_k \notin E} P(x_1,\ldots,x_n)}$$

Porém, o cálculo desta expressão é computacionalmente oneroso, pois os somatórios devem ser feitos sobre todos os valores das variáveis de $W - E - \{X_i\}$ e $W - E$. Além disso, os valores de $P(x_1, ..., x_n)$ são obtidos mediante a fórmula $P(x_1, ..., x_n) = \prod_i P(x_i|x_j;\ X_j \in U_{X_i})$, já que numa rede Bayesiana a distribuição é dada pelas probabilidades condicionais. De fato, esse problema de atualização é NP–difícil [COOPER, 1987].

Associando este processo de propagação das evidências ao DAG que representa a rede Bayesiana, obtemos um método que atualiza os valores das variáveis de forma localizada, enviando mensagens pela rede para propagação do impacto das atualizações.

PEARL [1988] apresenta um método de propagação de evidências através de troca de mensagens entre os nós da rede. Cada nó X, para atualizar o seu valor, recebe de seus pais um suporte causal $\pi(x)$, recebe de seus filhos um suporte de diagnóstico $\lambda(x)$, e calcula a nova medida da crença em $X = x$ dada por:

$$BEL(x) = \alpha\lambda(x)\ \pi(x)$$

onde α é uma constante normalizadora independente de x.

Esta abordagem de troca de mensagens é válida apenas para redes Bayesianas cujo grafo não–dirigido subjacente não contém ciclos, ou seja, é uma árvore. Em particular, se na rede cada nó possui no máximo um pai, então o cálculo de $\lambda(x)$ e $\pi(x)$ pode ser realizado eficientemente apesar da intratabilidade do caso geral.

Se o grafo não–dirigido subjacente apresentar ciclos, os esquemas de propagação local não são corretas e as mensagens podem circular indefinidamente através dos ciclos, sem que a rede atinja um estado estável.

Os métodos descritos abaixo constituem alternativas para contornar os problemas ocasionados pelos ciclos nos mecanismos de propagação de evidências. São eles: agrupamentos de variáveis; condicionamento de variáveis; e simulação estocástica.

O método de agrupamento de variáveis consiste em juntar variáveis num mesmo nó, formando variáveis compostas, de tal forma que o grafo não–dirigido subjacente torne–se uma árvore. As variáveis que devem ser agrupadas são identificadas mediante os cliques presentes no grafo.

O problema desse método é que se a rede tem muitos ciclos, o agrupamento exaustivo até a obtenção de uma árvore obscurece completamente as relações causais que haviam sido expressas de forma intuitiva. Além disso, há um aumento significativo no tamanho das matrizes que expressam as probabilidades condicionais, tornando mais custoso o cálculo das mensagens π e λ.

O condicionamento de variáveis consiste na instanciação de certas variáveis que, escolhidas adequadamente, tornam o grafo não–dirigido subjacente uma árvore. Este método pressupõe o raciocínio por hipóteses em relação aos valores das variáveis escolhidas. Ou seja, escolhida uma variável X, cujo domínio é $\{x_1, ..., x_n\}$, instanciamos X com cada um dos seus valores e propagamos as evidências na rede. Finalmente, fazemos a média dos valores obtidos e calculamos a crença em X. Se a rede contém muito ciclos, será necessário instanciar muitas variáveis e nesse caso o método de condicionamento apresenta problemas análogos aos de agrupamento.

O terceiro método, da simulação estocástica, consiste em calcular as

probabilidades computando a freqüência de ocorrência dos eventos em um conjunto de simulações. No processo de simulação estocástica, cada variável $X_i \in W$ é atualizada um número k de vezes, de acordo com as probabilidades condicionais

$$P(x_i | x_j;\ j \neq i).$$

Registramos o número de vezes que um determinado valor de X_i ocorre, dividimos pelo número total de vezes k e obtemos uma estimativa do impacto das evidências sobre X_i, $P(x_i | x_j;\ X_j \in E)$.

Para utilizar este método não é necessário alterar a estrutura da rede, pois os cálculos para atualização do valor de uma variável X dependem em cada passo da simulação apenas dos valores dos seus vizinhos V_X e de uma escolha aleatória. Não importa, portanto, que a rede tenha ciclos, já que não há propagação de evidências diretamente vinculada à troca de mensagens.

A simulação estocástica resolve o problema da complexidade aparentemente exponencial apresentada pelos métodos anteriores, pois a complexidade do processo de propagação foi deslocada para o número de vezes que se repete a simulação. Ou seja, na propagação mediante a troca de mensagens, a complexidade do processo de atualização dos valores das unidades depende diretamente da topologia da rede (número de variáveis e número de pais de cada variável), enquanto que na simulação estocástica a complexidade depende da precisão (número de simulações) do resultado.

Além disso, a simulação estocástica é um método inteiramente paralelo. O processo de simulação pode ser executado em paralelo desde que se estabeleça um critério que impeça a atualização simultânea de variáveis que sejam vizinhas (como no Corolário III.2). Considerando que a rede especificada para o nosso problema apresenta vários ciclos e que os outros métodos são computacionalmente bastante onerosos, escolhemos o método de simulação estocástica para implementar a propagação das evidências na rede. Como utilizamos apenas variáveis binárias, descreveremos mais detalhadamente os cálculos necessários para simulação nesse caso.

O método de propagação é composto de dois passos que devem ser executados repetidamente para cada variável $X \in W$. Sejam $W_X = W - \{X\}$, $U_X = \{U_1, ..., U_n\}$ o conjunto de pais de X e $Y_X = \{Y_1, ..., Y_m\}$ o conjunto de filhos de X. Seja F_j o conjunto de pais de Y_j, $j = 1, ..., m$. Seja

$$V_X = U_X \cup Y_X \cup (\bigcup_{j=1}^{m} F_j).$$

Passo 1: Computação numérica local

$$P(X=1|W_X) = P(X=1|V_X) = \alpha P(X=1|u_X) \prod_j P(y_j|f_j(1))$$

$$P(X=0|W_X) = P(X=0|V_X) = \alpha P(X=0|u_X) \prod_j P(y_j|f_j(0))$$

Passo 2: Escolha aleatória

Escolha aleatoriamente um valor para X favorecendo 1 na razão de $\frac{P(X=1|V_X)}{P(X=0|V_X)}$.

Para exemplificar, suponhamos que X seja um nó de uma rede tal que $U_X = \{U_1, U_2, U_3\}$, Y_X $\{Y_1, Y_2\}$, $F_1 = \{Z_1, X\}$, $F_2 = \{X\}$, conforme ilustrado na Figura III.5.

Suponhamos que a rede encontra–se na seguinte configuração:

$U_1 = 1$, $U_2 = 0$, $U_3 = 1$, $Y_1 = 1$, $Y_2 = 0$, $Z_1 = 0$, $X = 0$

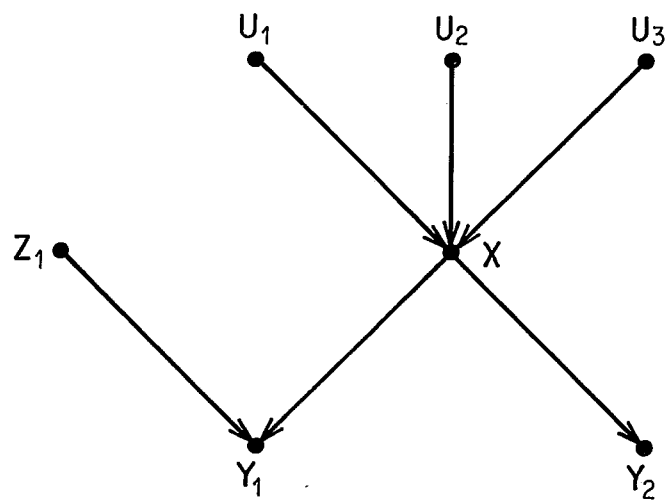

Fig. III.5 – Fragmento de uma rede Bayesiana mostrando a vizinhança que deve ser consultada para a atualização de uma unidade da rede

Então para fazer uma atualização de X temos que:

1) Calcular

$$P(X=1|V_X) = \alpha \quad P(X=1|U_1=1,U_2=0,U_3=1)$$
$$P(Y_1=1|Z_1=0,X=1)$$
$$P(Y_2=0|X=1)$$

e

$$P(X=0|V_X) = \alpha \quad P(X=0|U_1=1,U_2=0,U_3=1)$$
$$P(Y_1=1|Z_1=0,X=0)$$
$$P(Y_2=0|X=0)$$

utilizando as matrizes de probabilidades condicionais associadas a X, Y_1 e Y_2.

2) Escolher aleatoriamente um novo valor para X favorecendo 1 na razão obtida $\dfrac{P(X=1|V_X)}{P(X=0|V_X)}$.

Este ciclo (passo 1 seguido do passo 2) repete–se para todas as variáveis e pode ser executado em paralelo para variáveis que não sejam vizinhas.

Para controlar o processo de execução em paralelo sem violar a restrição de que variáveis vizinhas não podem ser atualizadas concorrentemente, usa–se uma política de controle chamada "reversão de arcos" [BARBOSA & GAFNI, 1989] que será mais bem explicada no Apêndice.

Se, após a repetição deste ciclo um número qualquer de vezes, desejarmos saber a distribuição posterior de X, calculamos o número de vezes que X obteve o valor 1 e dividimos pelo número total de vezes.

A convergência da rede para o valor correto, ou seja, $P(X=1|W_X)$ é garantida se a distribuição P é positiva, pois nesse caso temos que qualquer configuração da rede pode ser alcançada. Em outras palavras, se P é uma distribuição positiva, dado um par de configurações (i,j) qualquer, existe sempre uma probabilidade positiva de alcançar j a partir de i [FELLER, 1950].

Para concluir este capítulo faremos algumas observações sobre a capacidade de representar dependências probabilísticas mediante DAGs:

1) Os DAGs não captam completamente as dependências dos modelos probabilísticos. O critério de independência nos modelos probabilísticos é definido entre conjuntos e a independência de conjuntos não se traduz ponto a ponto, ao contrário do critério de d–separação em DAGs.

2) Os DAGs satisfazem uma forma de transitividade fraca que não necessariamente está presente nos modelos probabilísticos. Dados X, Y, Z conjuntos disjuntos de nós de um DAG, vale a seguinte propriedade:

$$<X|Z|Y>_D \text{ e } <X|Z\cup\gamma|Y>_D \Rightarrow <X|Z|\gamma>_D \text{ ou } <\gamma|Z|Y>_D, \text{ onde } \gamma \in W$$

Finalmente, seja P_1 a classe de modelos probabilísticos isomórficos a um DAG com o critério de d–separação. Seja P_2 a classe de modelos probabilísticos isomórficos a um grafo não–dirigido com o critério usual de separação de grafos. Então os modelos probabilísticos que pertencem a intersecção $P_1 \cap P_2$ constituem a classe de modelos isomórficos a um grafo cordal.

CAPÍTULO IV

A SOLUÇÃO COM REDES BAYESIANAS

IV.1 – INTRODUÇÃO

Um dos problemas cruciais na especificação de um sistema de compreensão de uma linguagem natural é a interligação dos processos de análise sintática e semântica. Nos sistemas que utilizam um universo de discurso restrito (micromundos), a questão da ambigüidade não se coloca como um problema fundamental porque, na verdade, ao restringir–se o contexto eliminam–se muitas fontes de ambigüidade, principalmente a léxica. Além disso, a ambigüidade estrutural também é mais facilmente resolvida (vide, por exemplo, WINOGRAD [1972]).

No entanto, para se abordar o problema da ambigüidade de forma geral, a questão da interação dos processos de análise sintática e semântica é decisiva. CHARNIAK [1983b] discorre sobre essa questão e conclui que ela se constitui num "quebra–cabeças" até hoje sem solução: "... semantics requires functional relations, functional relations require syntax, but semantics can be done without syntax".

Esta afirmação é, no nosso entender, parcialmente correta, já que a análise semântica para ser efetiva requer quase sempre uma análise sintática da sentença, ou pelo menos uma informação sobre a ordem das palavras na frase. De fato, por exemplo, frases simples como, "João ama Maria" não são perfeitamente compreendidas sem a seqüência das palavras. Assim como, as frases "João bateu a porta" e "João bateu na porta" são distinguidas apenas pela presença da preposição.

A afirmação de que a análise semântica pode ser feita sem a sintaxe tem a sua razão de ser, segundo Charniak, se considerarmos sentenças completamente agramaticais, como, "fire match arson hotel", e que no entanto são "compreendidas" pelas pessoas. Frases como estas não são realmente sentenças da linguagem mas podem ser consideradas como um exercício da livre associação e resultam na instanciação de um certo "frame" que define um contexto de ações e fatos possíveis ou adequados. Por outro lado, é inegável que há uma associação semântica entre as palavras da frase acima e que essa associação é obtida sem o auxílio da sintaxe.

As considerações acima mostram que os processos de análise sintática e semântica devem interargir cooperativamente, de modo que um processo não elimine alternativas que sejam viáveis para o outro. Além disso, é importante que as decisões sintáticas sofram influência das decisões semânticas e vice–versa.

Os sistemas dirigidos pela sintaxe são mais facilmente estruturados, dada a natureza algorítmica do processo de análise sintática. No entanto, como a solução da ambigüidade dá–se prioritariamente na análise semântica, os sistemas dirigidos pela sintaxe examinam, antes de obter uma interpretação semântica para a sentença, várias interpretações sintáticas, que são descartadas por não serem viáveis semanticamente. Considerando que a maioria dos sistemas dirigidos pela sintaxe são seqüenciais, a opção por privilegiar a sintaxe representa uma limitação ao desempenho dos sistemas.

Assim, os sistemas que acoplam os testes semânticos às regras sintáticas, tais como os baseados em DCGs e ATNs [ALLEN, 1987] apresentam um mau desempenho ao analisar sentenças ambíguas. Isto decorre da necessidade de realizar constantes "backtracks" para experimentar regras sintáticas alternativas.

As frases do tipo "garden path", como "The old man the boats", ou as frases sintaticamente ambíguas, como "Time flies like an arrow", constituem exemplos típicos desse problema.

Alternativamente, os sistemas dirigidos pela semântica estão fundamentados em modelos lingüísticos menos formalizados, que têm origem no trabalho de FILLMORE [1968] ou em modelos oriundos da psicologia baseados na teoria de "spreading activation" (vide QUILLIAN [1968]). A abordagem dirigida pela semântica reflete toda a complexidade inerente às inúmeras possibilidades de interpretação semântica.

Os sistemas orientados pela semântica são em geral mais adequados para tratar a ambigüidade. Porém, em contrapartida, às vezes apresentam deficiências na análise de sentenças simples, nas quais a informação sintática é decisiva.

Recentemente, surgiram várias propostas de integração dos processos de análise sintática e semântica utilizando modelos conexionistas. Nestas abordagens o processo de compreensão de uma sentença emerge da integração cooperativa

entre a sintaxe e a semântica.

Seguindo esta orientação de integração dos processos de análise sintática e semântica, as redes de Bayes apresentam, ao nosso ver, várias vantagens no tratamento da ambigüidade.

A flexibilidade oferecida pelos modelos conexionistas nos seus aspectos não–algorítmicos, tais como a propagação de decisões locais e os efeitos colaterais, tem como contrapartida as dificuldades de um modelo pouco formalizado e pouco estruturado. Tais características, se por um lado facilitam a implementação de certos efeitos importantes na análise semântica (como o de "semantic primming"), por outro dificultam enormemente a implementação dos processos que exigem a aplicação de regras, ordenação e realização de inferências (vide SELMAN [1985] e FERNANDES [1989].

As restrições necessárias para lidar com estes problemas nos modelos conexionistas complicam e tornam pouco genéricas as soluções apresentadas.

Os problemas mencionados têm uma solução automática nas redes de Bayes, pois os efeitos de competição ("winner take all") e inibição lateral ("lateral inibhition"), que devem ser cuidadosamente ajustados nos modelos conexionistas, são conseqüências naturais do funcionamento do modelo Bayesiano. Além disso, as relações de causa/efeito e a propagação de evidências inerentes ao modelo permitem uma integração harmoniosa dos processos de análise sintática e semântica.

IV.2 – DESCRIÇÃO GERAL DA ABORDAGEM

A rede especificada para compreender sentenças da linguagem natural pode ser construída para sentenças de tamanho arbitrário. Porém, na prática, os modelos paralelos para compreensão das linguagens naturais analisam sentenças de tamanho máximo ℓ, para algum $\ell > 0$.

A rede para tratamento da ambigüidade léxica está dividida em duas sub–redes: a sintática e a semântica. A Figura IV.1 mostra os principais módulos do sistema. A rede possui dois conjuntos de unidades de entrada (cada um composto de ℓ grupos) responsáveis pelas entradas sintática e semântica. A entrada é feita em paralelo e todas as unidades de entrada devem ser grampeadas (com

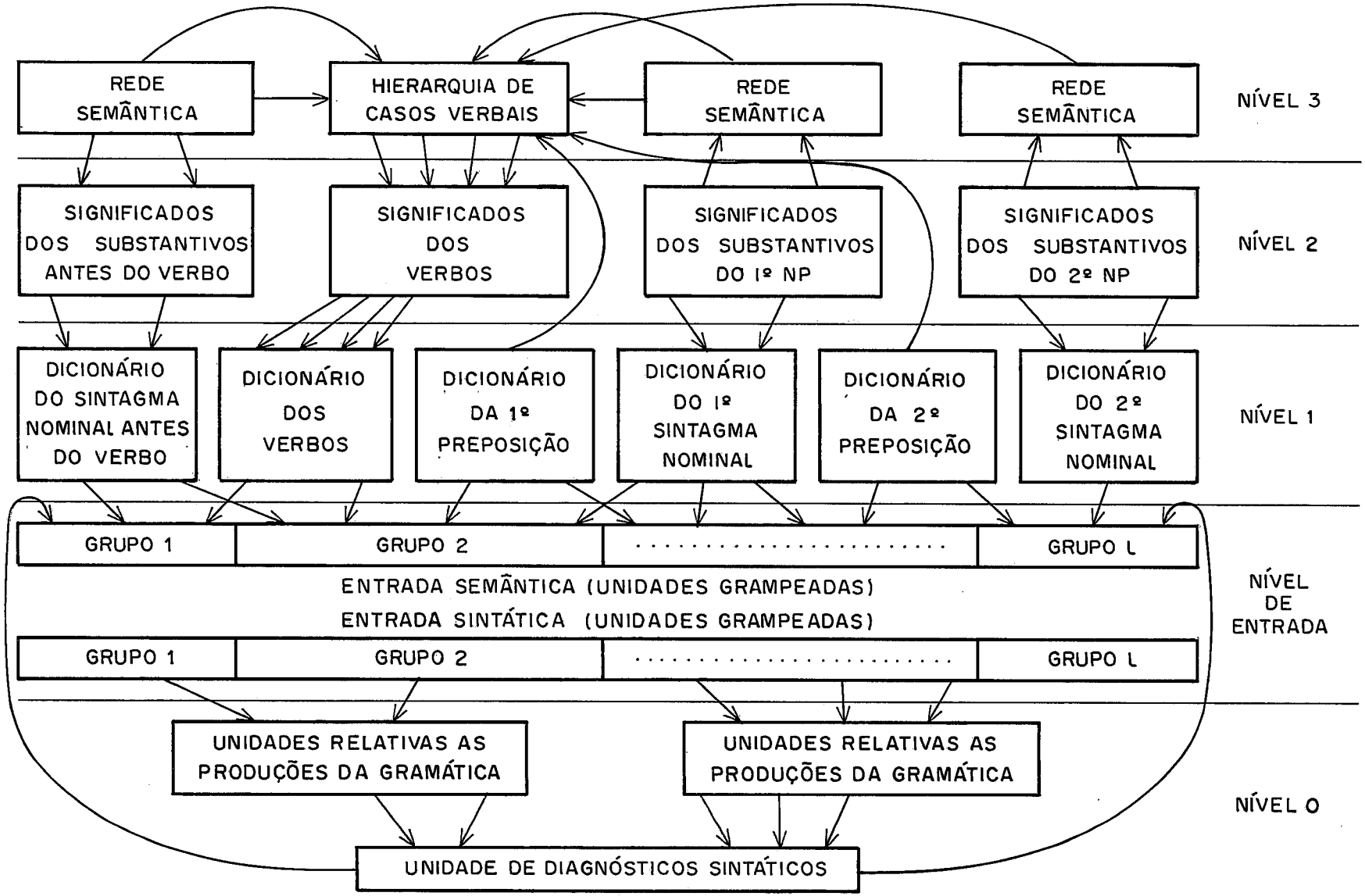

Fig. IV.1 – Descrição geral da rede

valor 1 ou 0, significando respectivamente a presença ou ausência daquela evidência).

A rede sintática (nível 0 da Figura IV.1) verifica quais sentenças pertencem à linguagem do ponto de vista estritamente sintático, definido por uma gramática livre de contexto. O resultado da análise sintática é sumarizado em um conjunto de diagnósticos sintáticos mediante os quais se estabelece a ligação entre a sintaxe e a semântica.

A rede semântica está estruturada em três níveis. No nível 1, as unidades representam a associação de uma palavra com seus possíveis papéis sintáticos: sintagma nominal antes do verbo, verbo principal, primeiro sintagma nominal depois do verbo, preposição referente ao primeiro sintagma, etc. Cada unidade representando a dupla <palavra, papel sintático> está ligada ao nível 2 que é composto de unidades representando os vários significados da palavra naquele papel sintático. Assim, por exemplo, uma unidade do nível 1 <"banco", sintagma nominal antes do verbo> é filha de duas unidades de significado da palavra "banco" como substantivo, pois "banco" pode ter pelo menos dois sentidos:

ban 1 – objeto para sentar
ban 2 – instituição bancária

Similarmente, uma unidade do nível 1 <"bater", verbo> é filha das unidades do nível 2 que representam os diversos significados do verbo "bater".

As unidades do nível 2 representam os diversos significados das palavras (substantivos e verbos). Os significados dos substantivos estão ligados à rede semântica*. utilizada para representar o conhecimento. Os significados verbais estão ligados a uma rede que representa os casos verbais hierarquicamente.

A rede semântica e a rede hierárquica de casos verbais constituem o nível 3 da rede. O nível 3 da rede representa a interpretação semântica da sentença, obtida pelo preenchimento dos casos obrigatórios do verbo principal da sentença.

---

* O termo rede semântica neste contexto significa a forma de representação do conhecimento, conforme WOODS [1975]. Segundo PEARL [1988] as redes de Bayes são adequadas para implementar esquemas de representação do conhecimento, em particular, as redes semânticas.

IV.3 – INTERLIGAÇÃO DAS SUB–REDES (FUNCIONAMENTO GERAL)

As unidades que compõem a entrada sintática representam as categorias gramaticais das palavras (evidências sintáticas). Estas unidades não têm pais e são sempre grampeadas. Seus descendentes constituem a rede sintática que é composta de unidades do tipo filha–e ou filha–ou conforme os pais sejam partes de uma produção da gramática ou referentes a produções distintas de um mesmo não–terminal.

Consideremos a sentença "João bateu no menino". A entrada sintática consiste em grampear com valor 1 as unidades "substantivo 1", "verbo 2", preposição 3", "artigo 4" e "substantivo 5" e as demais unidades de entrada sintática (como verbo 1, substantivo 2, preposição 2) com valor 0. O resultado da análise sintática é um conjunto de um ou mais diagnósticos sintáticos mais prováveis para a sentença.

A existência de mais de um diagnóstico ocorre devido à ambigüidade de categorias gramaticais ou como resultado da própria gramática livre do contexto, que pode ser ambígüa. Se a sentença contém palavras que possuem mais de uma categoria gramatical as unidades referentes às categorias possíveis devem ser grampeadas em 1 no mesmo grupo de entrada. Por exemplo, na sentença "João ama Maria" a palavra "ama" obriga o grampeamento com valor 1 das unidades "substantivo 2" e "verbo 2" na entrada sintática. As unidades que sumarizam os diagnósticos sintáticos são pais das unidades de entrada semântica.

A entrada semântica é composta de unidades (evidências semânticas) que representam cada palavra associada com a sua posição na frase. Estas unidades não têm filhos e são sempre grampeadas. Por exemplo, a entrada semântica para a sentença "João ama Maria" consiste em grampear com valor 1 as unidades "João 1", "ama 2" e "Maria 3" e com valor 0 as demais unidades de entrada semântica.

A integração da sintaxe com a semântica faz–se mediante as evidências semânticas. Estas unidades têm como pais os diagnósticos sintáticos e o nível 1 da rede que executa a análise semântica. A concordância de um determinado diagnóstico sintático com um diagnóstico semântico dá–se com a instanciação das evidências semânticas. Assim, o nível 1 do módulo semântico contém as associações possíveis das palavras com um ou mais papéis sintáticos. Por exemplo,

a unidade "ama 2" tem como pais os diagnósticos sintáticos e as unidades "ama NP0", "ama V", "ama NP1", "ama NP2" representando as duplas <ama, sintagma nominal antes do verbo>, <ama, verbo>, etc.

Consideremos a sentença "João ama Maria". O diagnóstico sintático obtido será o que sumariza a estrutura "subst 1 verbo 2 subst 3". Nesse caso, a instanciação da evidência semântica "ama 2" resulta numa configuração do nível 1 onde "ama V" ocorre com probabilidade alta e as demais unidades ("ama NP0", "ama NP1" e "ama NP2") ocorrem com probabilidade muito baixa.

No primeiro nível do módulo semântico obtêm–se, então, as associações mais prováveis das palavras aos seus papéis sintáticos, inclusive da palavra que é o verbo principal da frase. A identificação do verbo principal é fundamental para a obtenção de um diagnóstico semântico, já que esse é resultado do preenchimento dos casos obrigatórios do verbo principal.

A Figura IV.2 mostra um fragmento da rede apenas com as unidades necessárias para explicar a análise da sentença "João bateu em Pedro".

As unidades de entrada semântica "João 1", "bateu 2", "em 3" e "Pedro 4" devem ser grampeadas com valor 1. A unidade "porta 4" e as demais unidades de entrada semântica devem ser grampeados em valor 0.

As unidades de entrada sintática "subst 1", "verbo 2", "prep 3" e "subst 4" também devem ser grampeadas com valor 1 e as demais com valor 0.

Consideremos, por exemplo, a unidade do nível 1 "bater V". Esta unidade é filha de todos os significados distintos do verbo bater. Suponhamos que as unidades "b1" e "b2" do nível 2 identifiquem os seguintes sentidos do verbo bater

b1 – surrar
b2 – golpear para emitir som

A unidade "bater V" é filha–ou das unidades "b1" e "b2". A instanciação da unidade "bater V" relaciona as ocorrências de "b1" e "b2", de forma que, a ocorrência mais freqüente de um dos significados resulta na diminuição da freqüência do outro resolvendo assim a ambigüidade.

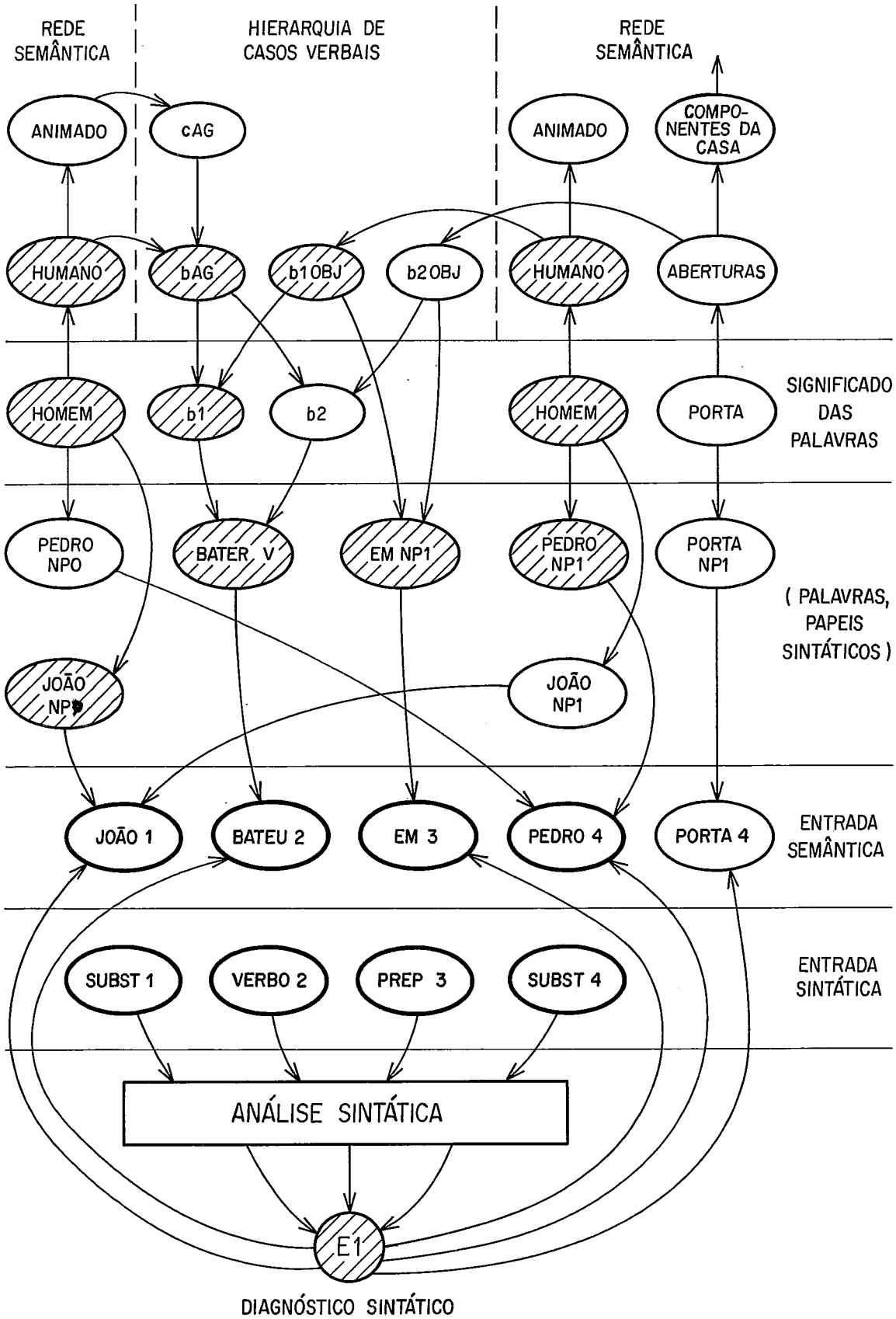

Fig. IV.2 – Organização dos diversos níveis da rede considerando a entrada "João bateu em Pedro"

As unidades "b1" e "b2" têm como pais as unidades do nível 3 que representam os seus casos. Estas unidades têm um pai comum "bAG" já que os dois significados do verbo têm agentes com características idênticas. Além disso, "b1" tem um pai "b1OBJ" e "b2" tem um pai "b2OBJ", representando os casos objeto que são obrigatórios e distintos.

As unidades representativas dos casos verbais são filhas de unidades classificatórias da rede semântica usada para representar o conhecimento (significados dos substantivos).

Assim, a unidade "bAG" é filha da unidade "humano" da rede semântica. Os casos dos verbos são preenchidos através da ligação dos significados dos substantivos com a rede semântica. No exemplo acima, a unidade "João 1" grampeada em 1 resulta num aumento da probabilidade de ocorrência das unidades "João NP0", "homem", "humano" e "bAG".

A distinção entre os dois significados "b1" e "b2" dá–se pela ocorrência mais freqüente de um dos casos objeto ("b1OBJ" ou "b2OBJ"), já que, dependendo da entrada, apenas um deles ocorrerá com maior freqüência. A interpretação semântica é dada então pela unidade de significado verbal e suas unidades de caso correspondentes. Para a sentença "João bateu em Pedro", o diagnóstico semântico é dado pelas unidades "b1", "bAG" e "b1OBJ".

As unidades desenhadas em negrito na Figura IV.2 são as evidências que estão grampeadas com valor 1. As unidades hachuradas são as que ocorrem com maior freqüência, dado este conjunto de evidências sintáticas e semânticas.

Observamos que a análise semântica utiliza a informação fornecida pelos diagnósticos sintáticos para localizar na frase o verbo principal e os sintagmas nominais correspondentes. A partir desta localização, se o verbo ou algum dos substantivos for ambíguo a solução da ambigüidade verifica–se mediante a conjunção de um significado verbal e dos casos adequados. Da mesma forma, se os diagnósticos mais prováveis oferecem várias alternativas para o verbo principal, a solução desta ambigüidade é eventualmente obtida na análise semântica.

Uma questão que não foi abordada é a da concordância gramatical. Na verdade, as regras de concordância gramatical não foram implementadas apenas

para não complicar desnecessariamente uma primeira experiência de resolver a ambigüidade de linguagens naturais utilizando o modelo Bayesiano. Hoje, no entanto, parece bastante simples acrescentar à rede estes testes de concordância gramatical.

Verificar a correção da concordância gramatical é importante para resolver alguns tipos de ambigüidade. Por exemplo, consideremos as sentenças:

1) Maria deu ao querido presentes
2) Maria gosta do lindo presente.

Na língua portuguesa (e por vezes em outras línguas) tanto o adjetivo como o verbo no particípio passado podem substantivar–se na presença do artigo. Então, ao escrever uma gramática livre do contexto para um subconjunto do português definimos a estrutura de um sintagma nominal incluindo as seguintes produções:

C → art A
C → art A N
C → N
A → adjet
N → subst

onde art, adjet e subst representam as categorias gramaticais de artigo, adjetivo, e substantivo respectivamente.

Com esta gramática as sentenças acima que têm a mesma entrada sintática (subst1, verbo2, prep3, art4, adjet5, subst6), resultam cada uma em dois diagnósticos sintáticos mais prováveis, a saber:

E1) C verbo prep C C e
E2) C verbo prep C

O diagnóstico correto seria decidido pela análise semântica de acordo com os casos (transitividade) dos verbos "dar" e "gostar". No entanto, se fosse testada a concordância do adjetivo com o substantivo a primeira frase obrigatoriamente resultaria apenas no diagnóstico sintático E1, já que o fato de estar o adjetivo no singular e o substantivo no plural impede que eles integrem o mesmo sintagma nominal.

Essa concordância pode ser testada de duas formas. Pode–se usar um modelo Bayesiano onde as unidades (variáveis) não assumem apenas valores binários ou então pode–se aumentar o número de unidades considerando gênero e número associados às unidades de entrada sintática.

Na primeira solução as unidades de entrada sintática art, subst e adjet ao invés de serem grampeadas em 0 ou 1, seriam grampeados em 0, 1, 2, 3 ou 4, conforme a convenção abaixo:

0 – não ocorre na frase
1 – ocorre como feminimo, singular
2 – ocorre como feminino, plural
3 – ocorre como masculino, singular
4 – ocorre como masculino, plural

Teríamos então a concordância estabelecida mediante as probabilidades condicionais associadas às unidades. A Figura IV.3 representa um fragmento da rede sintática relativo às produções C → art A N, C → art A e C → N.

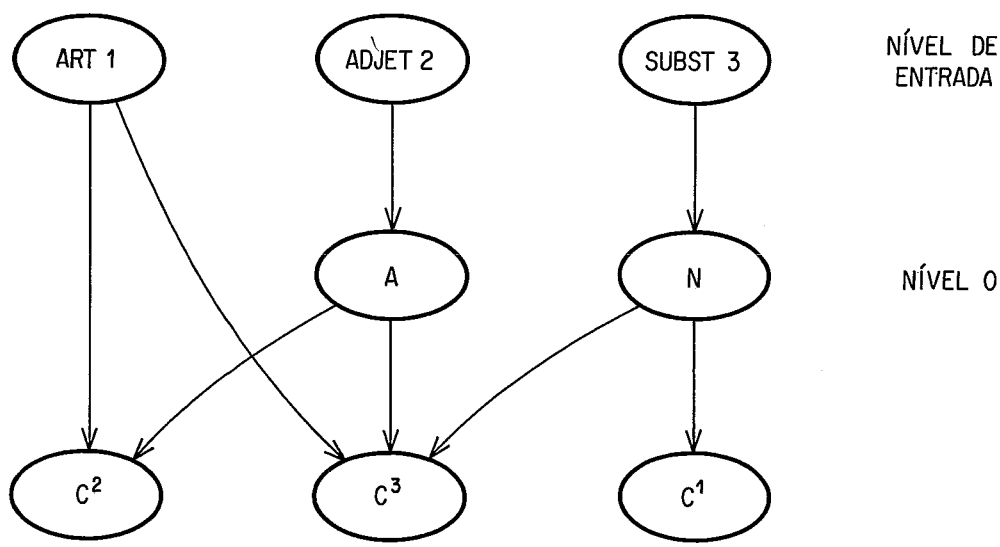

Fig. IV.3 – Fragmento de uma rede sintática modificada para incluir testes de concordância gramatical utilizando variáveis com cinco valores (0,1,2,3,4)

Fazemos:

$P(A = x | \text{adjet } 2 = x) = 1$, para $x = 0,1,2,3,4$
$P(A = x | \text{adjet } 2 = y) = 0$, para $x,y = 0,1,2,3,4$ e $x \neq y$

Analogamente:

$P(N = x | \text{subst3} = x) = 1$, para $x = 0,1,2,3,4$
$P(N = x | \text{subst3} = y) = 0$, para $x,y = 0,1,2,3,4$ e $x \neq y$

Na especificação das probabilidades condicionais das unidades C^1, C^2 e C^3 a concordância de gênero e número é então testada:

$P(C^3 = 1 | \text{art1} = x, A = x, N = x) = 1$, para $x = 1,2,3,4$

$P(C^3 = 1 |$ qualquer outra configuração de art1, A e N$) = 0$

$P(C^2 = 1 | \text{art1} = x, A = x) = 1$, para $x = 1,2,3,4$

$P(C^2 = 1 |$ qualquer outra configuração de art1 e A$) = 0$

$P(C^1 = 1 | N = x) = 1$, para $x = 1,2,3,4$

$P(C^1 = 1 | N = 0) = 0$

Com as probabilidades condicionais assim definidas, a primeira sentença resultaria em que apenas as unidades $C^1 = 1$ e $C^2 = 1$ ocorreriam com probabilidade alta, já que o valor mais freqüente da unidade A seria 3 e da unidade N seria 4, ocasionando $C^3 = 0$ com probabilidade alta. Portanto, o único diagnóstico sintático obtido com probabilidade alta seria E_1.

A solução mantendo a rede com unidades binárias e aumentando o número de unidades da rede é análoga. Criamos, para cada grupo de unidades de entrada referentes à posição i da frase, unidades binárias de gênero G_i e de número NU_i convencionando que $G_i = 1$ é masculino $G_i = 0$ é feminimo, $NU_i = 1$ é singular e $NU_i = 0$ é plural. A parte da rede sintática relativa às produções $C \rightarrow$ art A N, $C \rightarrow$ art A e $C \rightarrow$ N ficaria alterada conforme a Figura IV.4.

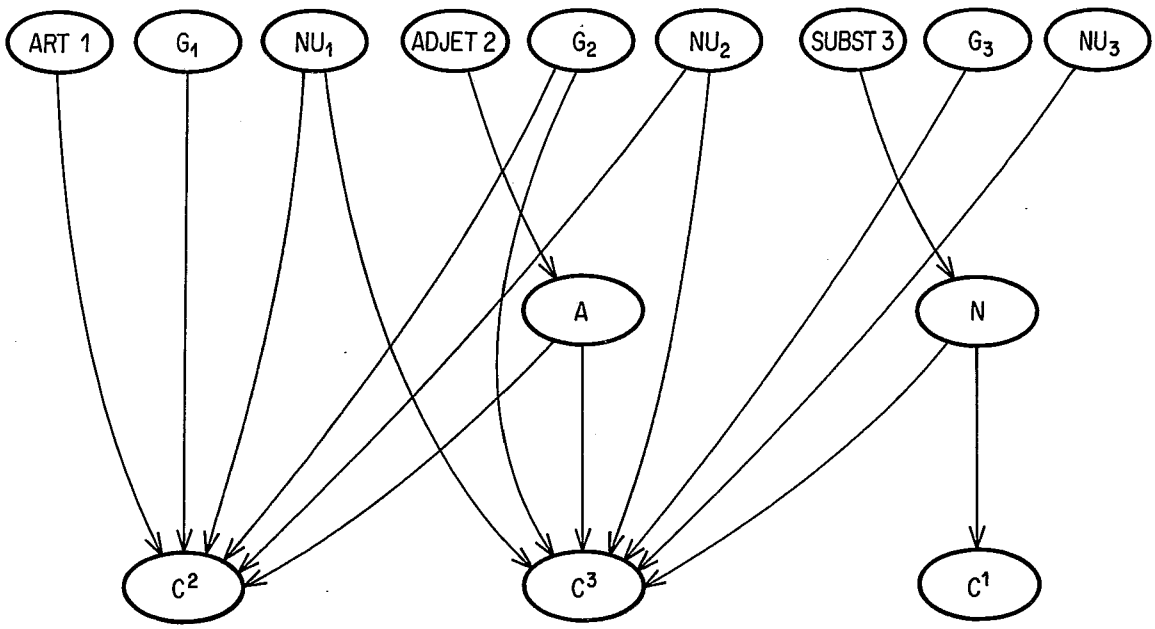

Fig. IV.4 – Fragmento de uma rede sintática modificada para incluir testes de concordância gramatical utilizando variáveis binárias

Estabelecendo adequadamente as probabilidades condicionais de C^1, C^2 e C^3 testamos a concordância. Por exemplo:

$P(C^2 = 1 \mid \text{art1} = 1, \text{adjet2} = 1, G_1 = x, G_2 = x, NU_1 = y, NU_2 = y) = 1$

para x, y = 0,1

$P(C^2 = 1 \mid \text{qualquer outra configuração dos pais}) = 0.$

CAPÍTULO V

A SUB–REDE PARA ANÁLISE SINTÁTICA

V.1 – INTRODUÇÃO

A análise sintática é executada por uma sub–rede que tem como evidências as categorias gramaticais das palavras e fornece como diagnóstico um conjunto de unidades que indicam os padrões sintáticos corretos associados à entrada dada. A rede sintática é uma rede de unidades do tipo filha–e/filha–ou constituída de unidades de entrada (referentes às categorias gramaticais), unidades intermediárias que constituem partes das árvores sintáticas obtidas a partir da entrada e unidades de diagnósticos que correspondem às produções do símbolo inicial da gramática.

O resultado da análise sintática é o conjunto de um ou mais diagnósticos mais prováveis associados à entrada. A ocorrência de mais de um diagnóstico indica ambigüidade que pode ser oriunda da própria gramática, quando ambígua, ou da ambigüidade de categorias gramaticais presentes na entrada.

A linguagem reconhecida pela rede é definida pela gramática $G=(N,\Sigma,P,S)$ onde $N = \{S, AV, C, A, N\}$ é o conjunto de símbolos não–terminais, S é o símbolo inicial, $\Sigma = \{$verbo, adjet, subst, verbopp, verboinf, art, prep, pronomp$\}$ é o conjunto de símbolos terminais (no caso categorias gramaticais), e P é o conjunto de produções mostrado na Figura V.1:

1) S → AV verbo C
2) S → AV verbo C prep C
3) S → AV verbo prep C
4) S → AV verbo prep C C
5) S → AV verbo prep C prep C
6) AV → C
7) AV → pronomp
8) AV → ϵ
9) C → art A A
10) C → art A N
11) C → art N A
12) C → art A
13) C → art N

14) $C \rightarrow N\ A$
15) $C \rightarrow A\ N$
16) $C \rightarrow N$
17) $A \rightarrow$ adjet
18) $A \rightarrow$ verbopp
19) $N \rightarrow$ subst
20) $N \rightarrow$ verboinf

Fig. V.1 – Gramática livre de contexto utilizada na geração da rede sintática

Os terminais correspondem às seguintes categorias gramaticais:

verbo – verbo flexionado
verboinf – verbo no infinitivo
verbopp – verbo no particípio passado
pronomp – pronome pessoal
prep – preposição
art – artigo
subst – substantivo
adjet – adjetivo

As formas nominais do verbo foram distinguidas por que podem substantivar–se ou adjetivar–se. Por essa razão, o não terminal N produz subst e verboinf enquanto que o não terminal A produz verbopp e adjet. A produção $C \rightarrow$ art A indica que tanto o verbo no particípio assado como o adjetivo podem substantivar–se na presença do artigo.

A gramática mostrada é deliberadamente ambígua refletindo ambigüidades que eventualmente só serão resolvidas na análise semântica. Por exemplo, dada a entrada:

s = subst verbo prep art verbopp subst

existem duas derivações mais à esquerda de s:

$S \xrightarrow{3}$ AV verbo prep C $\xrightarrow{6}$ C verbo prep C $\xrightarrow{16}$ N verbo prep C $\xrightarrow{19}$ subst verbo prep C $\xrightarrow{10}$ subst verbo prep art A N $\xrightarrow{18}$ subst verbo prep art verbopp N $\xrightarrow{19}$

$\xrightarrow{19}$ subst verbo prep art verbopp subst

e

S $\xrightarrow{4}$ verbo prep C C $\xrightarrow{6}$ C verbo prep C C $\xrightarrow{16}$ N verbo prep C C $\xrightarrow{19}$
$\xrightarrow{19}$ subst verbo prep C C $\xrightarrow{12}$ subst verbo prep art A C $\xrightarrow{18}$
$\xrightarrow{18}$ subst verbo prep art verbopp C $\xrightarrow{16}$ subst verbo prep art verbopp N $\xrightarrow{19}$
$\xrightarrow{19}$ subst verbo prep art verbopp subst

As duas derivações acima indicam que, de acordo com a transitividade do verbo principal da sentença, será utilizada uma ou outra interpretação na análise semântica. Então para a sentença

"Maria gosta do amado menino",

a primeira derivação é a correta, enquanto que para a sentença

"Maria deu ao amado presentes"

a segunda é a correta.

A obtenção de mais de um padrão sintático para uma dada entrada também pode ter origem em sentenças nas quais há palavras que têm mais de uma categoria gramatical. Por exemplo, as palavras "ama" e "casa" podem ser classificadas como verbo ou substantivo dependendo do contexto. Neste caso, são fornecidas como entrada todas as categorias gramaticais associadas à palavra e o resultado são os padrões sintáticos correspondentes a cada combinação.

Assim, para a sentença

"A bela ama casa no campo"

os padrões sintáticos possíveis são

"art adjet subst verbo prep art subst"

tomando "ama" como substantivo e "casa" como verbo.

e

"art adjet verbo subst prep art subst"

tomando "ama" como verbo, "casa" como substantivo e considerando que o adjetivo "bela" está substantivado.

V.2 – A REDE SINTÁTICA

A rede sintática está organizada em níveis. O nível de entrada da rede é composto de grupos indicativos da posição do terminal na sentença a ser analisada. Cada grupo do nível de entrada é constituído de unidades que correspondem aos terminais que podem ocorrer naquela posição de acordo com as regras da gramática. Assim, de acordo com a gramática da Figura V.1 temos, por exemplo, uma unidade verbo1 no primeiro grupo, mas não uma unidade prep1.

As unidades de entrada não têm pais e devem ser grampeadas em 0 ou 1 de acordo com a classe gramatical da palavra. Se a palavra possuir mais de uma classe gramatical todas deverão ser grampeadas no mesmo grupo em que ela aparece. A entrada é feita em paralelo e as possíveis interpretações sintáticas das entradas são obtidas concorrentemente.

O resultado da análise sintática efetuada pela rede está sumarizado em um grupo especial de unidades que correspondem a padrões sintáticos corretos de acordo com a gramática dada. Estas unidades são ligadas à rede que executa a análise semântica e sintetizam as informações de posição e função sintática necessárias para a análise semântica.

Para a descrição dos níveis intermediários da rede, consideremos a variante G' mostrada na Figura V.2 da gramática G.

1) S → C verbo C
2) S → C verbo prep C
3) S → C verbo prep C C
4) C → art N A
5) C → X N
6) C → X
7) C → N

8) X → art A
9) A → adjet
10) N → subst

Fig. V.2 – Variante da gramática da Figura V.1

A Figura V.3 mostra um fragmento da rede sintática que reconhece L(G'). Cada grupo do nível 1 contém as unidades correspondentes aos terminais que podem ocorrer naquela posição em sentenças sintaticamente corretas. Os grupos não estão completos.

O nível 2 contém unidades que correspondem a não–terminais que geram cadeias (sentenças ou prefixos de sentenças) de comprimento dois. As unidades levam o nome do não–terminal, o número do grupo de entrada no qual começa a cadeia, o comprimento da cadeia e o número da produção utilizada para construir a unidade. Assim, a unidade $X_8^{1,2}$ foi construída usando a produção 8 (X → art A) e é filha–e das unidades art1 e $A_9^{2,1}$ (começando no primeiro grupo de entrada).

A unidade $A_9^{2,1}$ é a unidade adjet 2 renomeada, pois a produção 9 (A → adjet) tem comprimento 1 e portanto não gera uma nova unidade (este processo de renomeação será explicado mais tarde).

No nível 3, as unidades $C_4^{1,3}$ e $C_5^{1,3}$ foram construídas de forma semelhante a unidade $X_8^{1,2}$. A unidade C^3 é filha–ou das unidades $C_4^{1,3}$ e $C_5^{1,3}$ e foi construída porque existe em G' mais de uma produção (4 e 5) que gera cadeias de comprimento 3 a partir de C.

A unidade $S_1^{1,3}$ é uma unidade de diagnóstico e sumariza as seguintes informações:

O verbo está na posição 2, a frase não tem preposição e possui dois sintagmas nominais.

Consideremos a sentença "João ama Maria". A entrada sintática consiste em grampear com valor 1 as unidades subst 1, subst 2, verbo 2, e subst 3, e com valor 0 as demais unidades do nível 1.

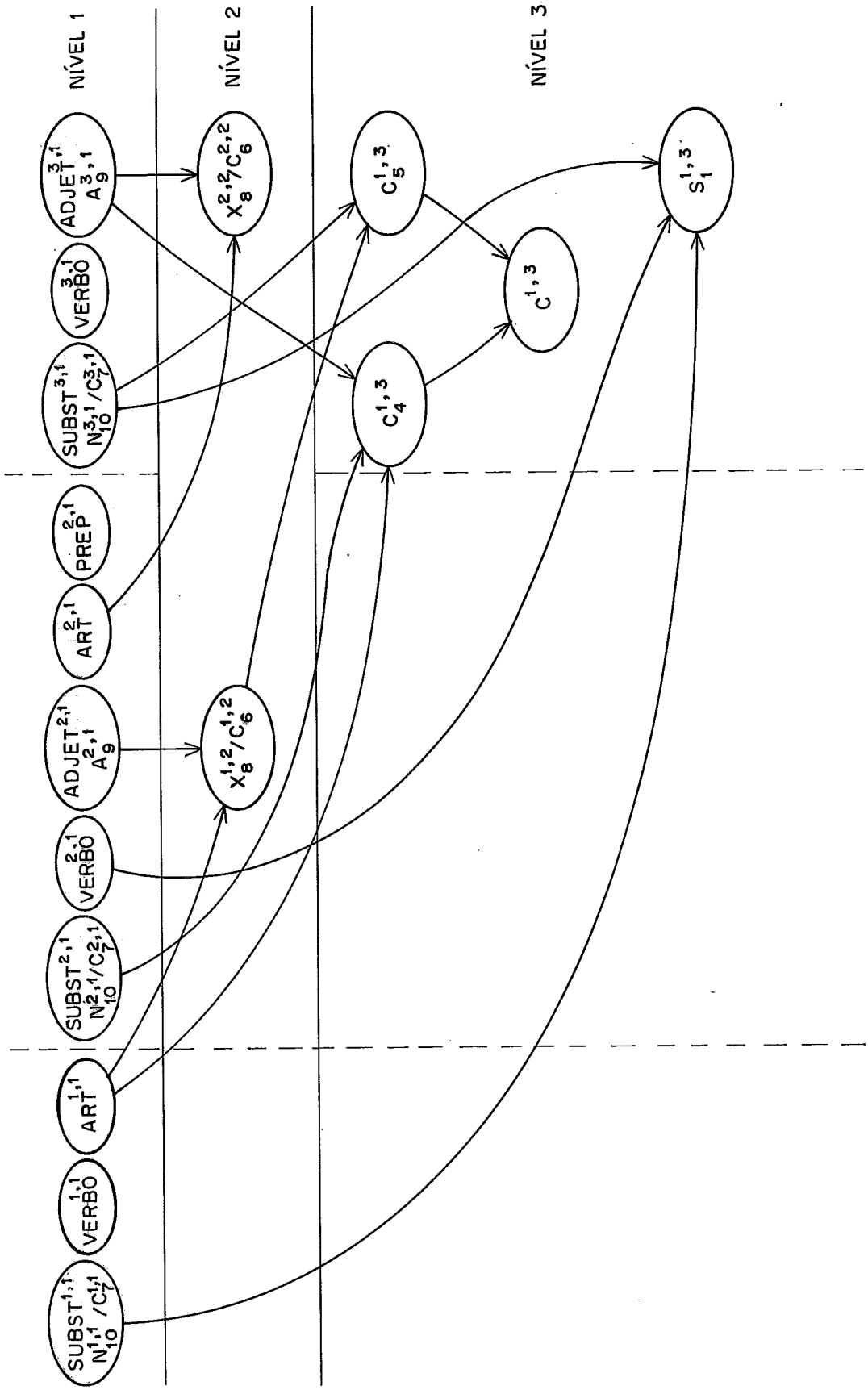

Fig. V.3 – Fragmento da rede sintática com três grupos de entrada incompletos

No segundo grupo de entrada foram grampeadas as unidades subst 2 e verbo 2 porque a palavra "ama" pode ser um substantivo ou o verbo amar flexionado. A unidade subst 2 grampeada não causará a ativação de nenhuma unidade de diagnóstico já que a única possibilidade de obter uma sentença sintaticamente correta é que "ama" seja o verbo da frase e portanto $S^{1,3}$ é o único diagnóstico sintático obtido.

Consideremos agora as frases:

"Maria deu ao vencedor presentes"
"Maria gosta do belo presente"

A entrada sintática para as duas frases seria subst 1, verbo 2, prep 3, art 4, subst 5, adjet 5, e subst 6, já que tanto a palavra "vencedor" como a palavra "belo" podem ser substantivo ou adjetivo. A Figura V.4 representa um fragmento da rede sintática apenas com as unidades de entrada necessárias para o exemplo.

Observamos que são obtidos nesse caso dois diagnósticos sintáticos $S_2^{1,6}$ e $S_3^{1,6}$ dependendo da interpretação com um ou dois sintagmas nominais depois da preposição. Esta ambigüidade só será decidida pela análise semântica de acordo com a transitividade do verbo e as características semânticas pedidas pelos casos verbais.

V.3– GERAÇÃO DA REDE SINTÁTICA A PARTIR DE UMA GRAMÁTICA LIVRE DE CONTEXTO

A gramática utilizada para construção da rede deve especificar a estrutura sintática do subconjunto da linguagem natural a ser analisado. Portanto, a gramática livre de contexto define do ponto de vista da sintaxe as sentenças que pertencem à linguagem.

Considerando que a análise sintática não é conclusiva mas apenas seleciona os padrões sintáticos corretos associados a uma entrada, é importante que, na especificação dos padrões sintáticos, fiquem claras quais as informações significativas para a análise semântica.

Para que a rede sintática espelhe esta informação é necessário que os

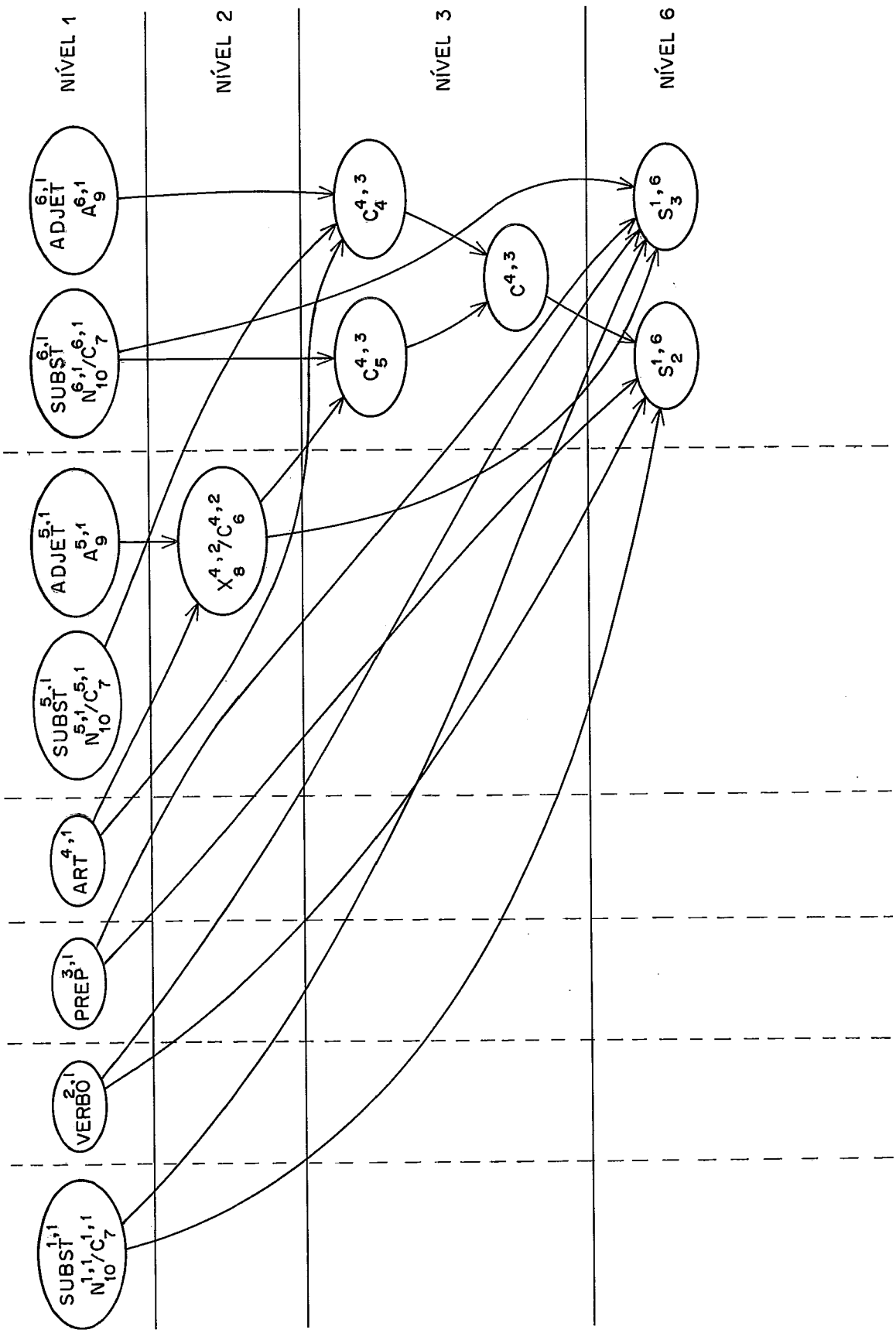

Fig. V.4 – Fragmento da rede sintática com seis grupos de entrada constituídos apenas das unidades referentes à entrada subst 1, verbo 2, prep 3, art 4, subst 5, adjet 5, subst 6 e adjet 6

marcadores sintáticos que definem um certo padrão ocorram nas produções que têm como lado esquerdo o símbolo inicial. Estas produções são tratadas de forma diferenciada na construção da rede.

Na gramática da Figura V.1 as primeiras cinco produções sintetizam as informações sobre a posição do verbo principal, o número de complementos e a ocorrência de uma ou mais preposições. Estas informações constituem o(s) diagnóstico(s) obtido(s) pela sintaxe e são usadas pela análise semântica para completar a interpretação da sentença.

A geração da rede sintática é feita em níveis. O primeiro nível da rede é a entrada da rede sintática, constituído de unidades que correspondem aos símbolos terminais da gramática. Os níveis seguintes são construídos a partir dos níveis anteriores usando as produções da gramática.

Dada uma gramática livre de contexto $G = (N, \Sigma, P, S)$, a rede é construída para reconhecer as sentenças de $L(G)$ de tamanho máximo ℓ, para algum $\ell > 0$.

A geração da rede é feita automaticamente de acordo com o Algoritmo V.1, que descrevemos a seguir e apresentamos mais tarde sob a forma de pseudo–código.

Inicialmente são eliminadas as produções ϵ da gramática. A seguir é construído o nível 1 da rede, $R_1 = (R_1^1, ..., R_1^\ell)$ constituído dos ℓ grupos de unidades $R_1^1, ..., R_1^\ell$, cada grupo contendo tantas unidades quantos forem os terminais da gramática.

O nível $R_m = (R_m^1, R_m^2, ..., R_m^{\ell-m+1})$, composto de ℓ–m+1 grupos, é construído a partir dos níveis anteriores $R_1, ..., R_{m-1}$. Cada unidade $X_p^{k,m} \in R_m^k$, $1 \leq k \leq \ell-m+1$ é referente a uma cadeia de m terminais gerados por uma seqüência de produções iniciada pela produção p: $X \rightarrow X_1 ... X_n$ e deslocada para iniciar no grupo k. Além disso, cada não–terminal X_i, $i = 1, ..., n$ que compõe o lado direito da produção p, por construção, ocorre em R_{j_i}, para algum $j_i < m$.

Consideremos, por exemplo, a gramática

1) $S \rightarrow C$ verbo C

2) C → art N
3) N → subst

A Figura V.5 mostra a rede sintática obtida. Conforme a descrição anterior, a produção 2: C → art N gera a unidade $C_2^{1,2}$ referente à cadeia de comprimento 2, "art subst", iniciando no grupo 1 e gera também as unidades $C_2^{2,2}$, $C_2^{3,2}$ e $C_2^{4,2}$ referentes a mesma cadeia deslocada para iniciar nos grupos 2, 3 e 4 respectivamente. As unidades $C_2^{1,2}$ e $C_2^{4,2}$ são pais da unidade de diagnóstico $S_1^{1,5}$.

Apresentamos a seguir a construção de uma unidade $X_p^{k,m} \in R_m^k$, a partir dos níveis anteriores já construídos $R_1, \ldots, R_{m-1}$.

Para construir a unidade $X_p^{k,m}$ no nível R_m é necessário identificar todas as formas de partir uma cadeia de terminais de comprimento m de maneira que as subcadeias sejam geradas pelos não terminais $X_1,\ldots,X_n$. Definimos então, o conjunto de índices J_n^m que é composto de n–uplas de inteiros menores do que m cuja soma é m.

Para cada n–upla de J_n^m verifica–se a existência de unidades correspondentes aos terminais $X_1,\ldots,X_n$ nos níveis compatíveis com esta partição. Em caso afirmativo, cria–se uma unidade $X_p^{k,m}$ filha–e das unidades referentes a $X_1,\ldots,X_n$ nos respectivos níveis. Como pode haver vários índices de J_n^m que satisfazem estas condições em relação à produção p, cada conjunto de unidades obtido a partir de um índice utiliza a mesma unidade $X_p^{k,m}$ como filha–e.

Se p for uma produção simples como $X \to X_1$ então não é criada uma nova unidade referente a X, apenas é acrescentado um novo rótulo $X_p^{k,m}$ à unidade referente a X_1.

Desta forma para cada produção p: $X \to X_1 \ldots X_n$, $n < m$ são obtidas as unidades $X_p^{k,m}$. Então definimos $R_m^k = \bigcup_{\substack{p \in P \\ X \in N}} \{X_p^{k,m}\}$.

Nas unidades $X_p^{k,m}$, o índice p identifica a produção utilizada para gerar a cadeia de terminais de comprimento m. No entanto, esta informação é

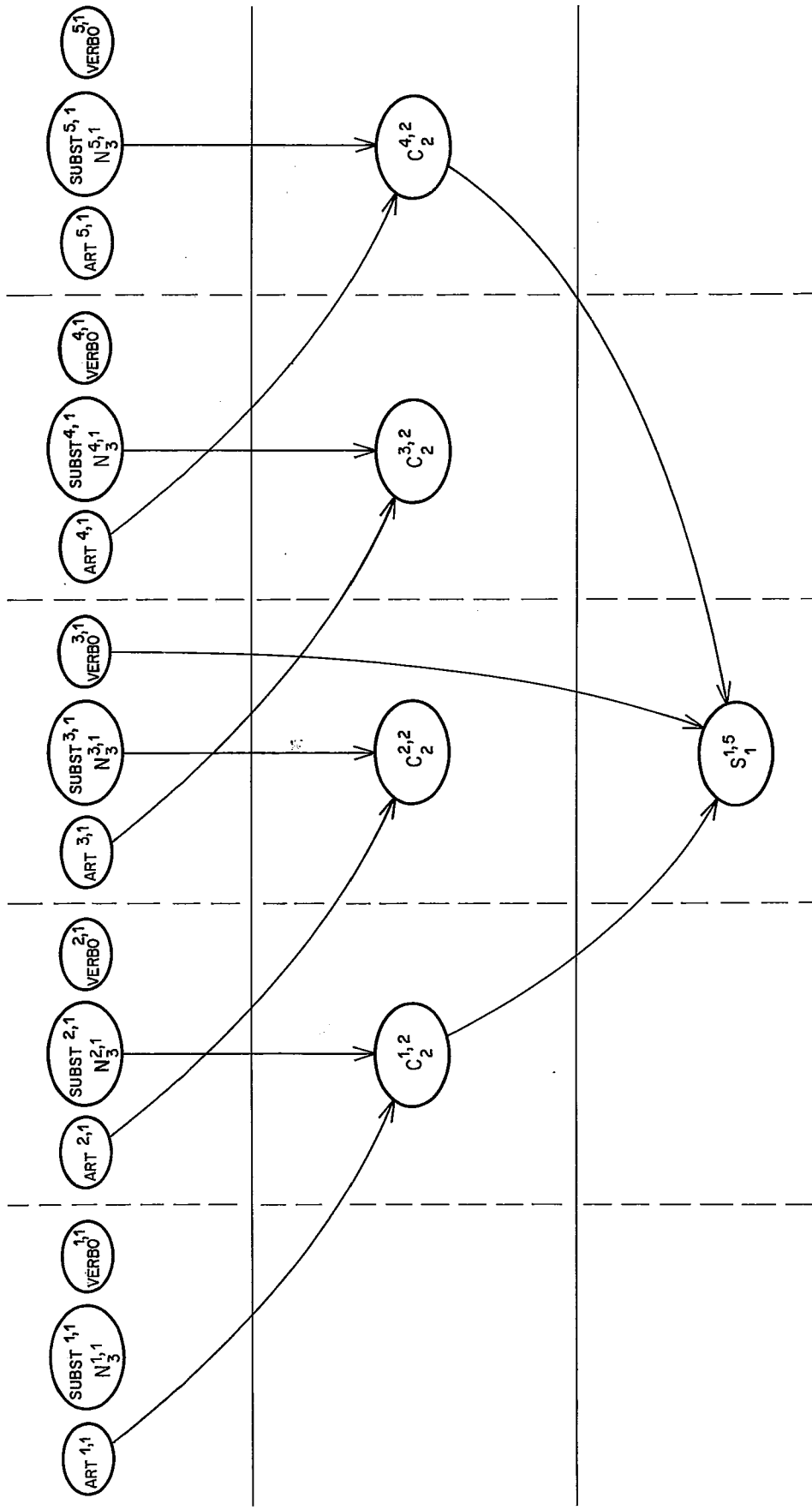

Fig. V.5 – Fragmento de uma rede sintática com cinco grupos de entrada mostrando a construção das unidades da rede para alguns terminais

desnecessária para concluirmos acerca da correção da entrada. Para tal, bastam o nome do não–terminal que gera a cadeia de comprimento m e a posição do seu primeiro símbolo (1,2, ..., ℓ–m+1).

Seja $C_x^{k,m} = \{X_p^{k,m} | p \in P \text{ e } p\colon X \rightarrow X_1 \ldots X_n\}$ o conjunto das unidades criadas no nível m para cada não terminal X e suas produções em P. Se $C_X^{k,m}$ tem somente uma unidade então existe apenas uma produção de X que gera cadeias de tamanho m. Nesse caso simplesmente renomeia–se a unidade $X_p^{k,m}$ com o rótulo $X^{k,m}$. Caso contrário, cria–se uma unidade $X^{k,m}$ filha–ou das unidades de $C_X^{k,m}$, e acrescenta–se esta nova unidade ao grupo R_m^k.

As unidades criadas para o símbolo inicial S e suas produções não são reunidas em uma filha–ou. Estas unidades contêm as informações de posição e categorias sintáticas que serão utilizadas pela análise semântica. Os diversos diagnósticos sintáticos estão diretamente ligados às produções de S e portanto estas informações não devem ser sumarizadas em uma única unidade.

A rede construída, $R = (R_1, \ldots, R_\ell)$, tem no máximo ℓ camadas e as unidades que resumem os diagnósticos sintáticos são $S_p^{1,m}$, $1 \leq m \leq \ell$ onde p é uma produção de S.

Algoritmo V.1

Entrada: Uma gramática livre de contexto $G = (N, \Sigma, P, S)$; $\ell \in \mathbb{N}$.

Saída: Uma rede Bayesiana $R = (R_1 \ldots R_\ell)$ que reconhece exatamente $L_\ell(G) = \{\omega \in L(G) \mid |\omega| \leq \ell\}$.

Método:

1) Elimine as produções ϵ usando o Algoritmo 2.10 em AHO e ULLMAN [1972]

2) Construa $R_1 = (R_1^1, \ldots, R_1^\ell)$ tal que $R_1^i = \{t^{i,1} | t \in \Sigma\}$ $i = 1, \ldots, \ell$.

3) Para todo $m = 1, \ldots, \ell$ faça:

3.1) Para todo $p \in P$, $p: X \rightarrow X_1 \ldots X_n$, faça:

Seja $J_n^m = \{j = (j_1, \ldots, j_n) \mid 1 \leq j_k \leq n \text{ e } \sum_{k=1}^{n} j_k = m\}$

Para cada $j = (j_1, \ldots, j_n) \in J_n^m$, sejam $m_1 = 0$ e

$m_i = \sum_{k=1}^{i-1} j_k$ para $1 < i \leq n$.

Se para todo $1 \leq i \leq n$ $X_i^{m+1,j_i} \in R_{j_i}^{m_i+1}$ então faça:

Para cada $1 \leq k \leq \ell - m + 1$
Se $n = 1$
então acrescente um rótulo $X_p^{k,m}$ à unidade $X_1^{k,m}$
senão se $X_p^{k,m} \in R_m^k$
então torne $X_p^{k,m}$ filha–e das unidades $X_1^{m_1+k,j_1}, \ldots, X_n^{m_n+k,j_n}$
senão faça:
crie uma unidade $X_p^{k,m}$ filha–e das unidades $X_1^{m_1+k,j_1}, \ldots, X_n^{m_n+k,j_n}$ e faça $R_m^k := R_m^k \cup \{X_p^{k,m}\}$.

Se $X = S$ e $k = 1$ então renomeie a unidade $S_p^{1,m}$ com $S_p^{j_1,\ldots,j_n}$.

3.2) Para cada $1 \leq k \leq \ell - m + 1$, $X \in N - \{S\}$ faça:

Seja $C_X^{k,m} = \{X_p^{k,m} \mid p \in P \text{ e } p: X \rightarrow X_1 \ldots X_n\}$ o conjunto de unidades criado para X em R_m^k.

Se $|C_X^{k,m}| = 1$
então renomeie a unidade $X_p^{k,m}$ com $X^{k,m}$
senão faça:

crie uma unidade $X^{k,m}$ filha–ou de todas as unidades de

$C_X^{k,m}$ e faça $R_m^k := R_m^k \cup \{X^{k,m}\}$

3.3) Faça $R_m := (R_m^1, ..., R_m^{\ell-m+1})$

4) Faça $R := (R_1, ..., R_\ell)$. Pare.

A rede construída através do Algoritmo V.1 reconhece a linguagem $L_\ell(G)$, porém possui uma série de unidades que são inúteis por representarem caminhos que não atingem nenhum diagnóstico $S_P^{1,m}$. Isto decorre do fato de que, em todos os grupos de entrada foram criadas unidades para todos os símbolos terminais, embora certos terminais, de acordo com a gramática, jamais ocorram em determinadas posições. Assim, no exemplo dado na Figura V.5 vemos que os terminais art e subst não ocorrem nas posições 2 e 3 respectivamente de uma sentença da linguagem. No entanto, a criação das unidades $art^{2,1}$ e $subst^{3,1}$ nos grupos de entrada, deu origem à unidade $C_2^{2,2}$, que não será ligada a nenhum diagnóstico sintático, sendo portanto inútil.

A construção dos grupos de entrada gerando apenas as unidades que efetivamente podem ocorrer naquela posição em uma sentença da linguagem é computacionalmente custosa. Por essa razão, optamos pela eliminação posterior dos caminhos inúteis através do Algoritmo V.2 descrito abaixo.

O algoritmo consiste em partir das unidades que são diagnósticos $S_P^{1,m}$ e marcar recursivamente seus pais, eliminando a seguir todas as unidades não marcadas.

Algoritmo V.2

Entrada: Uma rede Bayesiana $R = (R_1,...,R_\ell)$ construída usando o Algoritmo V.1 e que reconhece $L_\ell(G)$.

Saída: Uma rede Bayesiana sem unidades inúteis e que reconhece $L_\ell(G)$.

Método: 1) Para cada $S_P^{1,m}$, para $m = 1, ..., \ell$ marque todas as unidades da rede a partir das quais $S_P^{1,m}$ pode ser alcançado.

2) Elimine todas as unidades não marcadas.

Na construção da rede sintática através do Algoritmo V.1, as unidades são criadas de acordo com as produções da gramática. Para obter o máximo de compartilhamento das unidades como será visto a seguir, é importante que sejam identificadas as produções cujos lados–direitos possuem uma subcadeia comum de terminais e/ou não–terminais.

Assim, por exemplo, na gramática da Figura V.1 temos as produções 10 e 15 ($C \rightarrow$ art A N e $C \rightarrow$ A N), onde a subcadeia AN é comum. Esta cadeia pode ser substituída nas produções por um novo não–terminal X, acrescentando uma nova produção $X \rightarrow$ A N. Ou seja, reescrevemos as produções 10 e 15 como $C \rightarrow$ art X e $C \rightarrow$ X e acrescentamos a produção p: $X \rightarrow$ A N sem alterar a linguagem gerada pela gramática.

A Figura V.6 mostra um fragmento da rede gerada pelo algoritmo 1 considerando a gramática original e a Figura V.7 mostra a rede obtida a partir da gramática modificada.

Na Figura V.6 verificamos por exemplo, que a unidade $C_{10}^{1,3}$ não pode ser atualizada junto com as unidades $art^{1,1}$, $A^{2,1}$ e $N^{3,1}$ enquanto que na Figura V.7 a unidade $C_{10}^{1,3}$ não pode ser atualizada em paralelo apenas com as unidades $art^{1,1}$ e $X_p^{2,2}$. Assim, ao modificarmos a gramática, eliminando as subcadeias comuns, construímos uma rede mais eficiente do ponto de vista do paralelismo, já que diminuimos as vizinhanças significativas para atualização de cada unidade.

A seguir, descrevemos um algoritmo para eliminar as subcadeias comuns das produções de uma gramática sem alterar a linguagem gerada pela gramática.

Dada uma produção p: $X \rightarrow X_1 X_2 \ldots X_m$ definimos o conjunto de subcadeias de p de comprimento $0 < r \leq n$ por:

$$subc^r(p) = \{w \in (N \cup \Sigma)^* \mid |w| = r \text{ e } w = X_j \ldots X_{j+r-1} \text{ para algum } 1 \leq j \leq n-r+1\}$$

O algoritmo seleciona as produções que possuem uma subcadeia comum, começando a busca pelas cadeias mais longas, isto é, de n–1 até 2 onde n é o tamanho da maior produção.

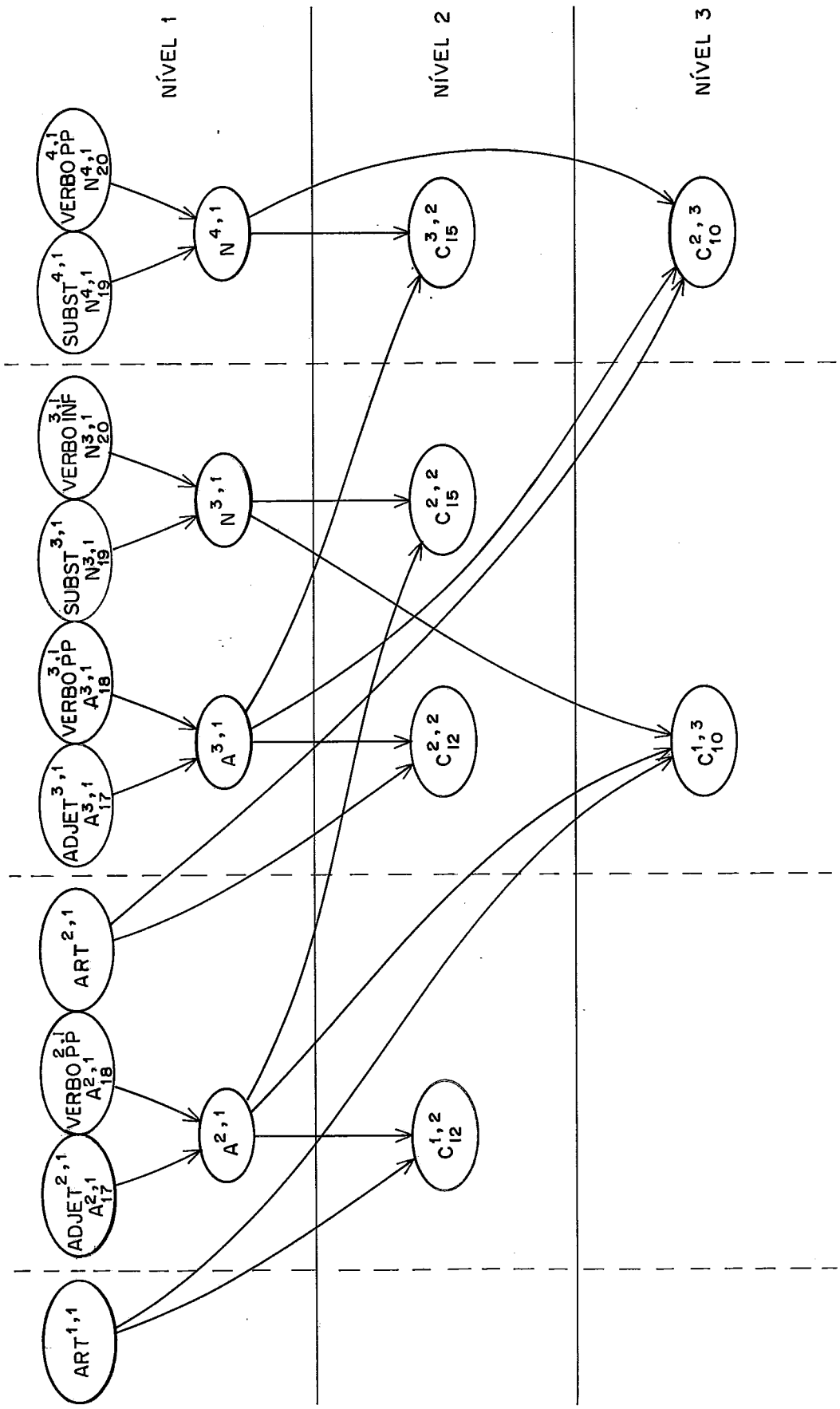

Fig. V.6 – Fragmento da rede sintática relativa à gramática da Figura V.1 e com quatro grupos de entrada

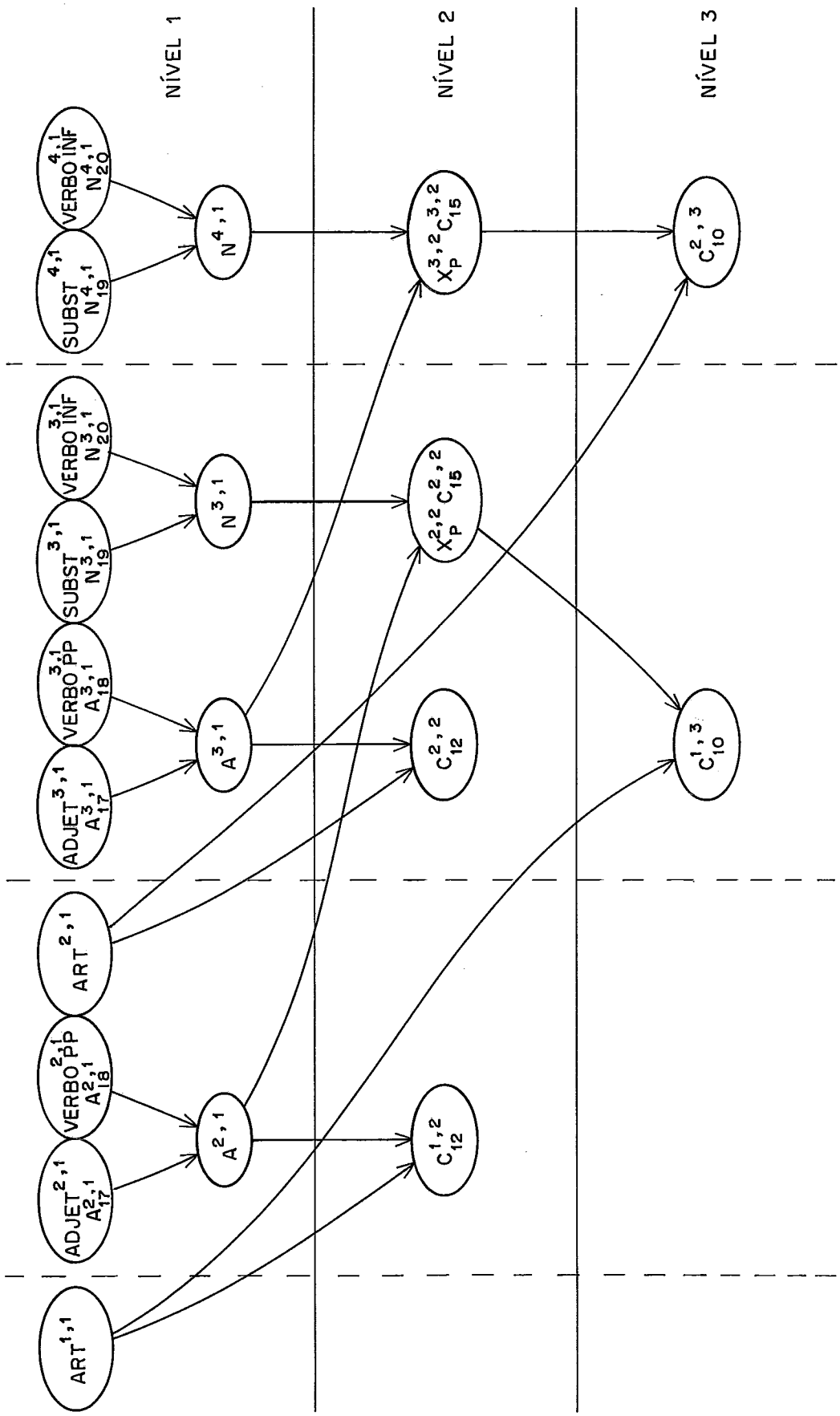

Fig. V.7 – Fragmento da rede sintática relativa à gramática da Figura V.1, modificada pelo algoritmo V.3

Inicialmente, as produções são ordenadas pelo tamanho e separadas em grupos de acordo com o tamanho. Então para cada $k = 2, \ldots m$, procura–se as produções que têm uma subcadeia comum w de comprimento k.

Fixado um k, verifica–se se existem duas ou mais produções de tamanho maior ou igual a k que tenham uma subcadeia comum w. Separado este conjunto T_w de produções que têm uma subcadeia comum w, cria–se um novo não–terminal X_w e acrescenta–se uma produção X_w. Além disso todas às produções que têm a subcadeia comum w são substituídas por uma produção de tamanho menor onde a subcadeia w foi substituída pelo não–terminal X_w.

Algoritmo V.3

Entrada: Uma gramática livre do contexto $G = (N, \Sigma, P, S)$

Saída: Uma gramática livre do contexto $G' = (N', \Sigma, P', S)$ com as sub–cadeias comuns eliminadas das produções e tal que $L(G') = L(G)$.

Método:

1) Seja $P_s = \{p \in P \mid p: S \rightarrow w\}$

2) Ordene as produções de $P - P_s$ por tamanho: $(P_1, \ldots, P_n)$ onde $P_i = \{p \in P - P_s \mid p: X \rightarrow w \text{ e } |w| = i\}$, $i = 1, \ldots, n$.

3) Se $n \leq 2$ pare (não existem sub–cadeias para eliminar).

4) Faça $N' := N$; $P'_i := P_i$, $i = 1, \ldots, n$;

5) Para $k = n–1, \ldots, 2$, faça:

5.1) $subc^k := \bigcup_{p \in P} subc^k(p)$;

5.2) Para todo $w \in subc^k$ faça:

5.2.1) Para todo $p, q \in \bigcup_{i=k}^{n} P'_i$ faça:
se $w \in subc^k(p) \cap subc^k(q)$

então $T_w := T_w \cup \{p,q\}$

5.2.2) Seja w_0 tal que $|T_{w_0}| = \max_w |T_w|$

5.3) Se $|T_{w_0}| > 0$ então faça

5.3.1) Crie um novo não–terminal X_{w_0}.

5.3.2) Faça $N' := N' \cup \{X_{w_0}\}$

5.3.3) Faça $P'_k = P'_k \cup \{X_{w_0} \rightarrow w_0\}$

5.3.4) Para cada $q \in T_{w_0}$ tal que $q : Y \rightarrow tw_0Z$ e $q \in P'_j$ para algum $j \geq k$ faça:

$P'_j = P'_j - \{q\}$

$P'_{j-k+1} = P'_{j-k+1} \cup \{Y \rightarrow t\, X_{w_0}\, Z\}$

6) Faça $P' = (\bigcup_{i=1}^{n} P'_i) \cup P_s$. Pare, $G' = (N', \Sigma, P', S)$ está com as subcadeias eliminadas e $L(G') = L(G)$.

CAPÍTULO VI

A SUB–REDE PARA ANÁLISE SEMÂNTICA

VI.1 – INTRODUÇÃO

No Capítulo IV foi apresentada a estrutural geral da rede incluindo uma descrição do módulo de análise semântica e sua interação com a rede sintática. Neste capítulo descreveremos com mais detalhes os vários níveis que compõem a rede responsável pela análise semântica. Mostraremos como devem ser estabelecidas as probabilidades condicionais que regulam o fluxo da informação na rede e as relações de causa/efeito.

Apresentaremos finalmente alguns exemplos completos de sentenças com diversos tipos de ambigüidade e descreveremos como a ambigüidade é resolvida.

VI.2 – DESCRIÇÃO GERAL

A interpretação semântica das sentenças é feita tendo como inspiração o modelo de FILLMORE [1968] que associa aos verbos uma certa estrutura de casos. Porém, da mesma forma que a maioria dos sistemas de compreensão das linguagens naturais desenvolvidos na área da Inteligência Artificial, utilizamos uma gramática de casos de forma pouca rigorosa e sem nenhuma pretensão de apresentar uma proposta lingüística. De fato, a gramática de casos, tal como foi definida por Fillmore, é muito geral e não incorpora aos casos características semânticas suficientes para que um sistema automático seja capaz de obter um diagnóstico semântico para uma sentença.

Utilizamos, então, os casos verbais de uma forma semelhante à proposta por COTTRELL [1985], ou seja, agrupando casos de verbos que tenham características semânticas semelhantes. Por exemplo, os agentes dos verbos "comer", "viver" e "dormir" são idênticos e correspondem a um ser animado. Então, definimos uma unidade "a–AG" para representar o caso agente destes verbos. Esta unidade é ligada à rede semântica usada para representar o conhecimento, como filha da unidade "ser–animado" que representa a classe dos seres animados.

Os verbos "estudar", "comprar", "pensar", têm como agente um ser humano. Representamos este caso por uma unidade "h–AG" que é filha da unidade

"humano" da rede semântica. Além disso, é óbvio que todo agente que satisfaz às restrições semânticas de "h–AG" também satisfaz às de "a–AG". Por isso, criamos uma hierarquia de casos, colocando a unidade "a–AG" como filha da unidade "h–AG".

Cada unidade que representa um determinado significado de um verbo tem como pais os seus casos correspondentes. Para que este significado ocorra com probabilidade alta é necessário que todos os seus casos obrigatórios estejam ocorrendo com probabilidade alta.

Este comportamento é obtido especificando convenientemente as probabilidades condicionais associadas às unidades de significado verbal.

Na Figura VI.1 reapresentamos a organização da rede que executa a análise semântica mostrando a interligação dos seus níveis. A seguir, descrevemos com mais detalhes a estrutura de cada nível e a forma como devem ser estabelecidas as probabilidades condicionais. Para tal, utilizamos a seguinte gramática livre de contexto que é uma restrição da gramática V.1:

1) S → C verbo prep C
2) S → C verbo C prep C
3) S → C verbo prep C C
4) C → art A N
5) C → art A
6) C → art N
7) C → N
8) A → adjet
9) N → subst

A rede sintática construída para esta gramática, considerando frases de no máximo seis palavras, tem as seguintes unidades de diagnóstico sintático:

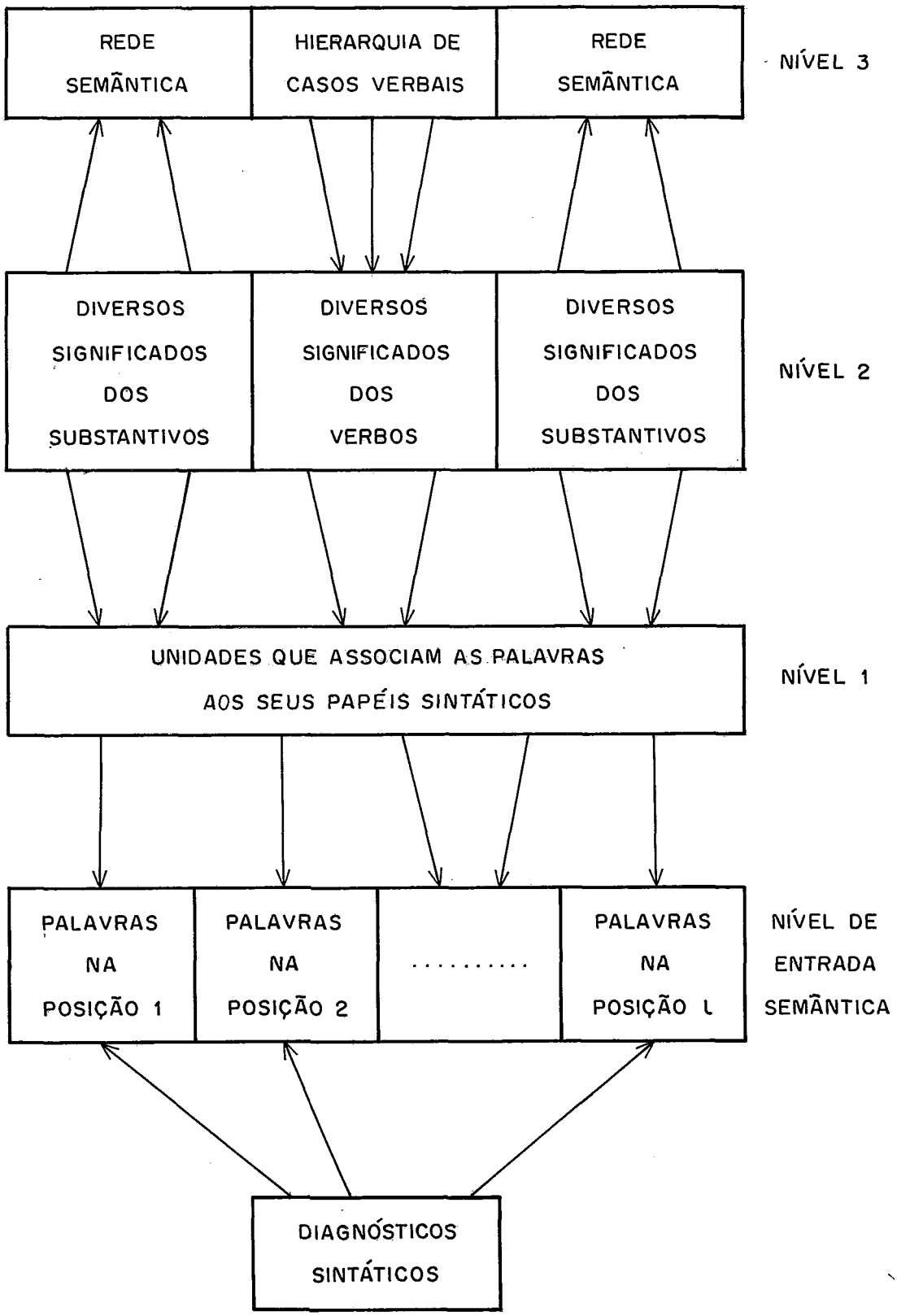

Fig. VI.1 – Organização geral da rede responsável pela análise semântica

UNIDADES		PAIS
E1	$S_1^{(1,1,1,1)}$	$C^{1,1}$, $verbo^{2,1}$, $prep^{3,1}$, $C^{4,1}$
E2	$S_1^{(1,1,1,2)}$	$C^{1,1}$, $verbo^{2,1}$, $prep^{3,1}$, $C^{4,2}$
E3	$S_1^{(1,1,1,3)}$	$C^{1,1}$, $verbo^{2,1}$, $prep^{3,1}$, $C^{4,3}$
E4	$S_1^{(2,1,1,1)}$	$C^{1,2}$, $verbo^{3,1}$, $prep^{4,1}$, $C^{5,1}$
E5	$S_1^{(2,1,1,2)}$	$C^{1,2}$, $verbo^{3,1}$, $prep^{4,1}$, $C^{5,2}$
E6	$S_1^{(3,1,1,1)}$	$C^{1,3}$, $verbo^{4,1}$, $prep^{5,1}$, $C^{6,1}$
E7	$S_2^{(1,1,1,1,1)}$	$C^{1,1}$, $verbo^{2,1}$, $C^{3,1}$, $prep^{4,1}$, $C^{5,1}$
E8	$S_2^{(1,1,2,1,1)}$	$C^{1,1}$, $verbo^{2,1}$, $C^{4,2}$, $prep^{5,1}$, $C^{6,1}$
E9	$S_2^{(1,1,1,1,2)}$	$C^{1,1}$, $verbo^{2,1}$, $C^{3,1}$, $prep^{4,1}$, $C^{5,2}$
E10	$S_2^{(2,1,1,1,1)}$	$C^{1,2}$, $verbo^{3,1}$, $C^{4,1}$, $prep^{5,1}$, $C^{6,1}$
E11	$S_3^{(1,1,1,1,2)}$	$C^{1,1}$, $verbo^{2,1}$, $prep^{3,1}$, $C^{4,1}$, $C^{5,2}$
E12	$S_3^{(1,1,1,2,1)}$	$C^{1,1}$, $verbo^{2,1}$, $prep^{3,1}$, $C^{5,2}$, $C^{6,1}$
E13	$S_3^{(1,1,1,1,1)}$	$C^{1,1}$, $verbo^{2,1}$, $prep^{3,1}$, $C^{4,1}$, $C^{5,1}$
E14	$S_3^{(2,1,1,1,1)}$	$C^{1,2}$, $verbo^{3,1}$, $prep^{4,1}$, $C^{5,1}$, $C^{6,1}$

Tais diagnósticos são pais das unidades de entrada semântica, que descrevemos a seguir.

VI.2.1 – Nível de Entrada Semântica

Este nível é composto de unidades que relacionam cada palavra com a sua posição na frase. Estas unidades não têm filhos e devem ser sempre grampeadas de acordo com a frase a ser analisada.

Consideremos uma palavra qualquer do dicionário, por exemplo "casa". Então, se a rede analisa sentenças de tamanho máximo seis construímos o nível de entrada composto de seis grupos e para cada i = 1,...,6 criamos uma unidade de entrada "casa i".

A palavra "casa" tem duas classificações gramaticais, a saber: verbo e substantivo. Então a palavra "casa" pode ser o verbo principal da sentença ou integrar um dos sintagmas nominais que compõem a sentença. A tabela abaixo

mostra os pais na rede semântica das unidades "casa i", i = 1,...,6. Estes pais são construídos de acordo com os diagnósticos sintáticos, conforme veremos mais adiante.

UNIDADE	FILHA DE
casa 1	casa NP0
casa 2	casa NP0, casa V
casa 3	casa NP0, casa V, casa NP1
casa 4	casa V, casa NP1
casa 5	casa NP1, casa NP2
casa 6	casa NP1, casa NP2

Por exemplo, a unidade "casa 3" tem a unidade "casa NP0" como pai porque a palavra "casa" como terceira palavra de uma frase pode integrar o sintagma nominal que ocorre antes do verbo. No entanto, "casa NP2" não é pai de "casa 3", porque, pelas regras da gramática, uma palavra na terceira posição da frase não pode fazer parte do segundo sintagma nominal depois do verbo.

Além disso, as unidades "casa i", i = 1,...,6, devem ter como pais os diagnósticos sintáticos nos quais é viável que ocorra substantivo ou verbo na posição i da entrada sintática. Estes são os diagnósticos sintáticos que têm chance de representar uma estrutura sintática coerente com a entrada semântica.

Se fixarmos, por exemplo, i = 3 e considerarmos a palavra "casa" temos que a unidade "casa 3" é filha dos seguintes diagnósticos:

E4, E5, E6, E7, E9, E10, E14.

Os diagnósticos E4, E5, E10 e E14 são pais de "casa 3" porque verbo é uma das classes gramaticais da palavra "casa" e a unidade da rede sintática "verbo^{3,1}" é pai destes diagnósticos.

O diagnóstico E6 tem como um de seus pais a unidade $C^{1,3}$ da rede sintática. Como $C \overset{*}{\rightarrow}$ art adjet subst, temos que existe um caminho entre a unidade de entrada sintática subst^{3,1} e o diagnóstico E6. Então, como uma das classes gramaticais da palavra "casa" é substantivo, devemos incluir E6 como pai de "casa 3".

Os diagnósticos E7 e E9 são pais de "casa 3" por razões idênticas às expostas acima.

Concluindo, temos que a unidade "casa 3" tem os seguintes pais:

"casa NP0", "casa NP1", "casa V", E4, E5, E6, E7, E9, E10 e E14.

Em geral, um diagnóstico E é pai de uma unidade de entrada semântica C_i se a palavra C pode ter a categoria gramatical X e há uma trajetória da unidade de entrada sintática $X^{i,1}$ até E.

A matriz abaixo representa os estados dos pais para os quais a probabilidade de ocorrência da unidade "casa 3" é alta.

	casa NP0	casa NP1	casa V	E4	E5	E6	E7	E9	E10	E14
1	x	y	1	1	0	0	0	0	0	0
2	x	y	1	0	1	0	0	0	0	0
3	1	x	y	0	0	1	0	0	0	0
4	x	1	y	0	0	0	1	0	0	0
5	x	1	y	0	0	0	0	1	0	0
6	x	y	1	0	0	0	0	0	1	0
7	x	y	1	0	0	0	0	0	0	1

onde $x, y \in \{0, 1\}$.

A escolha dessa matriz deve–se às relações entre um determinado diagnóstico sintático (dependente da posição das palavras na frase) e os possíveis papéis funcionais da palavra. Então, por exemplo, as configurações explicitadas pela linha 4 significam que a probabilidade de ocorrência da unidade "casa 3" é alta, se o diagnóstico sintático mais provável for E7 e a unidade do nível 1 mais provável for "casa NP1".

Os valores de "casa NP0" e "casa V" não são importantes nesse caso. De fato, isto significa que se "casa" é parte integrante do primeiro sintagma nominal depois do verbo e E7 é o diagnóstico sintático correto então "casa 3" tem uma probabilidade alta quaisquer que sejam os valores de "casa NP0" e "casa V". Se

uma destas unidades ocorrer muito freqüentemente isto significa que a palavra "casa" ocorre também em outra posição da frase. É o caso, por exemplo, de frases tais como: "Quem casa quer casa" e "O amo ama a ama".

A seguir, descrevemos o algoritmo para construir as ligações entre a rede sintática e a semântica, bem como para estabelecer as matrizes de probabilidade condicional associadas às unidades de entrada semântica.

Os diagnósticos sintáticos definidos no Capítulo V estabelecem as relações entre as categorias gramaticais e os constituintes estruturais da sentença (NP0, NP1, P1, etc.). Então, dado um diagnóstico sintático $S_p^{j_1,\ldots,j_n}$, $q_t = \sum_{k=1}^{t} j_k$ indica a posição na entrada do início de cada constituinte, $t = 1, \ldots, n$.

Dado o conjunto D de diagnósticos sintáticos para uma entrada de tamanho máximo ℓ, dada uma palavra "pal" e as suas possíveis categorias gramaticais, o Algoritmo VI.1 identifica para cada $i = 1, \ldots, \ell$ quais os diagnósticos devem ser pais de "pal_i" e quais unidades representando constituintes ("pal NP0", "pal V", etc.) devem ser criadas como pais de "pal_i". Quando estabelecemos os pais de "pal_i", criamos um conjunto $I(pal_i)$ que contém as duplas relacionando um dado diagnóstico e as unidades de constituintes que foram criadas associadas ao diagnóstico.

Assim, inicialmente, se existe um caminho ligando uma das categorias gramaticais "$cat^{q_i,1}$" de "pal_i" a um diagnóstico $S_p^{j_1\ldots j_n}$ então acrescentamos este diagnóstico como pai de "pal_i".

A seguir, verifica-se (de acordo com este diagnóstico sintático) a qual constituinte a categoria está associada. Cria-se então a unidade "palX", $X \in \{NP0, V, NP1, NP2, P1, P2\}$ correspondente. Além disso, acrescenta-se a $I(pal_i)$ a dupla $(S_p^{j_1\ldots j_n}, palX)$. O conjunto $I(pal_i)$ será utilizado no último passo do algoritmo para estabelecer os valores das probabilidades condicionais associadas a "pal_i".

Algoritmo VI.1

Entrada: O conjunto D de diagnósticos sintáticos de entrada de tamanho máximo ℓ, uma palavra pal do dicionário e suas categorias gramaticais $cat_1, \ldots, cat_m$.

Saída: Para cada $i = 1,\ldots,\ell$ o conjunto de pais de pal_i, $P(pal_i)$ e as configurações de $P(pal_i)$ que têm probabilidade alta.

Método:

1) Para cada $i = 1,\ldots,\ell$ faça $P(pal_i) = \phi$, $I(pal_i) = \phi$

2) Para cada diagnóstico sintático $S_p^{j_1 \cdots j_n} \in D$ faça:

2.1) Seja $q_t = \sum_{k=1}^{t} j_k$, $t = 1,\ldots,n$, a posição a que pertence o pai t de $S^{j_1,\ldots,j_n}$. Para cada $i = 1, \ldots,n$ e $n = 1, \ldots, m$ faça:

2.1.2) Se existe um caminho na rede sintática ligando a unidade de entrada sintática $(cat_r)^{q_i,1}$ a $S_p^{(j_1 \cdots j_n)}$ então faça:

2.1.2.1) $P(pal_i) = P(pal_i) \cup \{S_p^{(j_1,\ldots j_n)}\}$

2.1.2.2) Se cat_r = verbo, então X = 'V'

2.1.2.3) Seja $1 \leq h \leq n$ tal que a entrada sintática $verbo^{q_h,1}$ é pai de $S_p^{j_1 \cdots j_n}$.

2.1.2.4) Se cat_r = prep,

então se $i = h + 1$

então X = 'P1'

senão X = 'P2'

2.1.2.5) Se $cat_r \in$ {subst, adjet, verbopp, verboinf},

então se $i < h$,

então X = 'NPO'

senão se $h < i < n-1$

então X = 'NP1'

senão X = 'NP2'

2.1.2.6) palX := concat('pal',X)

2.1.2.7) Se ainda não existir uma unidade palX no nível 1 crie esta unidade

2.1.2.8) $P(pal_i) := P(pal_i) \cup \{palX\}$

2.1.2.9) $I(pal_i) := I(pal_i) \cup \{(S_p^{(j_1 \ldots j_n)}, palX)\}$

3) Para cada $i = 1, \ldots, \ell$ defina a matriz de probabilidade condicional de pal_i da seguinte forma:

Para cada $T = (T_1, T_2) \in I(pal_i)$ faça:

$P(pal_i = 1 | T_1 = 1,\ T_2 = 1,$ todos os outros pais que são diagnósticos sintáticos com valor 0, todos os pais restantes com valor qualquer) é alta.

VI.2.2 – Nível 1 (Relações Palavra x Papel Sintático)

As unidades que compõem este nível são obtidas mediante o algoritmo apresentado na seção anterior. Sendo assim, em geral, uma palavra "pal" que pode integrar um sintagma nominal aparece representada nesse nível com três unidades: "pal NP0", "pal NP1" e "pal NP2". Se a palavra for um verbo aparece como "pal V", se for preposição como "pal P1" e "pal P2".

Os pais dessas unidades correspondem aos possíveis significados da palavra com uma determinada classe gramatical. Assim, por exemplo, consideremos o

substantivo "banco". Este substantivo gerou três unidades no nível 1: "banco NP0", "banco NP1", "banco NP2".

O substantivo "banco" tem dois significados homônimos que devem ser representados no nível 2 da rede: ban 1 (objeto para sentar) e ban 2 (instituição financeira). Esta ambigüidade da palavra "banco" existe independente do sintagma nominal que a contém. Então, a construção das unidades de significado que será descrita abaixo deve ser repetida para cada sintagma nominal previsto pela gramática.

Tomemos, por exemplo, a unidade "banco NP0". Criamos duas unidades do nível 2 relativos aos significados da palavra "banco": "ban 1" e "ban 2". Estas unidades são pais da unidade "banco NP0".

Ao estabelecermos a matriz de probabilidades condicionais associadas à unidade "banco NP0" obrigamos a escolha de um dos significados como o adequado para o contexto. Especificamos então um valor alto para as probabilidades condicionais associadas a "banco NP0" nos seguintes casos:

ban 1	**ban 2**
1	0
0	1

Ou seja, a probabilidade de ocorrência da unidade "banco NP0" será alta se os estados mais prováveis dos seus pais são (ban 1 = 0 é ban 2 = 1) ou (ban 1 = 1 e ban 2 = 0). Portanto, dado um estado global da rede (definindo um contexto), a palavra "banco" deve assumir exclusivamente um dos significados "ban 1" ou "ban 2".

Os diversos significados de uma palavra como verbo também estão representados por unidades do nível 2 da rede. Assim, por exemplo, consideremos os seguintes significados polissêmicos do verbo "bater":

bat 1 – golpear para emitir som
bat 2 – surrar alguém
bat 3 – fechar (a porta, a janela)
bat 4 – vencer alguém (no jogo ou por mérito)

A unidade "bater V" tem como pais as unidades "bat 1", bat 2", "bat 3", e "bat 4", referentes aos sentidos do verbo "bater".

Fixada uma entrada, desejamos que um significado verbal destaque–se dos demais resolvendo assim a ambigüidade verbal de acordo com o contexto. Para isso, as probabilidades condicionais associados à unidade "bater V" são altas apenas para as configurações abaixo:

bat 1	bat 2	bat 3	bat 4
1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1

A Figura VI.2 mostra um fragmento da rede considerando seis grupos de entrada e as palavras "banco" e "bater".

Finalmente, cada preposição do nível de entrada dá origem a duas unidades no nível 1. Assim, por exemplo, para a preposição "em" temos no nível 1 as unidades "em P1" e "em P2" especificando se a preposição se refere ao primeiro ou ao segundo sintagma nominal depois do verbo (NP1 ou NP2), respectivamente.

As unidades do nível 1 referentes às preposições não têm pais no nível 2. Estas unidades estão ligadas diretamente às unidades do nível 3 relativas aos casos verbais.

VI.2.3 – Níveis 2 e 3

O nível 2 é composto pelas unidades de significado dos sintagmas nominais e do verbo principal, definidas na seção anterior. Conforme observamos, as unidades referentes aos significados dos sintagmas nominais devem ser repetidas de acordo com o número de sintagmas nominais. Existem, então, dois tipos de unidades no nível 2: as de significado verbal e as de significado nominal.

As unidades de significado nominal não tem pais. Estas unidades têm um filho do nível 1 (tipo "pal NP") e o(s) outro(s) filho(s) são unidades da rede semântica (nível 3) utilizada para representar o conhecimento. A figura VI.3 mostra a parte da rede referente ao substantivo "banco" como parte do sintagma

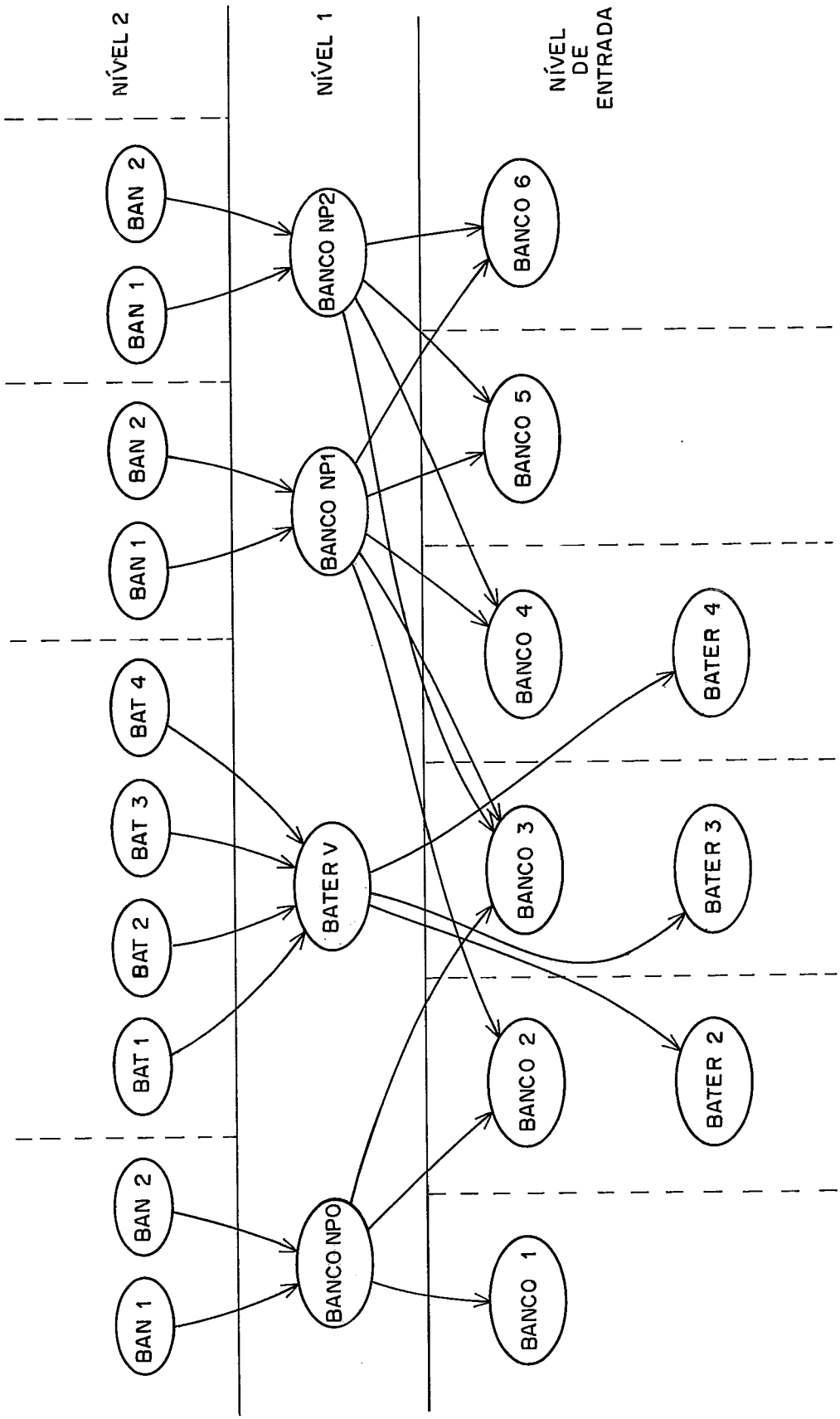

Fig. VI.2 – Fragmento da rede que executa a análise semântica considerando seis grupos de entrada e os níveis 1 e 2 da rede para as palavras "banco" e "bater"

nominal antes do verbo.

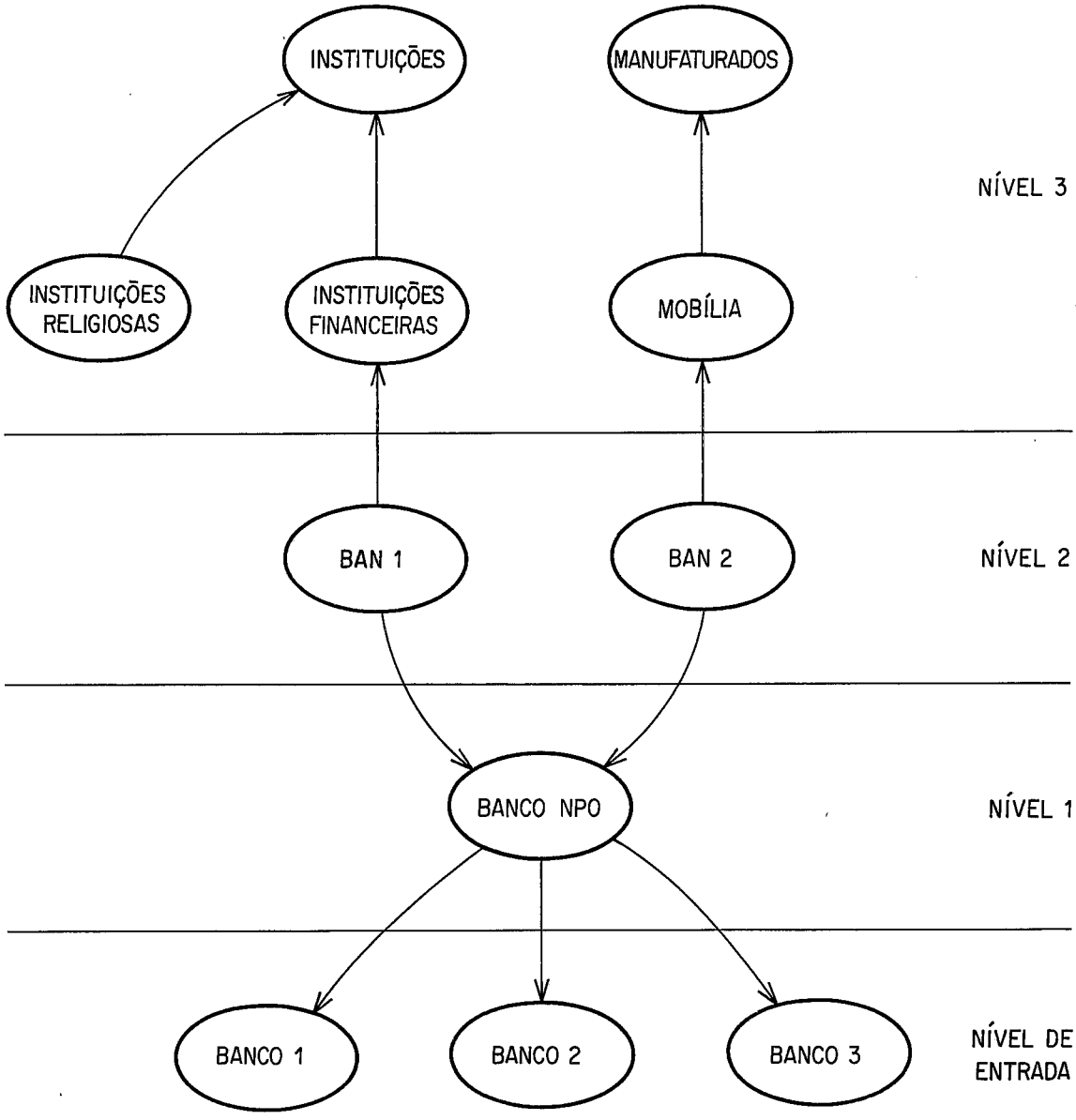

Fig. VI.3 – Parte da rede referente ao substantivo "banco" como sintagma nominal antes do verbo e considerando os níveis 1, 2, e 3

Como vemos, as unidades "ban 1" e "ban 2" não têm pais, portanto, devemos dar um valor às probabilidades a priori destas unidades. A experiência adquirida na implementação do modelo indica que estes valores devem ser pequenos para que a rede atinja mais rapidamente um estado estável.

A representação do conhecimento ligada aos substantivos consiste em uma rede semântica onde o conhecimento está representado de forma classificatória. As unidades relativas aos casos verbais compõem uma hierarquia de casos. Estas unidades são filhas das unidades de classificação da rede semântica. Então, a

ambigüidade da palavra "banco" é resolvida de acordo com o verbo da frase. Por exemplo, consideremos as frases:

"Sentei num banco" e "Entrei num banco"

O verbo "sentar" tem como caso objeto o componente de uma mobília (mesa, cadeira, banco, etc.), enquanto que o verbo "entrar" pressupõe um caso objeto como uma construção ou um veículo. Então, dependendo do verbo, obtém–se um caminho de unidades que ocorrem freqüentemente ligando o verbo à unidade "ban 1" ou à unidade "ban 2".

As unidades de significado verbal têm como pais unidades que representam os casos associados a um determinado sentido do verbo. Consideremos, por exemplo, o verbo "bater" e o seu significado bat 2 (surrar alguém).

A Figura VI.4 mostra um fragmento da rede, descrevendo as ligações da unidade "bat 2" com seus respectivos casos. Verificamos que a ordem em que aparecem os casos verbais pode variar. Assim, poderíamos ter a frase "João bateu com o chinelo em Maria" e dessa forma o caso instrumento seria o primeiro sintagma nominal depois do verbo e o caso objeto seria o segundo. Este fato é que explica a existência de duas unidades para representar o caso objeto ("b–OBJ 1" e "b–OBJ 2") e duas unidades para representar o caso instrumento ("b–INST 1" e "b–INST 2").

As probabilidades condicionais associadas a unidade "bat 2" são altas apenas nas configurações descritas abaixo:

b–AG	b–OBJ1	b–OBJ2	b–INST1	b–INST2
1	1	0	0	1
1	0	1	1	0
1	1	0	0	0
0	1	0	0	1
0	0	1	1	0
0	1	0	0	0

O único caso obrigatório para o significado "bat 2" do verbo "bater" é o caso objeto. Por isso, em qualquer linha da matriz acima temos sempre

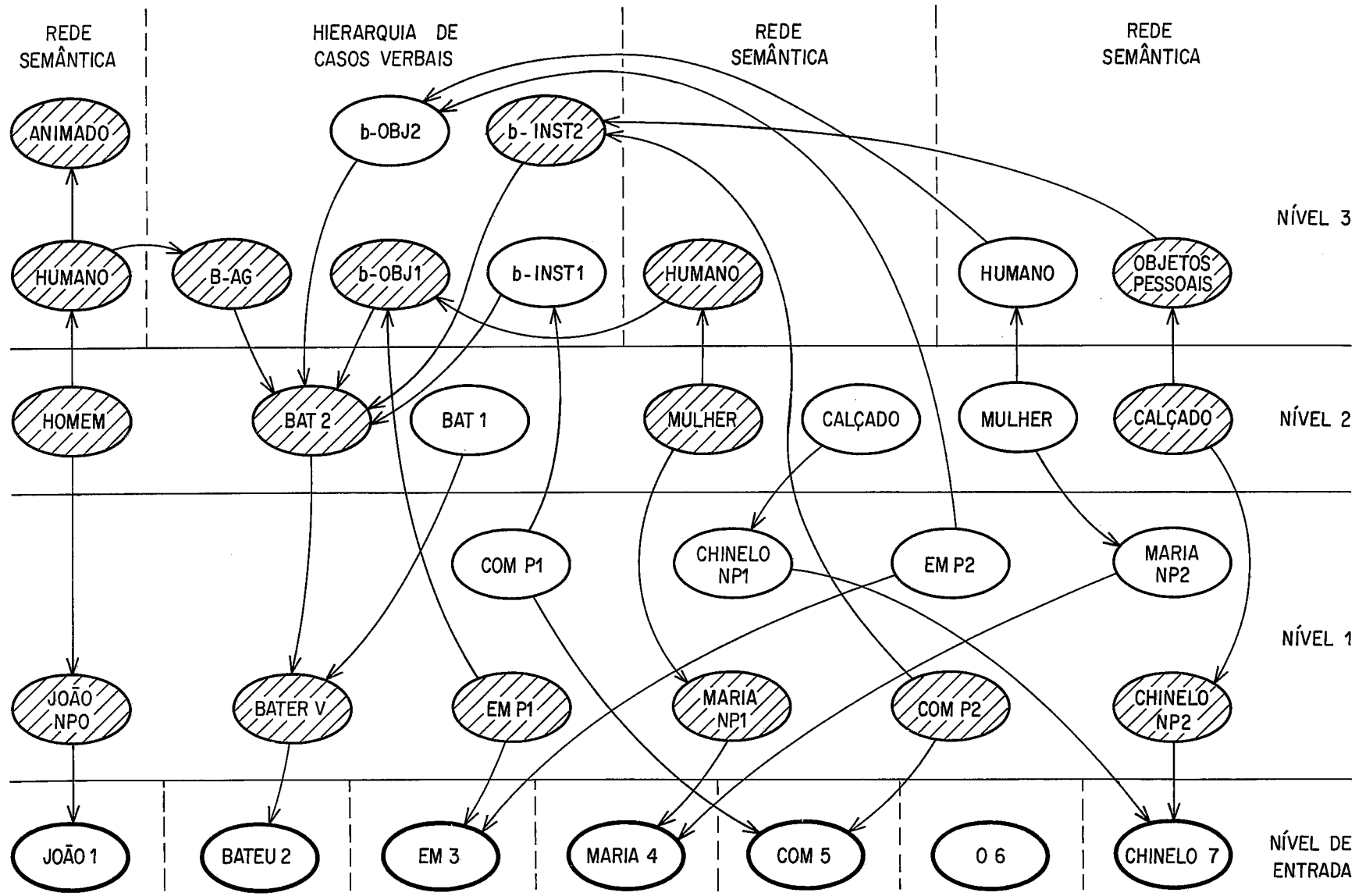

Fig. VI.4 — Fragmento da rede que executa a análise semântica considerando apenas a entrada "João bateu em Maria com o chinelo" com os diversos significados do verbo "bater" nos níveis 1, 2 e 3 da rede

"b–OBJ1" = 1 ou "b–OBJ2" = 1. Os casos agente e instrumento não são obrigatórios. Observamos que se o caso objeto se refere ao primeiro sintagma nominal depois do verbo ("b–OBJ1" = 1) então o caso instrumento, se existir, deve se referir ao segundo ("b–INST2" = 1).

As unidades hachuradas na Figura VI.4 são as que ocorrem mais freqüentemente para a entrada especificada (unidades em negrito). Vemos, pela figura, os caminhos que ligam as unidades de entrada com os casos verbais e finalmente com um determinado significado verbal (no caso bat 2). No exemplo, a configuração dos pais da unidade "bat 2" é a prevista na primeira linha da matriz.

VI.3 – ALGUNS EXEMPLOS DE SOLUÇÃO DA AMBIGÜIDADE

Exemplo 1:

Consideremos as frases:

"Maria gostou dos belos livros" e
"Maria deu ao amado livros"

A entrada sintática para estas frases é a mesma, a saber: subst1, verbo2, prep3, art4, adjet5 e subst6. A gramática apresentada na Figura VI.1 é ambígüa e para esta entrada a rede sintática fornece dois diagnósticos igualmente prováveis. No primeiro, E12, temos dois sintagmas nominais depois do verbo: "o amado" e "livros". No segundo, E3, teremos apenas um sintagma nominal depois do verbo: "os belos livros".

A ambigüidade de diagnósticos sintáticos é resolvida através da análise semântica conforme mostram as Figuras VI.5 e VI.6. Estas figuras representam o estado estável da rede para as entradas assinaladas. As unidades hachuradas ocorrem mais freqüentemente que as demais e as unidades de entrada em negrito são as que estão com valor 1.

O verbo "gostar" exige apenas um caso obrigatório, o caso objeto ("g–OBJ 1") e, além disso, a preposição associada a este caso é "de". Então, o diagnóstico semântico, para a primeira frase, reforça a ocorrência da unidade "livros NP1". Daí decorre que o diagnóstico sintático adequado, E3, ocorre mais

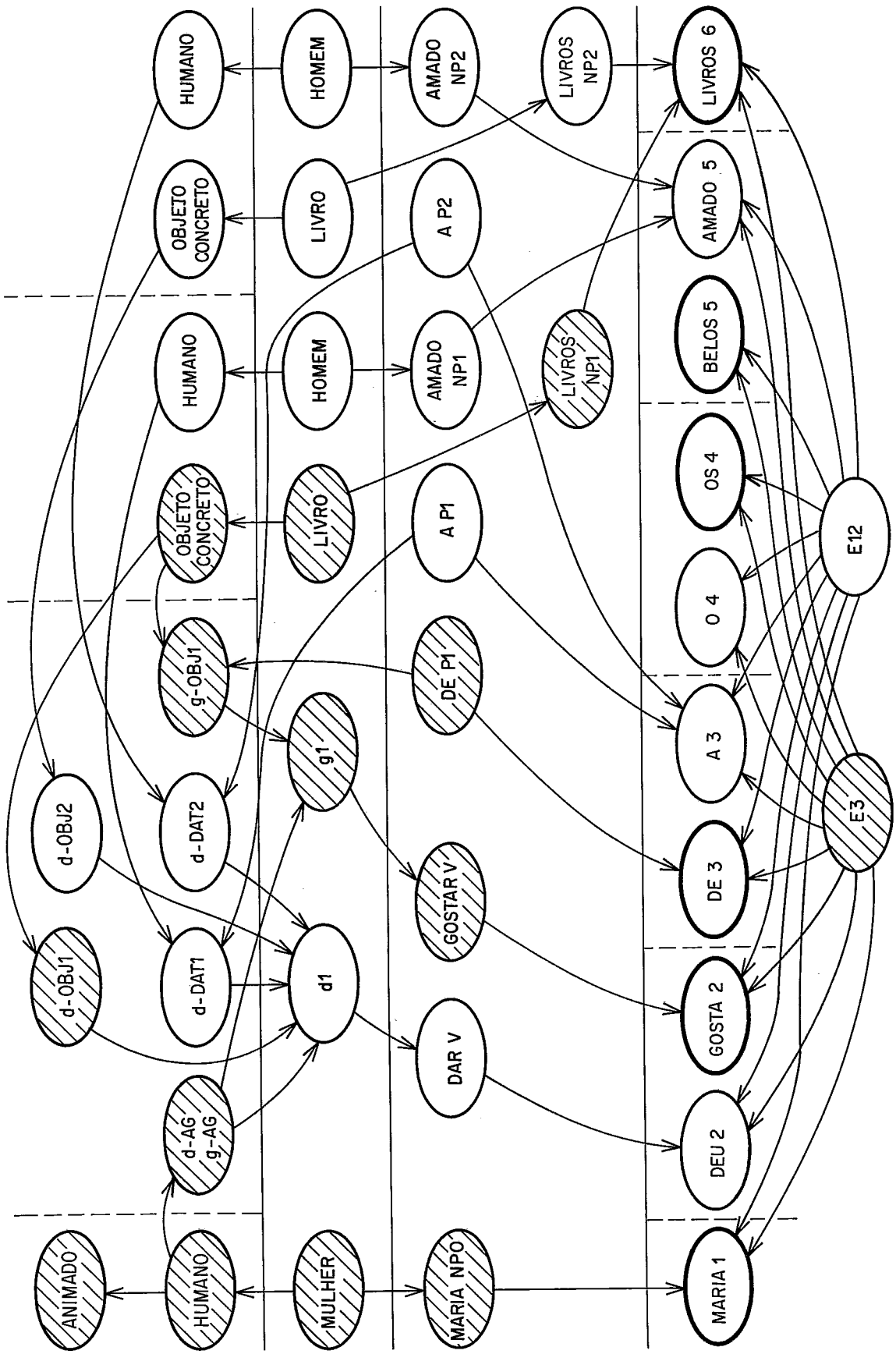

Fig. VI.5 – Exemplo do comportamento da rede que executa a semântica na solução de uma ambigüidade sintática da sentença "Maria gostou dos belos livros"

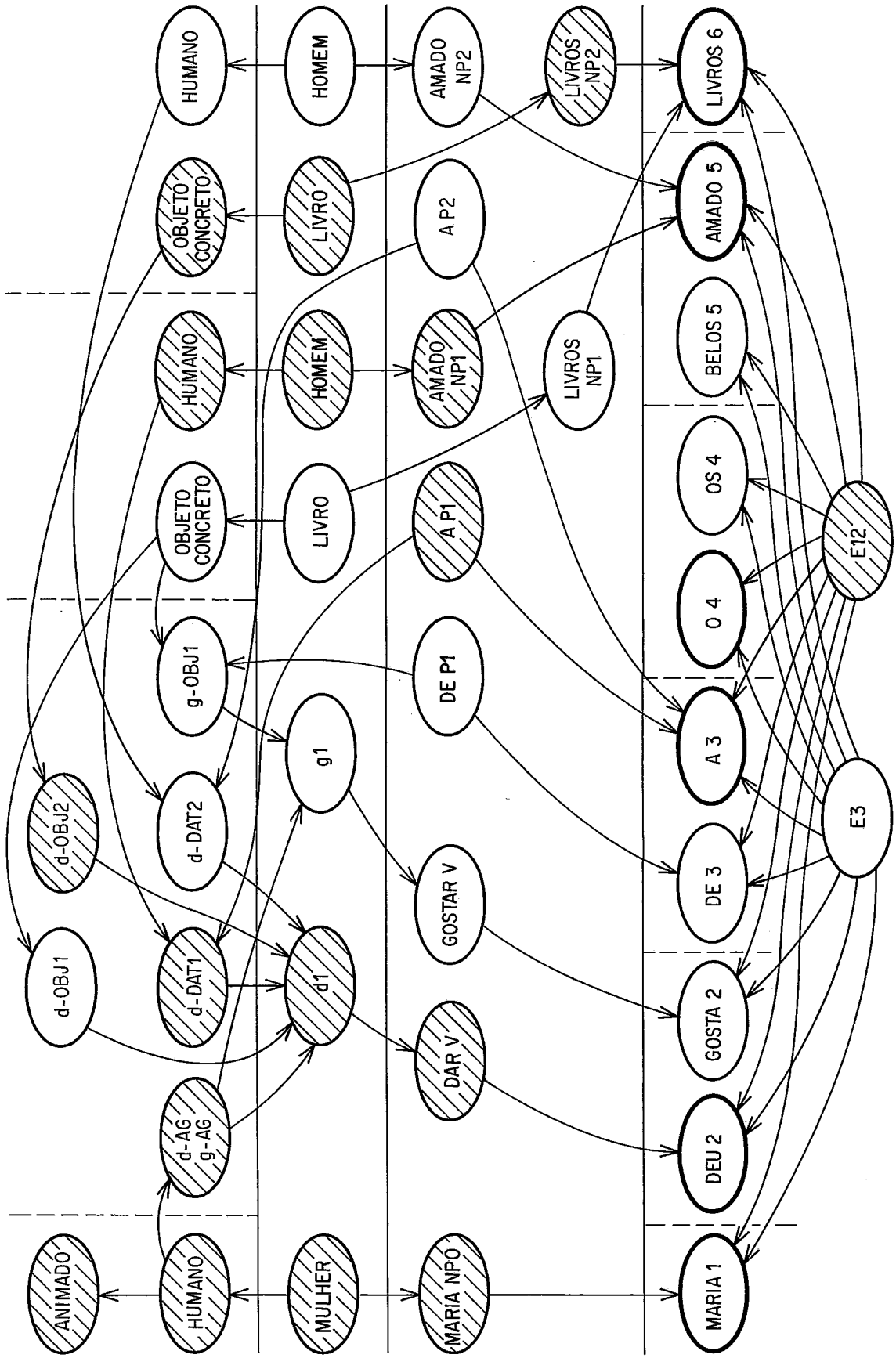

Fig. VI.6 – Exemplo do comportamento da rede que executa a semântica na solução de uma ambigüidade sintática da sentença "Maria deu ao amado livros"

freqüentemente que E12, espelhando a decisão da análise semântica. Por motivos semelhantes (o verbo "dar" é bitransitivo), a Figura VI.6 mostra a decisão semântica propagada ocasionando a escolha de E12 como diagnóstico sintático.

Exemplo 2:

Consideremos as frases:

"João bateu na porta",
"João bateu na Maria" e
"João bateu a porta"

O verbo "bater" tem sentidos distintos nas três frases. Na primeira usada no sentido de bat 1 (golpear para emitir som), na segunda no sentido de bat 2 (surrar alguém), e na última no sentido de bat 3 (fechar).

A Figura VI.7 mostra a rede referente aos três significados listados acima. Se são grampeadas com valor 1 as unidades referentes à primeira frase, a distinção entre os sentidos bat 1 e bat 2 dá–se unicamente pelas características semânticas do objeto. De fato, especificamos valores altos para as seguintes probabilidades condicionais:

P(bat 1–OBJ=1 | em P1 = 1, aberturas = 1)
P(bat 2–OBJ=1 | em P1 = 1, humano = 1)

A distinção entre os significados bat 1 e bat 3 verifica–se pela existência ou não da preposição "em". Então a única configuração para a qual a unidade "bat 3–OBJ" deve ocorrer com probabilidade alta é: "em P1" = 0 e "aberturas" = 1. Dessa forma, soluciona–se a competição entre "bat 1–OBJ" e "bat 3–OBJ".

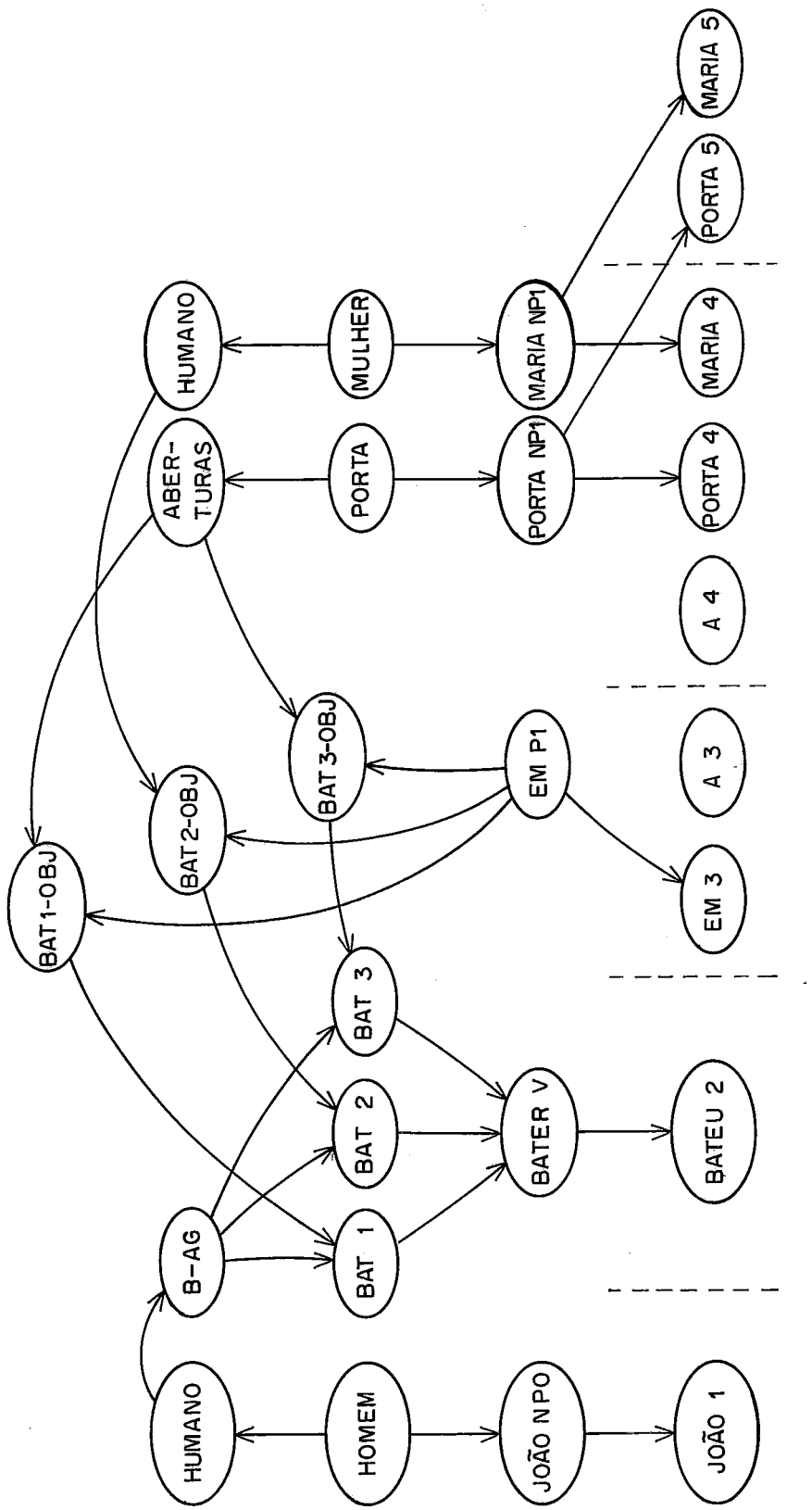

Fig. VI.7 – Fragmento da rede que executa a análise semântica representando os três significados do verbo bater.

CAPÍTULO VII

RESULTADOS EXPERIMENTAIS E CONCLUSÕES

Na formulação da proposta de tratamento da ambigüidade léxica apresentada neste trabalho partimos de alguns pressupostos. Em primeiro lugar, consideramos que a ambigüidade é uma característica central das linguagens naturais e permeia todo o processo de compreensão de uma sentença. Portanto, a solução da ambigüidade é indissociável do processo de compreensão e são necessárias informações de todas as fases do processo de análise da sentença para a obtenção de uma interpretação semântica.

Além disso, acreditamos que não se deve priorizar nenhuma fase do processo, pois a importância da informação de uma determinada fase (morfológica, sintática, semântica) depende diretamente da ambigüidade presente na sentença. Ou seja, em determinados casos a informação semântica é crucial e a sintática é irrelevante, em outros inverte–se a situação. Então, especificamos nossa rede de forma que o processo de compreensão resultasse da interação das informações sintáticas e semânticas.

Para implementar estas idéias foi necessário buscar um modelo computacional que permitisse o processamento paralelo e sem controle centralizado. Além disso, o modelo deveria ser adequado para implementar o processamento de regras (sintaxe) e os aspectos importantes da análise semântica ("semantic priming", tratamento de exceções, informações incompletas, etc.).

A escolha do modelo Bayesiano para abordar o problema da ambigüidade léxica mostrou–se bastante adequada para expressar as incertezas envolvidas na determinação dos significados das palavras. A especificação das relações causais mediante um grafo é bastante intuitiva e permite a determinação clara das relações entre as variáveis que representam os eventos. Além disso, o impacto das relações causais pode ser regulado diretamente mediante o estabelecimento de valores adequados para as probabilidades condicionais.

Acrescente–se às vantagens acima listadas, o fato de permitir uma implementação completamente paralela do problema e ainda a facilidade de regular o grau de precisão dos resultados mediante o número de iterações.

Estas observações são corroboradas pelos resultados da implementação descritos a seguir.

Os testes foram realizados utilizando o simulador seqüencial decrito no Apêndice. Inicialmente, foi utilizado o simulador sem "annealing"; no entanto, concluimos que o "annealing" acelera o processo de convergência da rede.

Sabemos, conforme mencionado no Capítulo III, que o fato de a distribuição ser estritamente positiva garante que uma configuração qualquer sempre pode ser alcançada durante a simulação. Isto é, dadas duas configurações quaisquer i, j, a probabilidade alcançar a configuração j a partir de i é sempre positiva. Porém, embora seja sempre positiva pode ser pequena. Então, se a rede parte de uma configuração inicial desfavorável (no sentido de ter uma probabilidade pequena de trocar de configuração), o processo de convergência pode ser retardado. Neste caso, a simulação com "annealing" parte de uma configuração inicial neutra, permitindo uma estabilização mais rápida da rede. Portanto, os testes sistemáticos foram realizados utilizando o simulador com "annealing".

Numa primeira fase, testamos a rede sintática separadamente; depois a rede sintática foi conectada à semântica e realizamos mais alguns testes.

A rede sintática foi construída para cinco grupos de entrada e utilizando a gramática da Figura V.1. A rede obtida tem 111 unidades, dentre as quais 28 correspondem a diagnósticos sintáticos.

Os valores das probabilidades condicionais associadas às unidades filha–e e filha–ou foram 0.995 e 0.005, referidos no Capítulo V como valores alto (próximo de 1) e baixo (próximo de 0), respectivamente. As únicas unidades da rede sintática que não têm pais são as evidências sintáticas (unidades de entrada sintática). Estas unidades estão sempre grampeadas e portanto não são importantes as suas probabilidades a priori.

Foram realizados testes com diversas entradas e a rede estabilizou–se, em geral, em torno de 5000 iterações. Utilizamos uma temperatura inicial de 1.000.000 e um fator de redução de 0.9.

Suponhamos que $S_1, \ldots, S_{28}$ representem as unidades de diagnóstico sintático e que S_i é o diagnóstico correto para uma dada entrada. Seja W_S o

conjunto de todas as unidades da rede exceto S. Após 5000 simulações obtivemos, em geral, $P(S_i = 1 | W_{S_i}) \cong 90\%$ e $P(S_j = 1 | W_{S_j}) \cong 3\%$ para $j \neq i$.

Nos casos de ambigüidade sintática, ou seja, quando existem dois ou mais diagnósticos sintáticos corretos associados à mesma entrada, obtivemos os diagnósticos com aproximadamente a mesma probabilidade de ocorrência. Então, se S_i e S_k são diagnósticos corretos associados a uma entrada ambígua, obtivemos $P(S_i = 1 | W_{S_i}) \cong P(S_k = 1 | W_{S_k}) \cong 90\%$ e $P(S_j = 1 | W_{S_j}) \cong 3\%$ para $j \neq i$ e $j \neq k$.

Numa segunda fase de testes a rede sintática foi conectada a um fragmento da rede que executa a análise semântica (significados de alguns substantivos e verbos) e a um fragmento da rede que representa o conhecimento e os casos verbais. Foram realizados inicialmente testes com sentenças não–ambíguas e a rede estabilizou na configuração correta com os mesmos parâmetros utilizados nos testes sintáticos (5000 iterações). As unidades referentes aos diversos significados dos substantivos não têm pais e portanto foi necessário estabelecer as suas probabilidades a priori. Atribuímos a mesma probabilidade a priori para todos os substantivos e verificamos mediante a experimentação que este valor deve ser baixo ($\cong 0.01$) para que a rede semântica funcione corretamente.

A seguir, testamos sentenças sintaticamente ambíguas nas quais a solução da ambigüidade dá–se na análise semântica. Nesse caso, obtivemos a convergência da rede após 10.000 iterações, mantidas a temperatura e o fator de redução. Obtivemos a configuração semântica correta (verbo e casos verbais) e o impacto dessa decisão refletiu–se nos diagnósticos sintáticos. Ou seja, o diagnóstico sintático correto manteve a sua probabilidade de ocorrência enquanto o outro (incorreto) caiu para 3%.

Realizamos, ainda, testes com sentenças completamente agramaticais, porém com significados relacionados, e obtivemos resultados bastante interessantes. Por exemplo, testamos a entrada semântica "vôo 1, avião 2, tempo 3" e a correspondente entrada sintática "subst 1, verbo 1, subst 2, subst 3". Obtivemos como diagnóstico sintático mais prováveis aqueles que têm o verbo na 1ª posição da frase. Como interpretação semântica obtivemos com maior probabilidade de ocorrência as unidades representando o significado do verbo "voar" referente a "veículo que voa pilotado por alguém" e o caso objeto do verbo voar ligado a um veículo.

Verificamos que a rede sintática sintetizou a única informação sintática coerente com a entrada: a única palavra que poderia ser verbo é "vôo". Além disso, o fato de o diagnóstico sintático não ser conclusivo não impediu que a rede semântica espelhasse a possível relação entre os significados.

Os resultados obtidos são bastante promissores, embora restem certos aspectos da questão que poderiam ser abordados dentro do modelo Bayesiano. Por exemplo, conforme sugerimos no Capítulo IV, poderiam ser acrescentados testes de concordância gramatical. Além disso, uma experiência interessante seria a de utilizar uma representação mais distribuída dos substantivos a experimentar valores diferenciados de probabilidades a priori, privilegiando significados mais usuais.

Como objeto de trabalho futuro, restam vários tópicos em aberto. Um problema pouco explorado neste trabalho e certamente de solução complexa é o da representação do conhecimento. No nosso trabalho representamos o conhecimento de forma classificatória sem utilizar o potencial do modelo Bayesiano para lidar com as exceções.

Outro tópico que não foi abordado é o da solução da ambigüidade estrutural. Nesta direção, acreditamos na possibilidade de estender o método de análise sintática anexando o tratamento da ambigüidade estrutural.

Finalmente, seria interessante incorporar à rede algum tipo de aprendizado. A questão do aprendizado em redes Bayesianas foi abordada por PEARL [1988], mas certamente está completamente em aberto em se tratando de uma aplicação concreta.

APÊNDICE I

ALGORITMOS PARA SIMULAÇÃO

Apresentamos neste apêndice os algoritmos utilizados para a simulação das redes Bayesianas. Descrevemos inicialmente o algoritmo que atualiza seqüencialmente as unidades da rede. A seguir, apresentamos as modificações feitas no algoritmo seqüencial para implementar o "annealing". Finalmente descrevemos o algoritmo para atualização das unidades em paralelo.

O Algoritmo A.1, descrito a seguir, é conseqüência do Teorema III.2 e está estruturado em dois passos principais. No passo 2 executam–se n iterações. Uma iteração consiste na atualização dos valores de todas as unidades da rede.

No passo 3, calcula–se a freqüência de ocorrência de cada unidade X da rede, dividindo–se o número de vezes que o evento $X = 1$ ocorreu pelo número de iterações n.

O processo de atualização de cada unidade X está dividido em duas partes. Inicialmente calcula–se a probabilidade P do evento $X = 1$ dados os valores dos pais de X, dos filhos de X, e dos pais dos filhos de X. Na segunda parte escolhe–se aleatoriamente um valor x para X favorecendo 1 com probabilidade p. Se $X = 1$ incrementa–se de um a variável que conta o número de vezes que ocorreu o evento $X = 1$.

Algoritmo A.1 (Simulação Seqüencial)

Entrada: Uma rede Bayesiana $D = (W, \vec{E})$, um conjunto de evidências $EV \subset W$, e o número de iterações n.

Saída: A freqüência de ocorrência de cada unidade $X \in W$, após n iterações.

Método:

1) Para cada variável $X \in W - EV$ faça $v_X := 0$;

2) Para i,..., n faça:

2.1) Para cada variável $X \in W - EV$ faça:

2.1.1) $c := P(X = 1 \mid \text{valores dos pais de } X)$;

2.1.2) $d^1 := \prod_{\substack{Y \in W \\ Y \text{ filho de } X}} P(Y=y \mid X=1, \text{ valores dos outros pais de } Y)$;;

2.1.3) $d^0 := \prod_{\substack{Y \in W \\ Y \text{ filho de } X}} P(Y=y \mid X=0, \text{ valores dos outros pais de } Y)$;

2.1.4) $a := c\, d^1$;

2.1.5) $b := (1-c)\, d^0$;

2.1.6) $p := \frac{a}{a+b}$;

2.1.7) Atribua a X o valor 1 com probabilidade p.

2.1.8) Se $X = 1$ então $v_X := v_X + 1$.

3) Para cada variável $X \in W - EV$ faça $p_X := \frac{v_X}{n}$

4) Pare, p_X é a freqüência de ocorrência da unidade $X \in W$ após n iterações.

A alteração do Algoritmo A.1 para introduzir o "annealing" no processo de simulação tem como objetivo acelerar a convergência da rede. Esta aceleração é obtida quando partimos de uma distribuição inicial "neutra", isto é, de uma distribuição próxima daquela em que todas as unidades são igualmente prováveis. Então, usamos a seguinte fórmula para o cálculo da probabilidade p do passo 2.1.6 do Algoritmo A.1:

$$p = \frac{1}{1 + e^{\ell n(b/a)/(1+T)}}$$

onde T é uma parâmetro que regula a distância da distribuição inicial da distribuição desejada.

Analisando a fórmula utilizada para o cálculo de p vemos que:

Se $T \to \infty$, a distribuição aproxima–se da distribuição uniforme, pois $P \cong \frac{1}{1+e^0} = \frac{1}{2}$.

Se $T \to 0$, a distribuição aproxima–se da distribuição desejada pois $P \cong \frac{1}{1+e^{\ell n(b/a)}} = \frac{1}{1 + \frac{b}{a}} = \frac{a}{a+b}$.

A simulação com "annealing" consiste em iniciar a simulação com uma temperatura T alta e estipular um fator de redução α a ser aplicado sobre T em cada ciclo de atualizações. Ou seja, deve–se acrescentar como entrada no Algoritmo A.1 a temperatura inicial T_0 e o fator de redução α. Ao final de cada ciclo onde são atualizadas todas as unidades da rede calcula–se uma nova temperatura T_i para ser utilizada no ciclo i de atualizações. Temos então, $T_i = \alpha T_{i-1}$, $0 < \alpha < 1$ e a temperatura final é dada por $T_f = \alpha^n T_0$.

A atualização das unidades de uma rede Bayesiana no processo de simulação estocástica pode ser feita em paralelo, desde que se observe a restrição de que vizinhos (no sentido do Corolário III.2) não podem ser atualizados concorrentemente. No processo de simulação estocástica, para atualização de cada variável X, a vizinhança V_X que deve ser consultada é composta dos pais de X, dos filhos de X, e dos pais de filhos de X. Então, as variáveis de V_X não podem ser atualizadas concorrentemente com X.

Para gerenciar o processo de atualização utilizamos uma estratégia mediante a qual unidades vizinhas (no grafo G' descrito abaixo) nunca são atualizadas concorrentemente. Então, é necessário acrescentar à rede Bayesiana arestas fictícias entre unidades que têm filhos em comum. Estas unidades serão doravante denominadas "mates".

Dada uma rede Bayesiana $D = (W, \vec{E})$, seja $G = (W, E)$ o grafo subjacente não–dirigido. Seja $G' = (W, E')$ o grafo não–dirigido obtido acrescentando ao grafo G arestas entre "mates".

Consideremos uma orientação acíclica qualquer em G' (esta orientação não

tem nenhuma relação com a orientação da rede Bayesiana). Nesta orientação distinguimos as unidades que só têm arestas convergentes (arestas que chegam na unidade), chamadas doravante "sinks". De acordo com o método de "edge–reversal" [BARBOSA e GAFNI, 1989], a cada instante podem ser atualizadas conjuntamente todas as unidades que são "sinks". De fato, duas unidades que são "sinks" não podem ser vizinhas, pois nesse caso uma delas não seria "sink".

Inicialmente são atualizadas todas as unidades "sinks" de acordo com a orientação acíclica inicial de G'. Após atualizarmos seus valores, estas unidades revertem todas as suas arestas. Desta forma obtém–se uma nova orientação acíclica para G' e um novo conjunto de unidades "sink". Repete–se sucessivamente o processo resultando em novas orientações acíclicas e novos conjuntos de "sinks". Assim, as unidades são atualizadas atendendo à restrição de que vizinhos não são atualizados simultaneamente. Além disso, mostra–se que todas as unidades são atualizadas com freqüência suficiente e que não ocorre "deadlock", já que toda orientação acíclica tem pelo menos um "sink" e o tempo necessário para uma unidade tornar–se um "sink" é igual ao tamanho da maior trajetória orientada da unidade a um "sink".

Apresentamos, a seguir, o algoritmo para a atualização de uma unidade da rede. Este algoritmo é uma adaptação para redes Bayesianas do algoritmo apresentado em BARBOSA e LIMA [1990] para a atualização de redes neuronais.

Existe um processo "mestre" que gerencia as informações globais da rede (topologia da rede, unidades que são evidências, configuração inicial da rede) e contabiliza a freqüência de ocorrência de cada unidade da rede. Para cada unidade X_i, chamamos de vizinhos "upstream" aqueles que têm arestas convergentes em X_i e chamamos de vizinhos "downstream" aqueles que têm arestas partindo de X_i. Cada unidade recebe do mestre a lista de seus vizinhos com as seguintes informações:

1) tipo de vizinho (pai, filho, ou "mate" em relação à orientação da rede Bayesiana); e

2) classificação ("upstream" ou "downstream" em relação à orientação acíclica).

Os passos 1 a 8 do Algoritmo A.2 apresentado abaixo correspondem à inicialização do processamento, e o passo 9 corresponde às atualizações da unidade X_i. O passo 10 é necessário apenas para encerrar o processamento.

Algoritmo A.2 (Algoritmo para X_i)

1) Receba lista de vizinhos (tipo, classificação) do mestre e informação se X_i é evidência ou não;

2) Receba x_i do mestre (valor inicial);

3) Receba matriz de probabilidades condicionais do mestre;

4) Marque os vizinhos "upstream";

5) Receba x_j de todo o pai X_j;

6) Envie x_i a todo filho;

7) Se $x_i = 1$ envie $p_i^1(j) = P(X_i = 1 \,|\, \text{valores dos pais com } X_j = 1)$ e $p_i^0(j) = P(X_i = 1 \,|\, \text{valores dos pais com } X_j = 0)$ para todo pai X_j não marcado.

   Caso contrário, envie $1 - p_i^1(j)$ e $1 - p_i^0(j)$ para todo pai X_j não–marcado.

8) Receba $p_j^1(i)$ e $p_j^0(i)$ de todo filho X_j marcado.

9) Para $k = 1,\ldots,K$ faça:

   Receba de todos os vizinhos "downstream" X_j:
   um sinal se X_j for "mate";
   x_j se X_j for pai;
   $p_j^1(i)$ e $p_j^0(i)$ se X_j for filho;

   Calcule

   $$a := P(X_i{=}1 \,|\, \text{valores dos pais}) \prod_{\text{filhos } X_j} p_j^1(i);$$

$$b := [1 - P(X_i = 1 \mid \text{valores dos pais})] \prod_{\text{filhos } X_j} p_j^0(i);$$

$$p := \frac{a}{a+b}$$

Se X_i não é evidência

então $X_i := 1$ com probabilidade p

senão $X_i := 0$.

Se $X_i = 1$

então envie $p_i^1(j)$ e $p_i^0(j)$ para todo pai X_j.

senão envie $1 - p_i^1(j)$ e $1 - p_i^0(j)$ para todo pai X_j.

Envie x_i a todo filho X_j e ao mestre.

Envie sinal aos "mates".

10) Receba mensagem de todo vizinho marcado, ignorando–a.

APÊNDICE II

COMPLEXIDADE DOS ALGORITMOS APRESENTADOS

Analisamos neste apêndice a complexidade dos algoritmos apresentados, principalmente a do Algoritmo V.1. Este algoritmo apresenta a construção de uma rede Bayesiana para executar a análise sintática de uma sentença, dada uma gramática livre de contexto $G = (N, \Sigma, P, S)$ e um inteiro $\ell \in \mathbb{N}$.

A complexidade do Algoritmo V.1 é analisada a partir do número de não–terminais $|N|$, do número de terminais $|\Sigma|$, do número de produções $|P|$, do tamanho da maior produção n e do inteiro ℓ.

A complexidade do Algoritmo V.1 é dominada pela complexidade do passo 3, que é executado para todo $m = 1, ..., \ell$. Para cada m, o passo 3.1 é executado $|P|$ vezes. Para cada $p \in P$, o passo 3.1.2 é executado $|J_n^m| = \begin{bmatrix} m-1 \\ n-1 \end{bmatrix}$ vezes. O passo 3.1.2 envolve operações sobre conjuntos de tamanho máximo $|P| + |N|$ quando $m > 1$ e $|P| + |\Sigma|$ quando $m = 1$. O binômio $\begin{bmatrix} m-1 \\ n-1 \end{bmatrix}$ é majorado por $\begin{bmatrix} m-1 \\ h(m) \end{bmatrix}$ onde $h(m) = \left\lfloor \frac{m-1}{2} \right\rfloor$. Então, a complexidade do Algoritmo V.1 é dada pela seguinte função:

$$\mathrm{Compl}(|N|, |\Sigma|, |P|, n, \ell) = |P| \begin{bmatrix} 0 \\ 0 \end{bmatrix} \log(|P| + |\Sigma|) +$$

$$+ \sum_{m=2}^{\ell} |P| \begin{bmatrix} m-1 \\ h(m) \end{bmatrix} (n+2\ell-2m+2) \log(|P| + |N|) =$$

$$= |P| \{(n+2\ell) \log(|P| + |\Sigma|) +$$

$$+ \log(|P| + |N|) \left[(n+2\ell+2) \sum_{m=2}^{\ell} \begin{bmatrix} m-1 \\ h(m) \end{bmatrix} - 2 \sum_{m=2}^{\ell} m \begin{bmatrix} m-1 \\ h(m) \end{bmatrix}\right]\}.$$

Segundo GRAHAM, KNUTH e PATASHNIK [1989] não é, em geral, conhecida uma forma mais simples de expressar o binômio $\begin{bmatrix} m-1 \\ h(m) \end{bmatrix}$, portanto não foi possível reescrever os somatórios em função de ℓ.

A complexidade do algoritmo possui componentes que são funções exponenciais da entrada. No entanto, isto não inviabiliza a utilização do algoritmo

dado que este será utilizado uma única vez na geração da rede sintática e mesmo assim com valores relativamente modestos para seus parâmetros de entrada.

Calculamos a complexidade do algoritmo para a gramática da Figura V.1 após a eliminação das produções ϵ e para 12 grupos de entrada. Então, para $|N| = 4$, $|\Sigma| = 8$, $|P| = 27$, $n = 6$ e $\ell = 12$, obtivemos $Compl(|N|, |\Sigma|, n, \ell) = O(10^6)$. Ou seja, é necessário realizar $O(10^6)$ operações significativas para a construção da rede.

A gramática da Figura V.1 apresenta valores típicos do problema, pois o número de terminais é dado pelas categorias gramaticais existentes, o número de não–terminais e o número de produções também não devem ser substancialmente alterados para acrescentar outras estruturas às reconhecidas pela gramática. Além disso, embora o tamanho da sentença possa ser qualquer, é limitado por um valor que representa o número máximo de palavras que pode ter uma frase compreensível da linguagem.

Os demais algoritmos (V.2, V.3 e VI.1) são polinomiais no tamanho da entrada.

REFERÊNCIAS BIBLIOGRÁFICAS

ADRIENS, G. e SMALL, S. L. (1988), "Word Expert Parsing Revisited", in Small, S. L., Cottrell, G. W., e Tanenhaus, M. K. (eds.), Lexical Ambiguity Resolution, San Mater, CA: Morgan Kaufmann Pub., Inc.

AHO, A. V. e ULLMAN, J. D. (1972), The Theory of Parsing, Translation and Compiling, Volume 1: Parsing, Englewood Cliffs, N. J.: Prentice–Hall, Inc.

ALLEN, J. (1987), Natural Language Understanding, Menlo Park, CA: The Benjamin Cummings Publishing Company, Inc.

BARBOSA, V. C. e GAFNI, E. (1989), "Concurrency in Heavily Loaded Neighborhood–Constrained Systems", ACM Transactions on Programming Languages and Systems, Vol. 11, no. 4, pp. 562–584.

BARBOSA, V. C. e LIMA, P. M. V., (1990), "On the Distributed Parallel Simulation of Hopfield's Neural Networks", Software Practice & Experience (aceito para publicação).

CHARNIAK, E. (1983a), "A Parser with Something for Everyone", in King, M. (ed.), Parsing Natural Language, London: Academic Press.

CHARNIAK, E. (1983b), "Passing Markers: A Theory of Contextual Influence in Language Comprehension", Cognitive Science, Vol. 7, no. 3, pp. 171–190.

CHARNIAK, E., GAVIN, M. K. e HENDLER, J. A. (1983), "The FRAIL/NASL Reference Manual", TR CS–83–06, Department of Computer Science, Brown University.

CHOMSKY, N. (1965), Aspects of the Theory of Syntax, Cambridge: MIT Press.

CHOMSKY, N. (1970), "Deep Structure, Surface Structure, and Semantic Interpretation", in Jakobovits, L. and Steinberg, D. (eds.), Interdisciplinary Reader in Philosophy, Linguistics, Anthropology and Psychology, Cambridge: Cambridge University Press.

COOPER, G. F. (1987), "Probabilistic Inference Using Belief Networks is NP–Hard", T.R. KSL–87–27, Medical Computer Science Group, Stanford

University, Stanford, CA.

COTTRELL, G. W. (1985), A Connectionist Approach to Word Sense Disambignation. Ph.D. Diss., Computer Science Department, University of Rochester, Rochester, N.Y.

CRYSTAL, D. (1988), Dicionário de Lingüística e Fonética, (tradução e adaptação Dias, M. C. P.), Rio de Janeiro: Jorge Zahar Editor.

FELDMAN, J. A. e BALLARD, D. (1982), "Connectionist Models and their Properties", Cognitive Science, Vol. 6, pp. 205–254.

FELDMAN, J.A., FANTY, M. A., e GODDARD, N. H. (1988), "Computing with Structured Neural Networks", Computer, IEEE.

FELLER, W. (1950). An Introduction to Probability Theory and Applications, Vol. 1 (3rd ed. 1968), New York: Wiley.

FERNANDES, M. A. (1989), Análise Sintática Baseada em Regras Usando um Modelo Conexionista: Uma Visão Crítica, Tese de mestrado, Engenharia de Sistemas e Computação, COPPE/UFRJ, Rio de Janeiro.

FILLMORE, C. J. (1968), "The Case for Case", in Bach, E. and Harmis, R. T. (eds), Universals in Linguistic Theory, Chicago: Holt, Rinehart and Winston.

GRAHAM, R. L., KNUTH, D. E. e PATASHNIK, O. (1989), Concrete Mathematics – A Foundation for Computer Science, Reading, MA: Addison Wesley.

HAYES, P. J. (1977), Some Association–Based Techniques for Lexical Disambiguation by Machine, TR 25, Computer Science Department, the University of Rochester, Rochester, N. Y.

HIRST, G. (1987), "Semantic Interpretation and Resolution of Ambiguity", (Studies in Natural Language Processing), Cambridge: Cambridge University Press.

HIRST, G. (1988), "Resolving Lexical Ambiguity Computationally with Spreading Activation and Polaroid Words", in Small, S. L., Cottrell, G. W., e

Tanenhaus, M. K. (eds.), Lexical Ambiguity Resolution, San Mateo, CA: Morgan Kaufmann Pub., Inc.

HOLANDA, A. B. DE (1986), Novo Dicionário Aurélio da Língua Portuguesa, Rio de Janeiro: Editora Nova Fronteira S. A.

LYTINEN, S. L. (1988), "Are Vague Words Ambiguous ?", in Small, S. L., Cottrell, G. W., e Tanenhous, M. K. (eds.), Lexical Ambiguity Resolution, San Mateo, CA: Morgan Kaufmann Pub., Inc.

MARCUS, M. (1980), A Theory of Syntactic Recognition for Natural Language, Cambridge: MIT Press.

McCLELLAND, J. L. e KAWAMOTO, A. H. (1986), "Mechanisms of Sentence Processing: Assigning Roles to Constituents", in McClelland, J. L., Runmelhart, D. E. and the PDP Research Group, Parallel Distributed Processing, Cambridge: MIT Press.

MILNE, R. (1988), "Lexical Ambiguity Resolution in a Deterministic Parser", in Small, S. L., Cottrell, G. W., e Tanenhaus, M. K. (eds.), Lexical Ambiguity Resolution, San Mateo, CA: Morgan Kaufmann Pub., Inc.

PEARL, J. (1988), Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, San Mateo, CA: Morgan Kaufmann Pub., Inc.

QUILLIAN, M. R. (1968), "Semantic Memory", in Minsky, M. L. (ed.), Semantic Information Processing. Cambridge: MIT Press.

SAMLOWSKI, W. (1978), "Case Grammar", in Charniak, E. and Wilks, Y. (eds.), Computational Semantics, North–Holland Pub. Co.

SELMAN, B. (1985), Rule–Based Processing in a Connectionist System for Natural Language Understanding, T.R. CSRJ–168, Department of Computer Science, University of Toronto, Toronto, Canadá.

SMALL, S. L. (1980), Word Expert Parsing: A Theory of Distributed Word–Based Natural Language Understanding, Ph.D. Diss., Computer Science Department, University of Maryland, College Park, Maryland.

WALTZ, D. L. e POLLACK, J. B. (1985), "Massively Parallel Parsing: A Strongly Interactive Model of Natural Language Interpretation", Cognitive Science, Vol. 9, no. 1, pp. 51–73.

WINOGRAD, T. (1972), Understanding Natural Language. New York: Academic Press.

WOODS, W. A. (1975), "What's in a Link ? Foundations for Semantic Networks", in Bobrow, D. and Collins, A. (eds.), Representation and Understanding, New York: Academic Press.