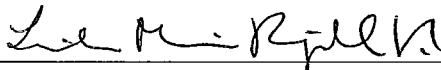


UMA APLICAÇÃO EM REDES SEMÂNTICAS
UTILIZANDO REDES BAYESIANAS

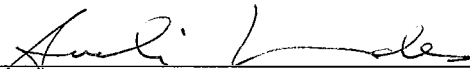
Eduardo de Paiva Alves

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 MESTRE EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

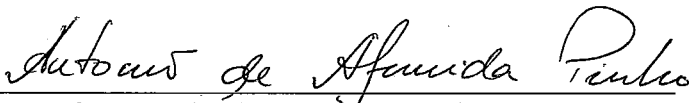
Aprovada por:



Prof^a. Leila Maria Ripoll Eizirik, D. Sc.
(Presidente)



Prof^a. Sueli Bandeira Teixeira Mendes, Ph. D.



Prof. Antônio de Almeida Pinho, D. Sc.

RIO DE JANEIRO, RJ - BRASIL
Agosto de 1993

ALVES, EDUARDO DE PAIVA

Uma Aplicação em Redes Semânticas utilizando Redes Bayesianas [Rio de Janeiro] 1993

vii, 56 p. 29,7 cm (COPPE/UFRJ, M. Sc., Engenharia de Sistemas e Computação, 1993)

Tese - Universidade Federal do Rio de Janeiro, COPPE.

1. Ambigüidade Léxica, Linguagem Natural, Redes Bayesianas

Aos meus pais
Aos meus irmãos
Ao meu Anjo

Aos meus pais
Aos meus irmãos
Ao meu Anjo

RESUMO DA TESE APRESENTADA À COPPE/UFRJ COMO PARTE DOS REQUISITOS NECESSÁRIO À OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS (M. Sc.)

**UMA APLICAÇÃO EM REDES SEMÂNTICAS
UTILIZANDO REDES BAYESIANAS**

Eduardo de Paiva Alves
agosto 1993

ORIENTADORA: Leila Maria Ripoll Eizirik

Programa: Engenharia de Sistemas e Computação

O trabalho apresenta uma aplicação de Redes Bayesianas ao problema da ambigüidade léxica. A rede é composta de uma sub-rede semântica e outra sintática. A parte sintática foi construída a partir de uma gramática livre de contexto. A sub-rede semântica consiste de uma hierarquia taxonômica contruída a partir de descrições estruturadas e um analisador semântico baseado em uma gramática de casos e estruturação de conceitos.

A interpretação semântica é resultado da interação das duas redes que funcionam paralelamente, ou seja, a interpretação é resultado da interação das duas partes.

Os testes foram feito utilizando simulação estocástica com annealing. A rede se mostrou adequada à solução da ambigüidade léxica, bem como possibilitou a interação entre informações semânticas e sintáticas.

ABSTRACT OF THESIS PRESENTED TO COPPE/UFRJ AS PARTIAL
FULFLLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER IN SCIENCES (M. Sc.)

AN APPLICATION OF SEMANTIC NETWORKS
USING BAYESIAN NETWORKS

Eduardo de Paiva Alves

August, 1993

Thesis Supervisor: Leila Maria Ripoll Eizirik

Department: Systems Engineering and Computer Science

This work proposes a Bayesian Network approach to Lexical Ambiguity Resolution. The network consists of two parts - the syntactic and the semantic sub-networks. The syntactic part is built from a context-free grammar. The semantic sub-network is made up of a taxonomic hierarchy build up of structured concepts and a semantic parser based on a case grammar using structured descriptions.

The sentence analysis is the result of interaction between the two sub networks and emerges from the global interaction of them.

Tests have been carried out using a Stochastic Simulation Algorithm using annealing. The net proved adequate to lexical ambiguity resolution using information from different level in a concurrent approach.

ÍNDICE

	pág.
INTRODUÇÃO	1
CAPÍTULO I - REVISÃO BIBLIOGRÁFICA	5
CAPÍTULO II - DESCRIÇÃO DO SISTEMA	18
CAPÍTULO III - REDES BAYESIANAS	24
CAPÍTULO IV - IMPLEMENTAÇÃO	31
CAPÍTULO V - RESULTADOS, CONCLUSÕES E EXTENSÕES	44
APÊNDICE A - MÉTODO DE SIMULAÇÃO	49
BIBLIOGRAFIA	55

INTRODUÇÃO

A ambigüidade é um dos problemas centrais para os sistemas de compreensão da linguagem, encontrando-se em todos os níveis. Uma frase é ambígua quando existe mais de uma interpretação possível para a mesma. Podemos classificar a ambigüidade em duas categorias: léxica e estrutural. A primeira se dá quando uma ou mais palavras possuem diferentes significados. A última, quando existe mais de uma maneira de agrupar os componentes da frase.

As ambigüidades podem também ser classificadas de acordo como tipo de informação necessária para solucioná-las. Podem ser informações sintáticas quando é possível determinar uma interpretação única para a frase a partir de regras e categorias gramaticais possíveis de cada palavra. Por exemplo:

João fixa a peça.

O item léxico *fixa* pode ser um adjetivo ou um verbo, contudo a partir das regras pode-se definir a interpretação correta para esta frase.

Em alguns casos, é possível resolver a ambigüidade utilizando-se informações semânticas locais, tais como o tipo de complemento exigido por um verbo. Por exemplo:

Vestiu o pijama com bolinhas.

Vestiu o pijama com pressa.

Embora as duas frases possuam a mesma estrutura gramatical, na primeira temos um sintagma verbal e um nominal, e na segunda temos um sintagma verbal e dois nominais.

Podem também ser necessárias informações do contexto no qual a frase está inserida para determinar entre várias interpretações qual a mais adequada. Na frase:

A manga está suja.

Não é possível determinar o significado correto da palavra *manga* sem levar em conta o contexto do enunciado. Por outro lado, é possível que a ambigüidade esteja na estrutura semântica quando não se pode definir o papel correspondente a um dado sintagma. No exemplo seguinte:

Pensou no Rio de Janeiro.

Apesar de *Rio de Janeiro* não ser ambíguo, seu papel na frase não fica definido pois o mesmo pode se referir tanto ao local da ação quanto ao objeto da ação.

Por último, há casos em que só se pode determinar a interpretação de uma palavra ou frase em função de dados extra linguísticos, como crenças e intenções. Em uma conversa telefônica, a pergunta:

Seu marido está?

Só pode ser interpretada conhecendo-se a intenção do falante pois pode corresponder tanto a uma simples pergunta como a um pedido para falar com o marido da pessoa com quem se fala.

Assim o processamento de linguagem natural requer informação distribuída em diferentes níveis - sintático, semântico, contextual e conhecimento do mundo. O nível sintático inclui regras e categorias gramaticais, o nível semântico, preferências semânticas que indicam a natureza dos participantes de uma determinada frase, bem como as possíveis acepções das palavras. Já o nível contextual corresponde às informações sobre uso das palavras nos diferentes contextos. Por fim o conhecimento do mundo inclui informações genéricas que os falantes devem ter, além das específicas a respeito do interlocutor, como crenças e objetivos.

Portanto para interpretar uma sentença em linguagem natural, precisamos além de um mecanismo de aplicação de regras uma base de conhecimento .Além disso, quando interpretamos uma frase, cada item isoladamente armazena apenas parte da informação necessária à compreensão da referida frase. Cada informação é portanto incompleta e pode conflitar com a informação armazenada em outra parte da estrutura. Assim é necessária uma estrutura que represente adequadamente informações incompletas e conflitantes e permita a aplicação de regras.

O modelo escolhido para especificação do sistema foi uma rede Bayesiana por possibilitar tanto a implementação de regras quanto a inserção de dados contextuais graças a sua estrutura em rede. Além disso, são adequadas ao tratamento de exceções e informações incompletas e conflitantes devido à sua natureza probabilística.

O sistema descrito na presente tese é capaz de analisar sintaticamente, selecionando dentre as possíveis categorias gramaticais de uma palavra a correta, bem como o padrão sintático correspondente à entrada. Além disso, identifica papéis semânticos bem como a acepção adequada correspondente a cada palavra de frases ambíguas selecionadas.

No capítulo I faremos uma curta revisão bibliográfica descrevendo alguns sistemas para compreensão de linguagem natural, ressaltando os trabalhos que serviram de base a presente tese como as redes de herança estruturada (Brachman [1985]), o sistema desenvolvido por Woods [1991] e as *estruturas conceituais* de Schank [1976] .

No capítulo II descrevemos o sistema conceitualmente, isto é, através de sua estrutura e independente da implementação, definindo as diretrizes básicas da especificação. A construção e o funcionamento de cada uma das partes constituintes do sistema são tratados. Isto inclui a especificação das unidades básicas e como estas formam estruturas mais complexas.

No capítulo III é feita uma breve exposição a respeito da escolha do modelo, bem como da teoria das redes Bayesianas e da construção e funcionamento do simulador.

No capítulo IV descrevemos a implementação. Justifica-se a escolha do léxico e a partir deste, obtemos as primitivas necessárias à sua definição. Enunciamos as regras construtivas da rede e demonstramos passo a passo a especificação da rede a partir das informações sintáticas e semânticas disponíveis.

No capítulo V apresentamos e analisamos os resultados experimentais, extraímos conclusões e propomos ampliações.

No apêndice A descrevemos o simulador estocástico com annealing e apresentamos o algoritmo, um resultado na forma de tabela e uma figura correspondente à sub-rede semântica.

CAPÍTULO 1

REVISÃO BIBLIOGRÁFICA

No presente capítulo descreveremos os sistemas de compreensão de linguagem natural que utilizam representações gráficas, genericamente conhecidas como redes semânticas. Esta revisão se restringirá aos trabalhos que contribuíram para o desenvolvimento das redes de herança estruturada.

A primeira tentativa de formalização das redes semânticas surgiu com o trabalho de Woods "What's in a Link" [1975]. Segundo Woods, as redes até então desenvolvidas utilizando estruturas de nós e arcos rotulados (grafos) não especificavam o que representavam os nós e os arcos nem as regras construtivas da rede.

Em seu trabalho Woods aborda os diferentes problemas representacionais das redes e introduz então a distinção entre arcos assertivos e definicionais. Os primeiros afirmam algo a respeito dos nós que relacionam. Os últimos, servem para relacionar partes de uma proposição ou descrição. Por exemplo:

telefone \Rightarrow preto

pode servir tanto para afirmar que todos telefones são pretos, como para descrever um telefone específico.

Em seu trabalho, Woods condena a busca de uma forma canônica para representação, pois segundo ele, mesmo que ela existisse, não seria adequada aos propósitos da representação do conhecimento pois os conceitos teriam que ser decompostos nestas formas todas as vezes que desejássemos acessar informações a respeito do mesmo.

O primeiro passo em direção às redes com capacidade expressiva superior à do cálculo de primeira ordem foi dado no trabalho de Hayes "On Semantic Nets, Frames and Associations" [1977]. A partir da noção de

frame, que são estruturas de dados que armazenam informações relativas ao reconhecimento de uma situação específica, Hayes idealizou a construção de nós complexos, formados a partir de outros nós, e chamados *depictions*. Além disso tem sua interpretação dependente do "ponto de vista", isto é, da maneira como são ativados. Deste modo, a descrição de um membro inferior de um animal quando ativada através da descrição de um indivíduo, é associada com a perna deste indivíduo. Além da relação parte-todo, a rede era também organizada em uma hierarquia taxonômica (sub/super conjunto). Tais características combinadas possibilitavam a exploração de duas das características mais importantes das redes semânticas: a herança e estruturação de conceitos.

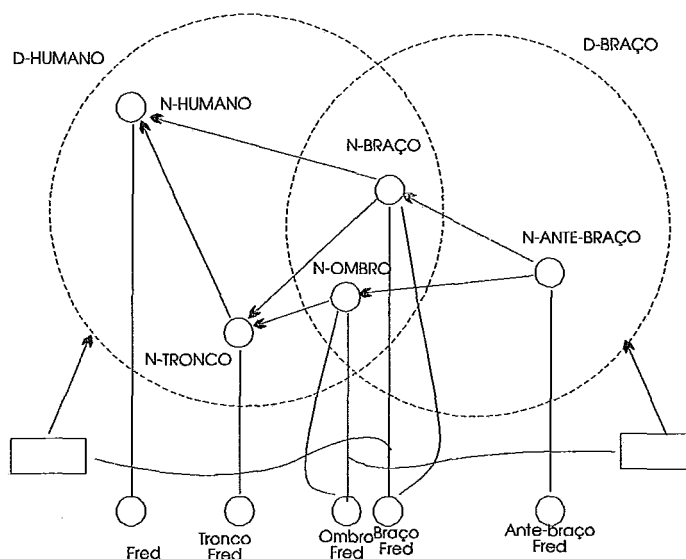


Fig. 1. Exemplo do uso de Depictions e Binders (Hayes (1977))

Na figura 1, os círculos correspondem a *depictions* para humano e braço. Os nós que aparecem na parte de baixo da figura são instâncias dos nós genéricos que aparecem na *depcition*. A herança das relações é feita através dos *binders* que são os retângulos.

No trabalho "Encoding Knowledge in Partitioned Networks" (Hendrix [1979]), utilizando o conceito de *frame*, Hendrix introduziu a noção de *partitioning* que é uma operação que permite agrupar nós em unidades chamadas *spaces*. *Vistas* são um conjunto de *spaces* utilizados para definir contextos. *Delineations* são similarmente *spaces* utilizados para representar o

mecanismo de quantificação em redes semânticas. Na figura abaixo representa a frase *Toda cidade tem um chefe de carrocinha que foi mordido por todos cães da cidade*. Os arcos rotulados *form* apontam para áreas que correspondem ao escopo de uma variável universalmente quantificada que é indicada pelo arco \forall . O nó *p* está implicitamente existencialmente quantificado e corresponde a um chefe da carrocinha específico para uma dada cidade.

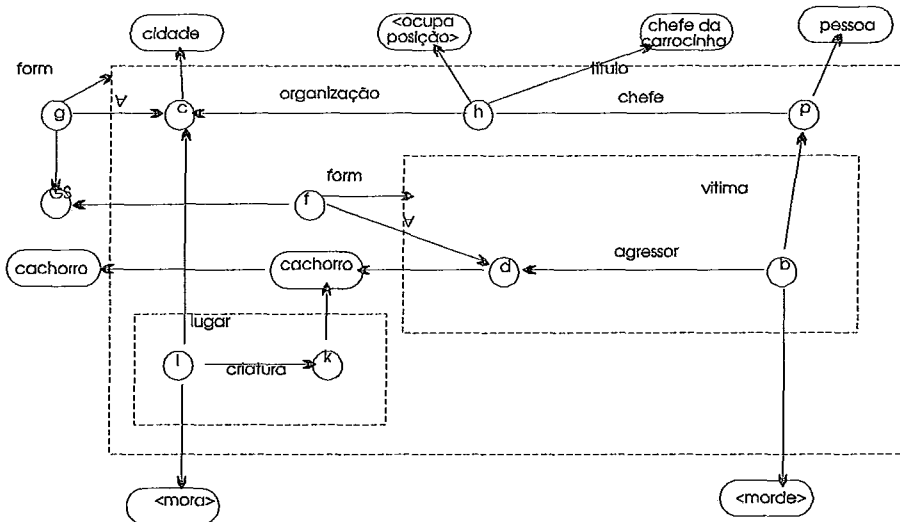
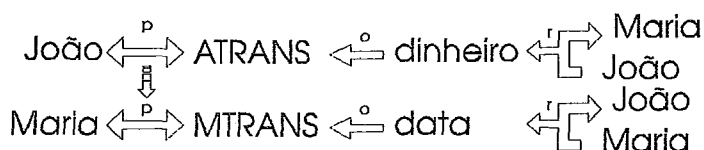


Fig. 2. Exemplo de partitions (Hendrix (1979))

Em seu artigo "Identification of Conceptualizations Underlying Natural Language" [1973], Schank propõe um formalismo onde as frases são analisados em estruturas conceituais, construídas a partir de conceitos e relações entre estes conceitos, relações estas chamadas de dependências. Os conceitos podem ser Nomes, Ações ou Modificadores. As Ações correspondem a um número reduzido de primitivas verbais.

Schank define um conjunto de regras conceituais que delimitam as dependências conceituais possíveis. Os casos conceituais são assim definidos como aqueles termos dependentes requeridos pela ação a que correspondem, e são eles: objetivo, receptivo, diretivo e instrumental, servindo para combinar aspectos do mesmo evento. Os casos diretivo e receptivo são complementares nunca aparecendo na mesma *estrutura conceitual*. A cada primitiva corresponde um conjunto definido de casos conceituais.

Existem ainda as relações conceituais que são aquelas entre *estruturas conceituais*, isto é, eventos distintos, que podem ser de causalidade, tempo e localização. Há ainda as flexões conceituais (conceptual tenses), que podem modificar as *estruturas conceituais* e relações conceituais. Por exemplo:



É a *estrutura conceitual* correspondente à frase *João deu dinheiro à Maria para que ela lhe fornecesse a data*. As setas correspondem às dependências conceituais Agente, Objeto e Recipiente (setas duplas, marcadas com *o* e com *r* respectivamente). A seta interrompida a relação conceitual, no caso de causalidade. Por fim, os *p* sobre as setas correspondem à flexão conceitual, no caso a passado. Os demais correspondem aos conceitos nominais (Nomes).

Uma *estrutura conceitual* não está completa até que todos os casos conceituais tenham sido explicitados. Ao contrário dos casos verbais de Fillmore [1968], os casos conceituais são partes da estrutura conceitual da qual dependem e deste modo estão sempre presentes. As mudanças de estado não são consideradas Ações. Para Woods toda Ação requer outra Ação como um caso instrumental e assim sucessivamente sendo representado apenas o que foi explicitamente mencionado como instrumento. Por exemplo, na frase: *João comeu sorvete*, existe um caso instrumento relacionando a Ação ingerir com a Ação levar à boca e esta com outra de mover o braço e assim sucessivamente.

Em seu artigo "On the Epistemological Status of Semantic Networks" [1979], Brachman propõe um formalismo no qual os conceitos podem ser estruturados e partes destas estruturas podem ser herdadas. As redes para a representação do conhecimento desenvolvidas até então são classificadas em níveis de acordo com os elementos representacionais primitivos (primitivas) utilizados nas definições dos conceitos. Esta classificação permite identificar

como a escolha das primitivas corresponde a diferentes abordagem do problema.

O nível implementacional corresponde a utilização da rede simplesmente como uma estrutura de dados, sendo os arcos ponteiros e os nós endereços, e não havendo nenhuma tentativa de utilizar a estrutura da rede em si para representar conhecimento. Tal abordagem corresponderia a utilização de uma estrutura em rede apenas para armazenamento de dados.

No nível lógico os arcos correspondem a relações lógicas e os nós a predicados. A estrutura da rede era utilizada para decompor o conhecimento bem como para expressar a estrutura do conhecimento. Nesse nível enquadram-se os trabalhos de Hendrix [1979] e Schubert [1976].

No nível conceitual os nós correspondem a conceitos e os arcos a relações semânticas ou conceituais. As redes neste nível se caracterizam pelo número reduzido de elementos conceituais e relações primitivas conceituais. Neste nível está incluído o trabalho de Schank [1973].

No nível linguístico os conceitos e as relações expressas pelos nós e arcos são específicos da linguagem representada. Existe um formalismo na literatura - OWL (Szolowitz et al. [1977]) que representa esta abordagem.

A estruturação de conceitos presente nos trabalhos de Hayes [1975] e Hendrix [1979] sugere a organização do conhecimento em unidades mais estruturadas do que nós e arcos isolados ou predicados e proposições. O uso de descrições intensionais possibilita a representação de conceitos que embora distintos se referem a mesma extensão não podendo portanto ser identificados com uma classe, como é o caso de *estrela da manhã e estrela da tarde* que possuem como extensão o planeta Vênus. Esta característica indica a existência de uma classe de relações que não é abrangida por nenhum dos quatros níveis acima. Estas relações são relações entre as partes de intensões e as intensões e entre as próprias intensões. Descrições intensionais podem ser relacionadas diretamente em virtude de sua estrutura ou indiretamente em virtude de suas extensões.

Outro traço comum às redes semânticas, a herança de propriedades, não é uma primitiva lógica e também não é considerada um relação semântica. Isto indica a existência de um nível entre o lógico e o conceitual. Desse modo, o nível epistemológico é definido como sendo aquele cujas relações são usadas para estruturar conceitos, como os casos, e para relaciona-los como a herança de sub-descrições entre descrições.

Brachman desenvolve então uma classe de linguagens de rede denominadas Redes de Herança Estruturada (SI-nets) cujos os arcos são primitivas epistemológicas, sendo KL-ONE um exemplo desta linguagem.

Os elementos básicos de KL-ONE são os Conceitos que representam objetos intensionais e nunca diretamente objetos extensionais. Conceitos podem ser genéricos ou individuais. Os conceitos genéricos representam classes de objetos descritos pelas características de um elemento típico da classe. Os conceitos individuais representam objetos individuais obtidos através da individuação de conceitos genéricos. Na figura abaixo vemos as relações entre o conceito genérico arco, o conceito individual Arco do Triunfo e o objeto extensional Arco do Triunfo, que são instanciação, denotação e individuação conforme a figura.

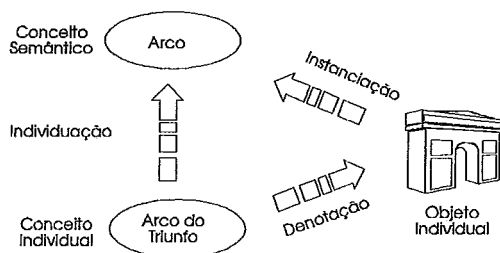


Fig. 3. Relações em KL-ONE

Como o objetivo deste formalismo é possibilitar descrições intensionais e herança de partes destas descrições, um objeto não pode ser representado adequadamente apenas como uma entidade atômica ou uma lista de propriedades. O Conceito tem que representar a estrutura interna do objeto, bem como o objeto como entidade. Essa característica é obtida através de Descrições Papel/Participantes e Descrição de Estruturas.

As Descrições Papel/Participantes (Papéis) representam os vários tipos de atributos, partes e etc. que os conceitos possuem, consistindo da entidade que desempenha o papel e a função por ela desempenhada.

As Descrições Estruturadas (SD) armazenam a informação de como os Papéis interagem entre si. Cada SD é um conjunto de relações entre dois ou mais Papéis.

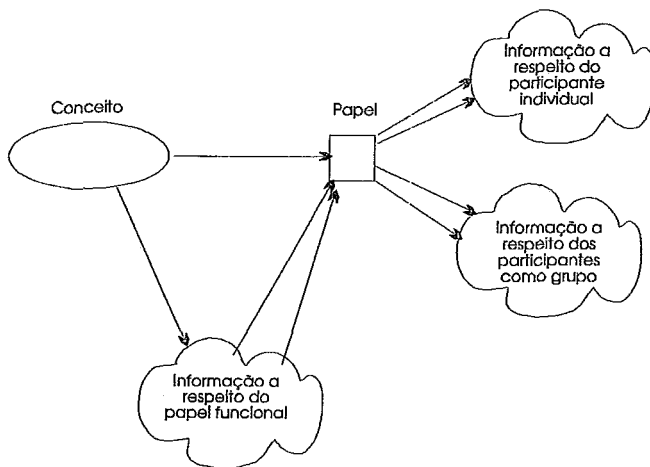


Fig. 4. Relações entre Papéis, Conceitos e Estruturas Conceituais

As relações epistemológicas podem ser entre Conceitos, Papéis, Descrições Estruturadas, internamente nas Descrições de Papéis e nas Descrições Estruturadas e entre Descrições de Papéis e Descrições Estruturadas. É importante ressaltar que a relação entre o Conceito e Descrição de Papéis não é a mesma que entre objeto representado pelo conceito e a coisa que corresponde ao papel para aquele dado objeto. Portanto não existem arcos rotulados *cor* que seria substituído por um Papel *PI* e uma relação Restrição de Valor para o Conceito *cor*.

Outra relação epistemológica se dá entre Conceitos-Conceitos, Papéis-Papéis e SD's-SD's, e é responsável pela herança. Brachman vê os arcos de herança como cabos que conduzem Papéis e SD's para o herdeiro, que podem não ser herdados intactos mas que podem ser modificados para dar ao sub-Conceito uma definição mais restrita que a do Conceito original. Esta modificação pode ser de satisfação, diferenciação e restrição no caso de Papel e preempção no caso de SD.

Em seu artigo "Understanding Subsumption and Taxonomy" Woods [1991] fornece ferramentas para descrições conceituais estruturadas que permitem uma classificação automática em uma hierarquia taxonômica. A classificação é possível pois sendo os conceitos estruturados, a sua localização na rede pode ser deduzida destas estruturas a partir de um mecanismo de herança formalmente especificado. Além da classificação, o sistema possibilita inferências em função do mecanismo.

Woods ressalta ainda a importância de se tratar os Conceitos como intensões e não como predicados de primeira ordem ou classes. Uma diferença é que enquanto os predicados são verdadeiros ou falsos as descrições são satisfeitas o que é mais adequado quando se trata de herança. Isto resolve também problemas com nomes próprios pois um objeto pode satisfazer uma descrição em uma situação. E podemos ainda usar conceitos para falar de entidades abstratas.

No sistema de Woods as descrições podem ser atômicas ou compostas. As atômicas possuem apenas um nome. Já as compostas são da forma:

$$c_1 \dots c_k / (r_1 : v_1) \dots (r_k : v_k) : (p_1 \dots p_t)$$

Onde c são descrições conceituais primárias, $r:v$ modificadores relacionais (pares relação : valor) e p condições gerais. Cada conceito primário pode ser pensado como um arco "kind-of" do conceito sendo definido para o conceito primário correspondente, cada par $r:v$ pode ser pensado como um arco relacional r do conceito sendo definido para o conceito v , e cada condição geral p como um arco satisfaz do conceito sendo definido para a descrição correspondente ao predicado que deve satisfazer.

Assim, a categoria *bebida* é definida como

$$\textit{subst liq} / (\textit{converse}(\textit{inger}) : \textit{animal})$$

Deve-se ressaltar aqui que a razão pela qual estas condições gerais foram incluídas é a necessidade de descrições complexas cujos argumentos

são como *slots* não podendo ser expressas por conceitos primários ou relações como por exemplo:

Menino cuja quarto a mãe arruma.

descreve uma entidade que possui determinadas relações com outras entidades cuja definição depende desta, isto é, no exemplo, *mãe* e *quarto* estão definidos em função da entidade que se quer definir.

As relações podem ser arbitrárias ao contrário do formalismo definido por Brachman. Outra diferença entre os dois sistemas é que no trabalho de Woods não existem relações de restrição, diferenciação, modificação, número etc, que são substituídas por operadores formadores de relação.

Woods volta a abordar a questão dos arcos assertivos e definicionais, ressaltando a não exclusividade dos dois. O tratamento quantitativo para os arcos assertivos é feito através de operadores formadores de relações que Woods preferiu incluir como "tags" destas relações. Outra característica marcante é o fato de serem as relações tratadas do mesmo modo que os conceitos.

Outros operadores como *converse*, *chain* e *restrict* são introduzidos. O primeiro corresponde a uma forma de voz passiva. O segundo a um encadeamento e o último a uma contra definição. Já os arcos estruturais tiveram seus próprios operadores definidos, incluindo outros como o de negação e o de restrição numérica.

A existência de definições alternativas, que também não eram possíveis no formalismo de Brachman, possibilita a criação de Abstrações, que podem ser pensadas como classes de definições. No sistema em questão existe um operador disjuntivo que toma uma coleção de descrições e produz uma nova que é satisfeita por qualquer uma delas. Por exemplo, podemos definir triângulo como uma abstração de duas definições alternativas: figura plana com três lados ou figura plana com três ângulos.

Podemos relacionar a noção de condições necessárias e suficientes com a abordagem analisada da seguinte forma: A conjunção de descrições alternativas de uma dada abstração corresponde a condição necessária, e cada uma das descrições sendo suficiente. Já numa definição disjuntiva, a condição necessária é uma condição que todas elas tenham em comum, e a suficiente é cada uma delas separadamente. Por outro lado nas definições parciais existe um vazio entre as satisfeitas pelas suficientes e o complemento das necessárias, isto é, podem existir entidades que não sejam nem incluídas pelas suficientes nem excluídas pelas necessárias. No sistema em questão existem operadores definicionais capazes de tratar condições necessárias e suficientes ou bicondicionais.

A partir de conceitos básicos é possível definir novos conceitos e relações que possam ser usados do mesmo modo que os primitivos. É importante ressaltar que definições e descrições compostas não são sinônimos dado que é possível que uma definição seja uma abstração de várias descrições ou mesmo um conceito parcialmente definido. Existem operadores definicionais que tomam um conjunto de condições e produzem uma definição.

De acordo com Woods, o sistema assim definido tem a capacidade de expressão de conjunções e disjunções em qualquer profundidade e quantificação envolvendo argumentos de uma relação binária. Os predicados n-ários podem ser definidos a partir de relações binárias e podem ser pensadas como papéis.

$$R/(r_1:x_1), \dots, (r_n:x_n)$$

Compreendida a estruturação dos conceitos, é possível então analisar a generalização. Informalmente, uma descrição composta generaliza outra por uma das seguintes razões ou uma combinação delas:

1 . Uma categoria primária em uma descrição é mais geral que uma categoria primária na outra

pessoa cujos filhos são médicos generaliza

mulher cujos filhos são médicos

2 . Um modificador relacional em uma descrição generaliza um modificador relacional na outra

pessoa cujos filhos são formados generaliza
pessoa cujos filhos são médicos

3 . Uma condição geral em uma descrição generaliza uma condição geral na outra

criança cujos pais limpa o quarto generaliza
criança cuja mãe limpa o quarto

4 . A descrição mais específica inclui uma categoria, relação ou condição não presente na descrição mais geral

pessoa cujos filhos são médicos generaliza
pessoa cujos filhos são médicos e joga golfe

É importante ressaltar aqui a respeito do item 2 que generalizar um elemento de um modificador relacional pode generalizar ou especializar a descrição resultante dependendo do tag quantificacional . Exemplo:

aluno que assiste a todos ao curso de matemática generaliza
aluno que assiste a todos os cursos

embora *curso de matemática* não generalize *todos os cursos*. Uma análise detalhada aparece em Woods [1991]

Apesar da ideia de generalização ser bastante simples a sua aplicação não tem sido transparente. É possível diferenciar pelo menos cinco significados que não tem sido percebidos nas discussões do tópico.

1. Generalização extensional, quando as extensões do conceito generalizador incluem as extensões do conceito generalizado.

2. Generalização estrutural onde a generalização é determinada por critérios generalizadores formalmente especificados.
3. Generalização gravada onde o conceito mais geral é explicitamente gravado como generalizador do menos geral na estrutura taxonômica armazenada
4. Generalização axiomática onde inclui-se como axioma da base de conhecimento o fato do mais geral generalizar o mais específico.
5. Generalização deduzida onde deduz-se a partir da base de conhecimento através de operações de inferência, que o mais geral generaliza o mais específico.

Enquanto o significado 3 é resultado da organização taxonômica, a acepção 2 é utilizada para testar e classificar novas descrições nesta estrutura. O primeiro significado serve apenas para determinar a completude e consistência do significado 2. O penúltimo significado fornece as bases sobre as quais está baseada a generalização estrutural. O último decorre de relações que não podem ser deduzidas pelo critério de generalização estruturada.

Não é necessário como em KRIPTON (Brachman [1983]), definir uma fronteira entre generalização terminológica usando Descrições Conceituais e generalização não-terminológica usando Dedução, mesmo porque como vimos, nem sempre as definições terminológicas são completas e bicondicionais como era de se esperar.

Tomando por princípio a noção de generalização intensional é possível definir o critério básico da generalização estrutural da seguinte forma:

$$c_1 \dots c_k / m_1, \dots, m_n$$

generaliza

$$c_1' \dots c_k' / m_1' \dots m_k'$$

se cada c_i generaliza algum c_j'

e cada m_i generaliza algum m_j'

onde $m_i = r_i : v_i$ generaliza $m_i' = r_i' : v_i'$ de acordo com regras de generalização de modificadores que como dizemos não obedecem a regra geral e podem ser vistas na bibliografia.

Embora esta regra não cubra todos os casos de generalização, parece corresponder às inferências feitas com facilidade por humanos, sendo utilizada no sistema como critério de generalização intensional. Relações que podem ser deduzidas logicamente são incluídas diretamente na rede como generalização gravadas tratando-as como axiomas.

Para tratar adequadamente conceitos disjuntivos, abstratos e parciais é necessário tratar diferenciadamente o generalizador e o generalizado com relação a definição de generalização. Isto é, para o generalizador basta olharmos separadamente as condições suficientes enquanto que para o generalizado é necessário analisarmos o conjunto das condições necessárias.

O sistema que descreveremos a partir do próximo capítulo, é uma aplicação dos trabalhos de Woods [1990] e Brachman [1985] apresentados acima, utilizando os casos conceituais de Schank [1973]

CAPÍTULO II

DESCRIÇÃO DO SISTEMA

O objetivo deste capítulo é descrever conceitualmente o sistema, isto é, independente da implementação que é assunto do capítulo seguinte. A partir do problema a ser resolvido, propomos uma solução em rede, consistindo de uma sub-rede sintática e outra semântica. A maneira como as diferentes partes do sistema se interligam e como é feita a entrada também são assunto deste capítulo.

O sistema descrito na presente tese é capaz de analisar sintaticamente e identificar papéis semânticos bem como a acepção adequada correspondente a cada palavra da entrada. A análise sintática é feita utilizando-se uma gramática formal e permite identificar os termos da frase. A análise semântica é feita a partir do léxico, identificando-se as possíveis acepções de cada palavra, classificando-as em uma hierarquia taxonômica e buscando-se aquelas que permitam um diagnóstico semântico. Os dois processos se interligam fornecendo informações um ao outro, isto é, uma informação sintática pode ajudar a retirar uma ambigüidade semântica e vice-versa.

A entrada no sistema consiste na identificação da posição onde ocorrem as palavras presentes na frase em questão bem como de suas possíveis categorias gramaticais. Assim na frase:

João fixa a peça.

a entrada consiste dos itens *João*₁, *fixa*₂, *a*₃, *peça*₄, *subst*₁, *verbo*₂, *adj*₂, *art*₃, *prep*₃, *verbo*₄, *subst*₄ que correspondem à identificação dos itens léxicos e das categorias com a posição de ocorrência das mesmas. A maneira como escolhemos fazer esta entrada bem como todos os demais detalhes de implementação são assunto do capítulo seguinte.

Para a análise sintática especificamos uma rede a partir de uma gramática livre de contexto. Nesta rede as diferentes categorias gramaticais possíveis referentes a cada item léxico da entrada são agrupadas

sucessivamente, de acordo com as regras da gramática escolhida em diagnósticos que correspondem às possíveis análises sintáticas da frase de entrada. Estes diagnósticos estão interligados com a parte semântica da rede possibilitando a troca de informações entre as duas partes. Os arcos portanto correspondem a aplicações das regras e os nós a diagnósticos.

A partir dos diagnósticos sintáticos possíveis para a frase de entrada identificam-se os termos participantes, isto é, os termos nominais e verbais. É importante ressaltar que a partir da entrada léxica são identificados os diagnósticos possíveis desde o início da simulação, que relacionam os itens léxicos com os termos da frase como *joão*np₁ por exemplo, caracterizando o paralelismo entre as diferentes partes da rede, sendo os diagnósticos inadequados inibidos durante a simulação.

Assim, na figura abaixo temos o trecho da rede sintática relevante à interpretação da frase *João fixa a peça*. A ilustração mostra como as categorias gramaticais são herdadas da entrada léxica, e então agrupadas sucessivamente até formar um diagnóstico adequado à frase. É importante ressaltar também que, como mencionamos anteriormente, as unidades que relacionam os termos léxicos com os grupos nominais estão ligadas a estes e portanto possibilitam um paralelismo aos processos sintático e semântico.

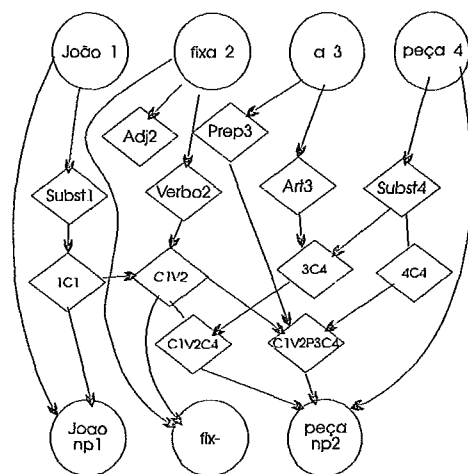


Fig. 5. Trecho da sub-rede sintática

A sub-rede semântica é capaz de classificar os termos nominais e, com base nessa classificação, selecionar a acepção do verbo que se ajuste às características dos demais participantes da oração. Para isto possui duas partes distintas: uma primeira que chamaremos de classificador, e outra que chamaremos de analisador.

O classificador consiste em uma estrutura taxonômica onde os termos nominais da frase são identificados com categorias que estão hierarquizadas de acordo com um critério de generalização adequado. Nesta hierarquia estão representadas apenas as categorias relevantes a especificação dos verbos definidos na rede. Para isso, a rede possui uma representação dos participantes requeridos por cada acepção verbal.

O classificador foi construído a partir das palavras selecionadas utilizando uma versão simplificada do sistema proposto por Woods [1991]. As descrições são da forma:

$$c : c_1, \dots, c_n / (r_1 : v_1) \dots (r_k : v_k)$$

Assim uma descrição c generaliza outra se:

1. Uma categoria primária é mais geral que uma categoria primária na outra descrição;
2. Um modificador relacional numa descrição generaliza um modificador relacional na outra descrição.
3. A descrição mais específica inclui uma categoria ou relação não presente na descrição mais geral.

As relações aparecem na frase na forma de verbos e são representadas com auxílio de Descrições de Estrutura como veremos a seguir.

O critério formal de generalização estrutural proposto por Woods [1990] não é usado no sistema pois a rede trabalha com um léxico definido à

priori não sendo necessárias inclusões através de um classificador automático na base de conhecimento ou inferências.

Além das descrições compostas, é possível incluir Abstrações utilizando-se um operador disjuntivo que toma um conjunto de descrições e gera uma nova que pode ser satisfeita por qualquer delas. Uma rede Bayesiana expressa perfeitamente a idéia de abstração através das dependências probabilísticas. Assim se uma categoria envolvida em uma abstração ocorre, provoca, em função da forma como construímos a rede Bayesiana, a ocorrência da categoria definida independente da ocorrência das demais categorias envolvidas do mesmo modo que nas descrições compostas. Caso as características sejam compatíveis sua ocorrência poderá ser simultânea no maior generalizador comum, caso contrário teremos inibição lateral.

Os diagnósticos semânticos tinham que depender probabilisticamente dos termos da frase para que pudéssemos selecionar as combinações possíveis entre os mesmos. Devido à esta escolha de orientação, necessária dada a natureza indiferenciada dos nós, as Abstrações resultam no mesmo tipo de generalização que as definições estruturadas, pois todos seus elementos podem ser herdados indistintamente, isto é, não importa se as categorias são necessárias ou suficientes, pois todas podem ser herdadas.

Assim, se houvesse na rede o conceito *triângulo* definido como uma Abstração, poder-se-ia herdar tanto *polígono de três lados* quanto *polígono com três ângulos*, do mesmo modo que para *bebida* pode-se herdar *substância líquida e ingerível*.

Além da generalização estrutural, há ainda casos de generalização axiomática quando incluimos no sistema o fato de uma categoria incluir outra independente da estrutura das descrições correspondentes. Assim, *vivo* generaliza *animal* que generaliza *humano*. A hierarquia resultante consiste das generalizações gravadas que corresponde a soma das generalizações estruturais e axiomáticas. Olhando os conceitos individualmente como faremos mais tarde, veremos como as definições permitiram classificar os conceitos na hierarquia taxonômica.

A orientação das setas do mais específico para o mais geral (generalização) caracteriza a herança (os conceitos mais específicos *satisfazem* os mais genéricos). Os arcos correspondem portanto à generalização gravada e caracterizam a herança.

A segunda parte da rede semântica consiste de um analisador semântico. Utilizamos as *Estruturas Conceituais* de Schank [1975] para identificar as dependências existentes nas frases. Estas estruturas foram representadas com auxílio das Descrições Estruturadas (SD's) e os Casos Conceituais correspondentes através das Descrições de Papel/Participantes (Papéis) semelhantes àqueles das Redes de Herança Estruturada (SI-Nets).

Enquanto as Descrições Estruturadas armazenam a maneira como os termos se relacionam, isto é, o caso a que cada um corresponde bem como a aceção correta do verbo e correspondem aos nossos diagnósticos semânticos, as Descrições de Papel/Participante indicam através dos arcos de dependência, qual o termo nominal que corresponde ao Caso Conceitual e qual o atributo requerido pelo mesmo representado pelo nó que o arco aponta.

Por exemplo, no caso dos diagnósticos semânticos, a ocorrência do nó *ag atrans obj* está vinculada à ocorrência de um *humano* no primeiro grupo nominal, dos verbos *levar* ou *tomar* e de *propriedade* no segundo grupo nominal.

Tentamos o máximo possível nos aproximar das noções de Schank [1973] de *Caso Conceitual*(agente, paciente/recipientes ou instrumento) e *Ação*, sendo que não representamos mudança de estado e só entidades animadas podem se agentes. Os verbos escolhidos são atômicos, isto é, consideramos alguns verbos expressando estas primitivas (*Ações*). O caso instrumental que corresponde a uma dependência entre duas *estruturas conceituais* aparece na nossa rede como um *Caso Conceitual* por motivo de simplificação. Pelo mesmo motivo omitimos os *Casos Conceituais* que, embora requeridos pela *Ação*, não foram mencionados explicitamente.

A rede em questão se enquadra perfeitamente no nível *epistemológico* da classificação de Brachman [1985], por dois motivos: primeiro os conceitos são estruturados e as relações se dão entre partes das descrições dos conceitos, e segundo que a partes destas descrições podem ser herdadas.

No caso de herança múltipla, tal como ocorre em situações de ambigüidade , podemos resolver o conflito no *maior generalizador comum*, que é o conceito mais específico que pode ser herdado de ambos. Atribuimos então probabilidades de acordo com o comportamento esperado. As exceções que também podem aparecer devido à esta orientação, podem do mesmo modo ser tratado no conceito onde surge o conflito usando negação, isto é, atribuindo uma probabilidade baixa para a ocorrência deste, dada a da exceção. Exemplos deste comportamento podem ser vistos nos resultados.

No capítulo seguinte faremos uma breve apresentação da teoria e construção das redes Bayesianas para facilitar a compreensão do capítulo referente à implementação.

CAPÍTULO III

REDES BAYESIANAS

No presente capítulo descreveremos o modelo Bayesiano, ressaltando suas principais características, bem como a construção da rede Bayesiana a partir do modelo probabilístico correspondente às relações causa/efeito. Uma descrição detalhada das redes Bayesianas pode ser encontrada em Pearl [1988]. O simulador aparece no apêndice A.

A análise de uma frase em um sistema dirigido prioritariamente pela sintaxe pode ser feita eficientemente nos modelos seqüenciais já que estes estão baseados em regras que definem uma gramática e existem algoritmos para buscar a análise adequada a uma frase dada uma gramática. Por outro lado nos modelos seqüenciais a utilização do contexto na análise semântica e a interligação dos processos sintático e semântico é dificultada pois um paralelismo é desejável.

Os modelos paralelos por sua vez, são adequados ao tratamento de informações conflitantes, pois possibilitam que sejam consideradas simultaneamente todas informações pertinentes a uma determinada decisão. pois é possível que uma decisão inadequada só seja detectada muitos passos adiante quando a informação conflitante é considerada.

Nos modelos simbolistas a representação de informações incompletas é extremamente complexa. O modelo conexionista possibilita tratar informações distribuídas de maneira adequada e explorar o paralelismo facilmente sendo no entanto, a implementação de regras de grande complexidade.

Dentre os modelos paralelos, a abordagem em rede é extremamente indicada para a representação do conhecimento pois associa ao paralelismo do processamento, informações espaciais que correspondem as interrelações entre as informações disponíveis que uma implementação puramente paralela não possui necessariamente.

As redes neuronais quando utilizadas para aplicações de mais alto nível, tais como processamento de linguagem natural, carecem de um procedimento de atribuição de pesos e dos limites de excitação das unidades adequados ao funcionamento e nem sempre tem a convergência para os resultados globais desejados garantida (ou seja, que representem o que efetivamente se deseja). O modelo escolhido para especificação do sistema foram as redes Bayesianas pois possibilitam tanto a implementação de regras quanto a inserção de dados contextuais graças a sua estrutura em rede. Além disso, permitem o tratamento de exceções e evidências devido à sua natureza probabilística.

Redes Bayesianas são grafos dirigidos e acíclicos onde os nós representam unidades e os arcos representam as dependências diretas entre essas unidades. A probabilidade de ocorrência de um determinado nó é dada mediante as probabilidades condicionais associadas a ocorrência de seus pais (origem dos arcos dirigidos a um nó)

Os modelos probabilísticos se caracterizam pela possibilidade de regular o impacto das relações causa/efeito através dos valores das probabilidades associadas às ocorrências das variáveis possibilitando o controle da aplicação das regras.

Além disso, determinados modelos probabilísticos possibilitam "abductive reasoning": se A implica B , então o fato de B ser verdadeiro torna A mais provável. Como consequência comportam-se em conformidade com a regra "explaining away": se A implica B e C implica B , e B e C são verdadeiros A se torna menos provável. O primeiro possibilita o tratamento das evidências enquanto o segundo o tratamento de exceções.

Os modelos fundamentados teoricamente além disso, possibilitam particularizações e a identificação das causas de possíveis falhas e realização dos ajustes adequados ao comportamento desejado.

Para expressar as dependências condicionais presentes no problema através de um grafo 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 1

Seja $U = \{\alpha, \beta, \dots\}$ 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 do dado Z se e só se $P(y/z) > 0 \Rightarrow P(x/y,z) = P(x/z)$ para quaisquer configurações X, Y, Z , respectivamente.

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

Consideramos o critério de separabilidade usual para grafos não dirigidos, isto é, dado um grafo G e três subconjuntos de nós do grafos X, Y, Z dizemos que Z separa X de Y (e usamos a notação $\langle X/Z/Y \rangle_G$) 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 $\langle X/Z/Y \rangle_G$ implica $I(X,Z,Y)_P$ e vice versa para quaisquer $X, Z, Y \subseteq U$.

Dentre os modelos probabilísticos buscamos aqueles capazes de representar esta dupla implicação, cujo critério de separação seja capaz de representar dependências e independências. No nosso caso, desejamos que o modelo expresse as dependências induzidas que podem ser exemplificadas do seguinte modo. Considerando o exemplo da fig. 5, vemos que uma dada entidade pode ser física ou abstrata, temos que a ocorrência da unidade *entidade* depende da ocorrência de uma das duas. No entanto, a ocorrência

da unidade *entidade* cria uma dependência entre as outras duas unidades já que estas não podem ocorrer simultaneamente.

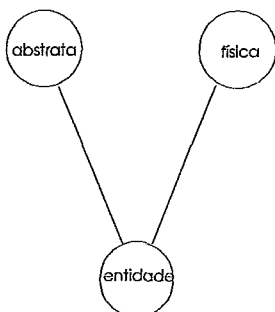


Fig. 5. Grafo não dirigido

Com os grafos não-dirigidos não é possível expressar estas dependências induzidas, portanto é necessária a construção de uma representação gráfica menos restritiva em relação às dependências do modelo probabilístico associado.

Definição 2

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

$$\langle X/Z/Y \rangle_G \iff I(X,Z,Y)_P$$

Pela definição vemos que, se um grafo G é um mapa-I de P então dois vértices separados em G correspondem a variáveis separadas em P ; mas não assegura que vértices ligados em G correspondam, obrigatoriamente, a variáveis dependentes em P .

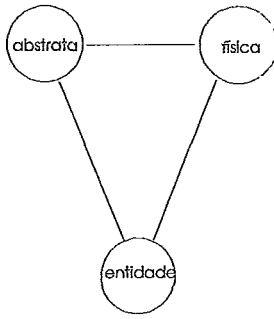


Fig. 6. Grafo não dirigido completo

No entanto os grafos completos como o da figura 6 não expressam adequadamente as independências presentes no modelo pois aparece uma dependência não presente no problema original. Para representar estas dependências induzidas e não-transitivas, utilizam-se grafos dirigidos acíclicos com um novo critério de separação.

Definição 3

Seja $D = (V, \overset{r}{E})$ um grafo dirigido e acíclico. Sejam X, Y, Z subconjuntos disjuntos de V . Dizemos que Z d-separa X de Y (e usamos a notação $\langle X|Z|Y \rangle_p$) 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 completamente contida em Z .

Definição 4

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

$$\langle X/Z/Y \rangle_G \rightsquigarrow I(X,Z,Y)_P$$

Um DAG é um mapa-I minimal de P se a retirada de quaisquer de suas arestas faz com que não seja mais um mapa-I de P.

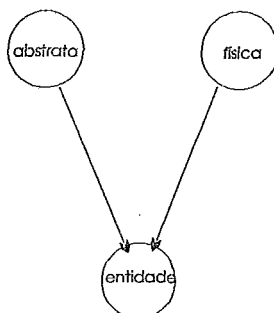


Fig. 7. Grafo dirigido

Definição 5

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

A construção da rede Bayesiana a partir de uma distribuição P é feita determinando-se para uma dada ordem das variáveis U, o menor conjunto que torna cada variável independente das demais. Esta construção é feita de modo que para cada variável X, o conjunto de seus pais Π_X d-separa X de todos seus não descendentes.

Existem resultados na teoria Pearl [1988] que enunciam as condições que deve satisfazer um DAG $D = (V, \vec{E})$ para que este seja uma rede Bayesiana de uma dada distribuição P sobre um conjunto de variáveis U.

Do mesmo modo existem condições sobre o ordenamento das variáveis de U que se cumpridas na construção de um DAG $D = (V, \vec{E})$ resultam na obtenção de uma distribuição P da qual D é uma rede Bayesiana

Na prática 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 obtemos uma distribuição P e um DAG que por construção é uma rede Bayesiana de P.

A garantia de que obtivemos uma rede Bayesiana é que, na construção seguimos uma ordem δ do conjunto de vértices tal que

$$\delta = \{x_1, \dots, x_n\}$$

$$P(x_i | x_1, \dots, x_{i-1}) = P(x_i | \Pi x_i) \text{ para todo } i=1..n.$$

Transcrevemos abaixo outro resultado importante que estabelece a dependência de X com uma vizinhança, que será utilizado pelo algoritmo de atualização da rede descrito no Apêndice A.

Teorema 1

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

$$P(x|w_x) = \alpha P(x|u_x) \prod_{j \in \mathcal{Y}_j} P(y_j | f_j(x))$$

onde α é uma constante normalizadora.

CAPÍTULO IV

IMPLEMENTAÇÃO

No presente capítulo, descreveremos a implementação em detalhes, desde a escolha da gramática até os diagnósticos semânticos. Todos os procedimentos são especificados passo a passo, bem como as escolhas justificadas.

A entrada léxica consiste da repetição para todas as posições dos itens léxicos selecionados. Estes itens foram escolhidos a partir das frases-exemplo definidas em função da sua ambigüidade como veremos adiante.

A rede sintática foi construída nos mesmos moldes de Eizirik [1990]. A gramática escolhida foi uma capaz de gerar frases com até dois complementos, preposicionados ou não, sempre depois do verbo. Além disso, os termos nominais podem possuir ou não artigos e adjetivos, sempre antes do substantivo. A gramática assim definida pode ser descrita pelas seguintes regras:

1. $S \Rightarrow C \text{ verbo } C$
2. $S \Rightarrow C \text{ verbo prep } C$
3. $S \Rightarrow C \text{ verbo prep } C \ C$
4. $S \Rightarrow C \text{ verbo } C \text{ prep } C$
5. $S \Rightarrow \text{ verbo } C$
6. $S \Rightarrow \text{ verbo prep } C$
7. $S \Rightarrow \text{ verbo prep } C \ C$

8. $S \Rightarrow verbo C prep C$

9. $S \Rightarrow verbo prep C prep C$

10. $C \Rightarrow art subst adj$

11. $C \Rightarrow art subst$

12. $C \Rightarrow subst$

As maiores sentenças que são geradas pela gramática dada são: *art adj subst verbo prep art adj subst prep art adj subst*, possuindo 12 posições, portanto estipulamos este comprimento máximo para as entradas léxicas. A entrada consiste então das categorias gramaticais cujas ocorrências são possíveis em cada posição, e aparece na primeira linha da figura 8.

Deste modo foram criadas unidades correspondentes a aplicação das regras aos termos gramaticais e às sub-unidades assim construídas até a obtenção de todos diagnósticos sintáticos possíveis a partir dos itens léxicos em questão. As setas entre os termos integrantes e os subtermos têm que ser orientadas de modo a permitir diferentes combinações entre aqueles, portanto a ocorrência dos diagnósticos finais ou parciais depende da ocorrência dos termos que os integram. Graficamente isto corresponde a orientar as setas dos termos mais simples para os mais complexos e atribuir adequadamente a matriz de probabilidade associada a cada nó.

A rede sintática foi construída para entradas até tamanho doze. A partir das categorias gramaticais (terminais da gramática), aplicam-se das regras 10,11 e 12 que agrupam os termos nominais. O resultado da aplicação das regras 1 a 9 são os diagnósticos sintáticos. Existem também sub-unidades intermediárias que não correspondem a diagnósticos. A nomenclatura adotada para os termos nominais identifica a posição inicial e a posição final do grupo. Já para os demais grupos, aparece apenas a posição inicial de cada termo.

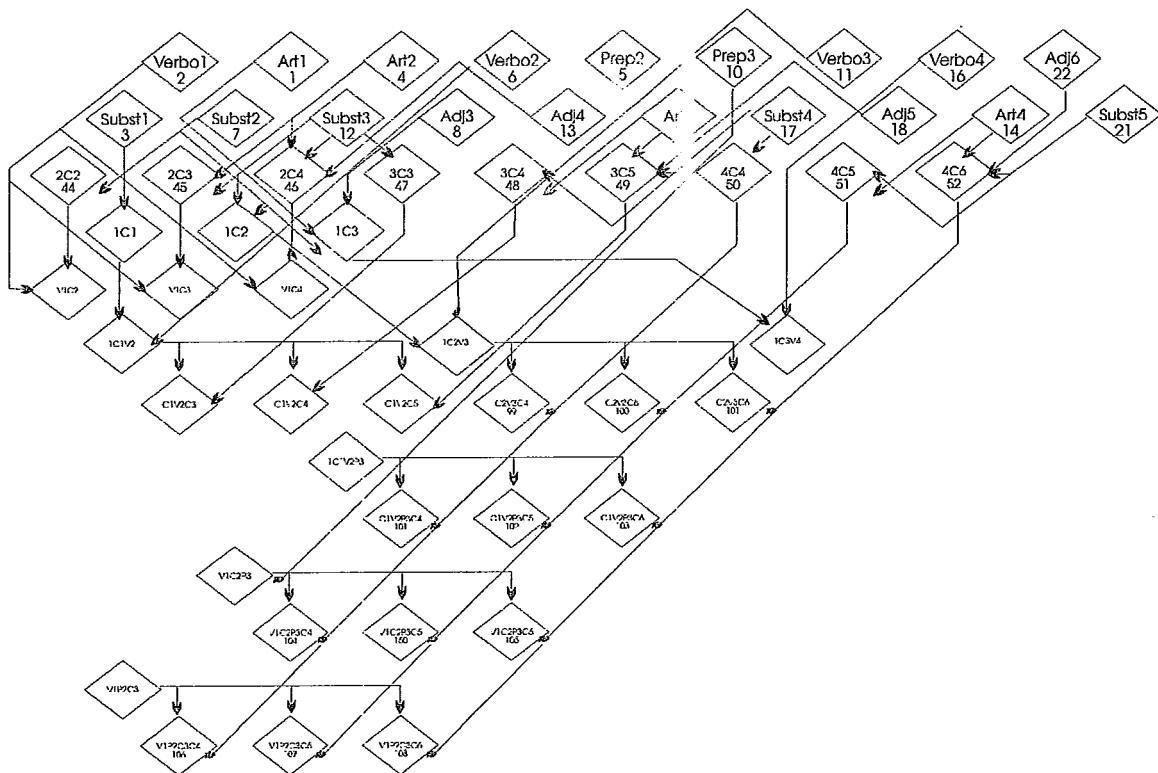


Fig.8 Trecho da sub-rede sintática mostrando diagnósticos até tamanho 6

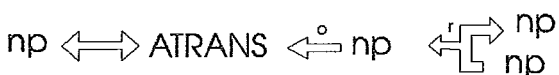
No exemplo *João fixa a peça* selecionamos as categorias *subst1*, *verbo2*, *adj2*, *art3*, *prep3* e *subst4*. As unidades *art3* e *subst4* se agrupam na unidade 3C4 que corresponde a um sintagma nominal (np) começando na posição três e terminando na posição quatro. Esta por sua vez se agrupa com as unidades *subst1*, *verbo2* na unidade S1 V2 C4 que corresponde ao diagnóstico sintático correto para a sentença em questão. Note-se que as unidades *prep3* e *adj2* não integram o diagnóstico tratando-se de uma solução de ambigüidade a nível sintático.

A interligação entre as redes merece consideração mais detalhada. Para identificar um dado item da entrada *joão2_com* o termo ao qual corresponde *joãonp1* por exemplo, é necessário informações da rede sintática, a informação de que o substantivo da primeira posição pertence ao primeiro grupo nominal. Criamos então os nós *subst_{ij}npj* que identificam o papel com a posição na rede.

Para a rede semântica utilizamos os casos, as primitivas e as categorias de Schank [1975]. Seleccionamos um conjunto de verbos que representasse as classes definidas pelas primitivas. Foram escolhidos verbos que mais se aproximassem das primitivas de modo a reduzir a complexidade da rede, isto é, verbos que correspondessem ao sentido expresso por uma primitiva. Como o teste da rede será a solução da ambigüidade semântica, tentamos encontrar verbos que possuíssem mais de uma acepção desde que literal.

Para cada verbo seleccionado foi construída uma estrutura contendo uma primitiva e os casos conceituais correspondentes, bem como os conceitos e as dependências.

A primitiva *ATRANS* requer três casos conceituais: agente, objeto e recipiente. A *estrutura conceitual* correspondente é a seguinte:



É importante ressaltar que o *np* da esquerda (*agente*) pode coincidir tanto com a *doador* do caso *recipiente*, quando o *objeto* é transferido do *agente*; ou com o *receptor*, quando o *objeto* é transferido para o *agente*. Dentre os verbos escolhidos *fornecer*, *perder* no sentido de transferir pertencem ao primeiro caso, enquanto *tomar* e *levar*, também no sentido de transferir corresponde ao segundo caso. as estruturas resultantes são as seguintes:

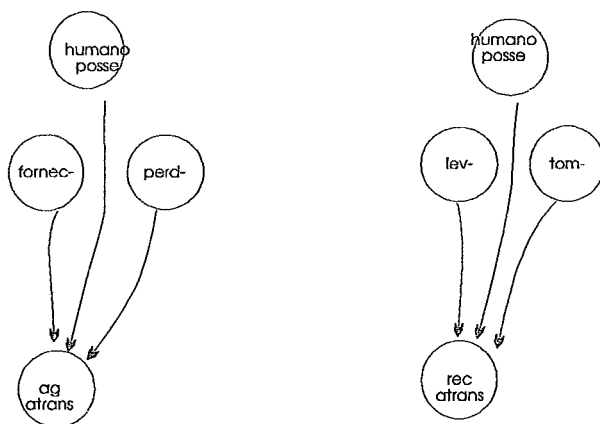
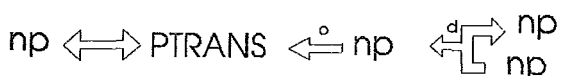


Fig. 9 Estruturas correspondentes a PTRANS

A primitiva *PTRANS* requer três casos conceituais: agente, objeto e recipiente. A *estrutura conceitual* correspondente é a seguinte:

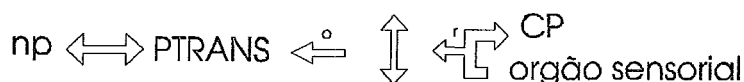


É importante ressaltar que o *np* da esquerda (*agente*) pode coincidir tanto com a *origem* do caso *diretivo*, quando o *objeto* é transferido do *agente*; ou com o *destino*, quando o *objeto* é transferido para o *agente*. Dentre os verbos escolhidos *fornecer*, *perder* no sentido de transferir pertencem ao primeiro caso, enquanto *tomar* e *levar*, também no sentido de transferir corresponde ao segundo caso. as estruturas resultantes são as seguintes:



Fig. 10 Estruturas correspondentes a ATRANS

A primitiva *MTRANS* requer três casos conceituais: agente, objeto e recipiente. A *estrutura conceitual* correspondente é a seguinte:



Caso a origem seja externa ao *agente*, teremos entrada de informação. Caso contrário, haverá fornecimento. Dentre os verbos escolhidos *sentir* no sentido de perceber corresponde ao primeiro caso, enquanto *fornecer*, no sentido de transferir *informação* corresponde ao segundo caso. As estruturas resultantes são as seguintes:

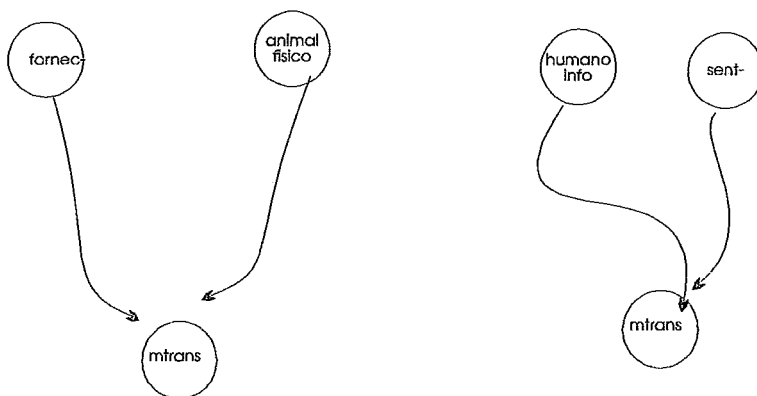


Fig. 11 Estruturas correspondentes a MTRANS

A primitiva *GRASP* requer três casos conceituais: agente, objeto e recipiente. A *estrutura conceitual* correspondente é a seguinte:

np \longleftrightarrow GRASP \leftarrow° np

O verbo que expressa esta primitiva é *fixar* no sentido de impedir o movimento. A estrutura contendo um caso *instrumento* corresponde de acordo com Schank [1975] a uma estrutura complexa, contendo uma ação, *PTRANS* no caso. As estruturas resultantes são as seguintes:

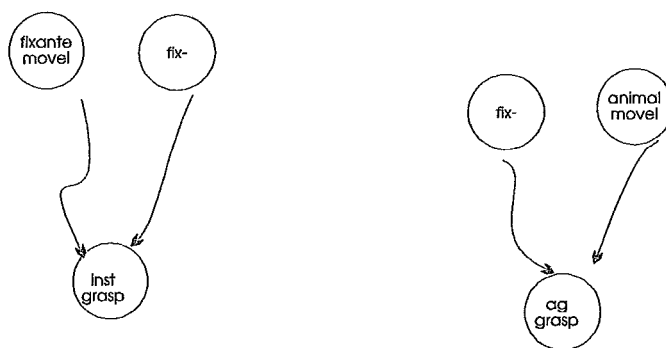


Fig. 12 Estruturas correspondentes a GRASP

A primitiva *PROPEL* possui caso *recipiente* ou *diretivo* conforme o caso. A *estrutura conceitual* correspondente é a seguinte:



O verbo selecionado para exprimir o significado desta primitiva foi *mover* no sentido de impulsionar. Mais uma vez o aparecimento do caso *instrumento* corresponde a uma estrutura complexa, que neste caso envolve uma mudança de estado de um objeto. As estruturas resultantes são as seguintes:

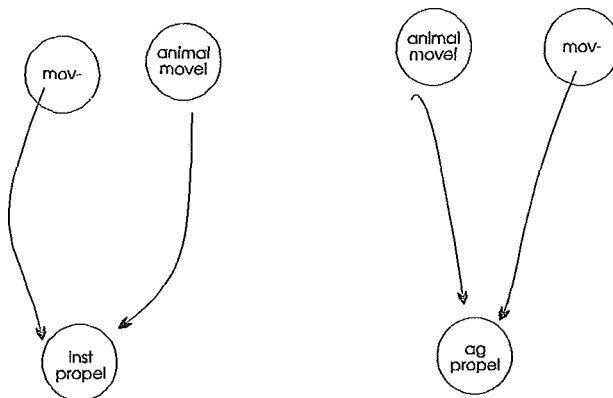
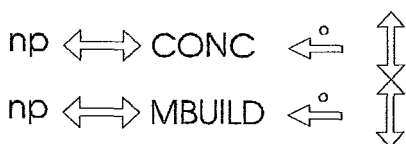


Fig. 13 Estruturas correspondentes a MTRANS

A primitiva *CONC* pode expressar os verbos *imaginar*, *sonhar*, *refletir* e *ter em mente* entre outros, dependendo da *Ação* que ocupa o caso *objetivo*. A primitiva *MBUILD* possui significado semelhante, envolvendo no entanto As *estruturas conceituais* correspondentes são as seguintes:



Os verbos selecionados para representar a primitiva foram *sentir*, no sentido de vivenciar e *pensar* significando ter em mente. Para não aumentar a complexidade, escolhemos para *objeto* *np* e não *Ações*. As estruturas assim construídas são as seguintes:

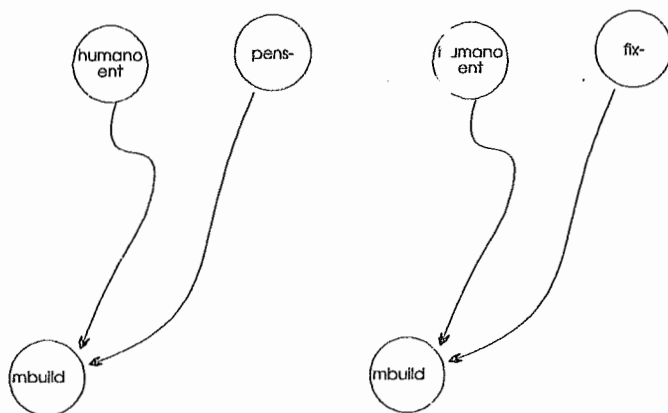


Fig. 14 Estruturas correspondentes a MBUILD

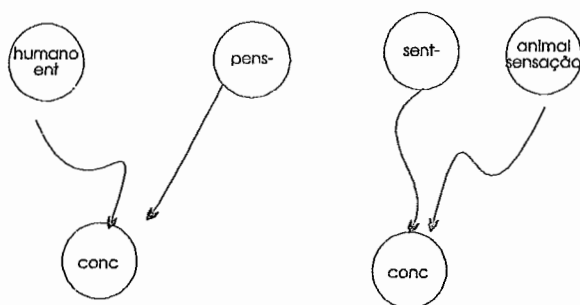
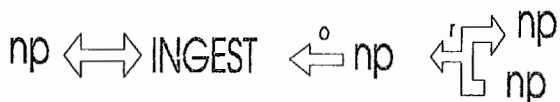
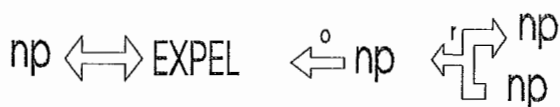


Fig. 15 Estruturas correspondentes a CONC

As primitivas *expel* e *ingest* possuem os mesmos casos, sendo que agente se confunde com o *doador* do caso *recipiente* ou com o *receptor*. As estruturas conceituais correspondentes são:



Os verbos escolhidos para expressar estas primitivas foram *perder* significando expelir e *tomar* significando ingerir. As estruturas correspondentes são:

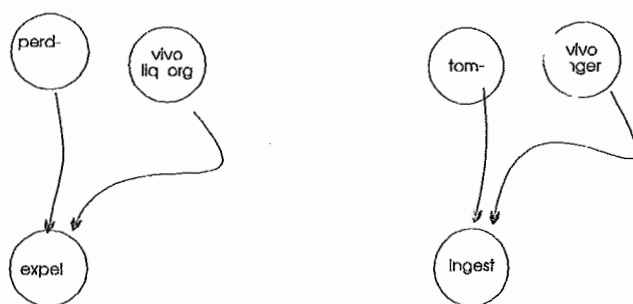


Fig. 16 Estruturas correspondentes a EXPEL e INGEST

Criamos então unidades de diagnóstico semântico que classificam os nps em agente e identificam as primitivas, possibilitando assim a solução da ambigüidade tanto do papel semântico do termo nominal como da acepção do verbo. Para reduzir a complexidade inserimos nós intermediários que associam nps e são compartilhados por diferentes diagnósticos.

Na figura abaixo estão representados os radicais dos verbos selecionados, bem como os possíveis casos para cada um. A rede semântica foi então construída a partir desta camada com os fragmentos referentes aos verbos servindo de base para interligação dos demais itens léxicos.

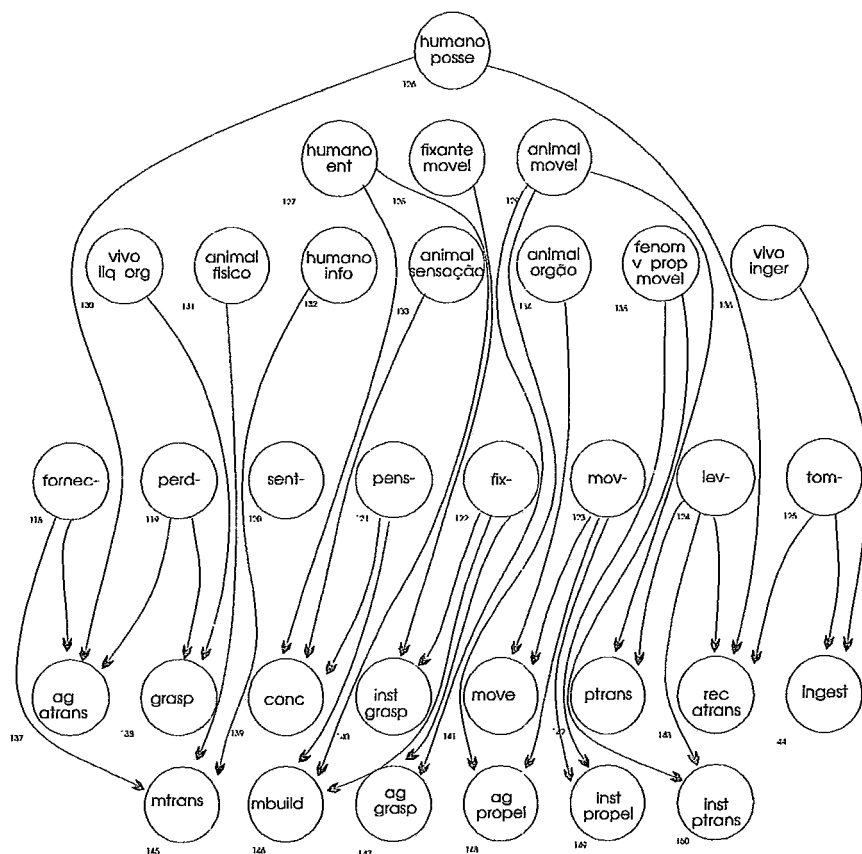


Fig. 17 Analisador semântico

A parte nominal da rede semântica foi definida em função das micro características relevantes à compreensão dos verbos. Outro critério foi representar apenas as unidades necessárias à classificação dos itens nominais em questão.

Definimos então um léxico adequado ao problema a partir da escolha das frases exemplo, uma para cada acepção dos verbos escolhidos. Estas frases aparecem a seguir:

João fornece a data

João fornece dinheiro.

O vento move o barco.

João toma o dinheiro

João move a mão.

João fixa a peça.

João fixa a data.

O vento leva o dinheiro.
Maria perdeu dinheiro.
Maria sente o vento.
João leva o dinheiro.
João sente calor.
João toma água.
Maria perde sangue.
João pensa numa data

Os itens léxicos escolhidos são então definidos e classificados usando o formalismo descrito por Woods. Primeiramente cada item léxico foi identificado com uma categoria, conforme a figura abaixo.

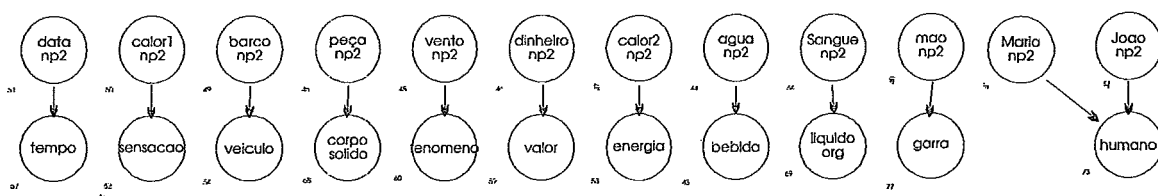


Fig. 18 Inclusão do léxico em categorias

Estas categorias são definidas a partir de categorias mais gerais como por exemplo:

veículo / engenho, propelido

engenho / físico, artificial

bebida / inger, subst liq

liq org / org, subst liq

org / subparte, corpo

A partir destas definições, utilizando critério de generalização estrutural enunciado no capítulo II, deduzimos as generalizações possíveis que são então incluídas na rede como generalizações gravadas. A estas

adicionamos as generalizações axiomáticas, que conforme mencionamos anteriormente são aquelas incluídas diretamente na base de dados pelo especificador.

A rede resultante, denominada classificador aparece na figura 19. As probabilidades foram atribuídas de acordo com o caso, pois às vezes duas categorias de cuja ocorrência dependa uma terceira podem ocorrer simultaneamente.

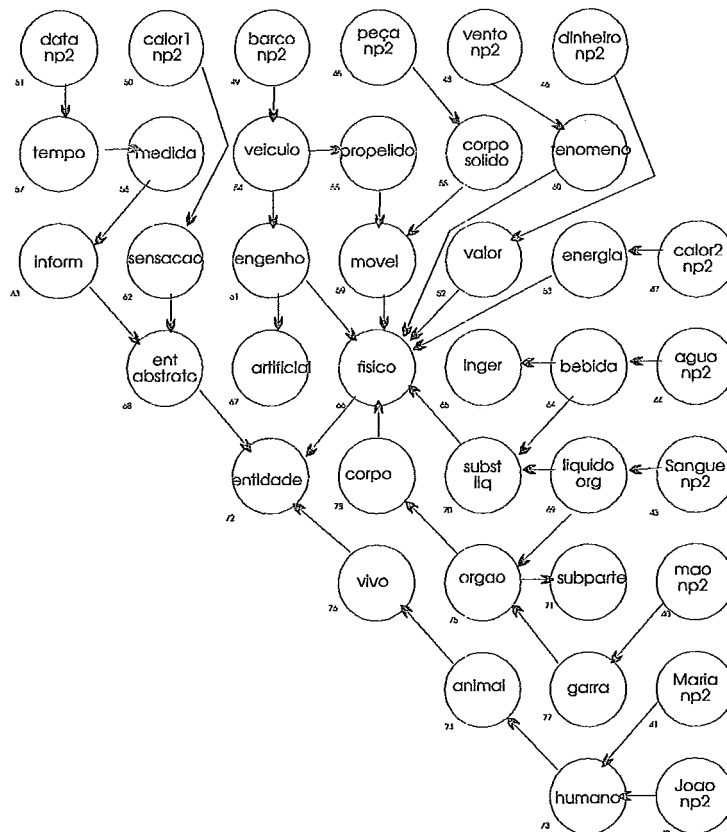


Fig. 19 Classificador para os termos nominais

Esta estrutura aparece repetida na rede semântica para cada np, estando interligada como vimos à entrada através dos itens do léxico e ao analisador semântico que utiliza esta classificação para determinar o diagnóstico adequado à frase. Um esquema completo da rede semântica, incluindo o classificador e o analisador aparece na figura 19, onde a parte central corresponde à figura 17 e as rotuladas np_i à figura 18.



Fig. 20 Esquema geral da rede semântica

CAPÍTULO V

RESULTADOS, CONCLUSÕES E EXTENSÕES

No presente capítulo, apresentaremos os resultados das simulações analisando a performance do sistema na resolução da ambigüidade léxica. Verificamos se o comportamento de cada uma das partes da rede está de acordo com a especificação. Apreciamos então a adequação do sistema projetado ao problema em questão. Por fim sugerimos ampliações e melhoramentos.

Foram feitos testes na sub-rede semântica e na rede completa. Os testes se concentraram na parte semântica, pois testes da sintática podem ser encontrados na bibliografia. Para a sub-rede semântica, grampeamos as unidades cuja ativação decorre dos diagnósticos sintáticos, isto é, os nós tipo *itemléxicon* p_j e os nós verbais. Grampear uma unidade significa atribuir-lhe um valor (0 ou 1) e impedir que este varie durante a simulação.

A simulação da parte semântica com temperatura inicial $T = 1.000$, e fator de redução $\alpha = .95$ estabilizou a partir de 150 iterações quando $T < 10^{-1}$ como podemos ver na tabela que aparece no apêndice B junto com uma representação da sub-rede semântica. Esta tabela corresponde à simulação da frase *João sente o vento*.

Assim, de acordo com a tabela, podemos verificar que as unidades 3, 48 e 120 correspondentes a *Joãonp1*, *ventonp2* e *sent-* estão grampeadas, isto é, não podem variar durante a simulação. Na figura a seguir aparecem as unidades cujo valor ao fim da simulação era um, ainda de acordo com a tabela. Estas unidades correspondem à classificação de João como agente, vento como objeto e sentir com a acepção correspondente à primitiva MTRANS, isto é, como um verbo de percepção.

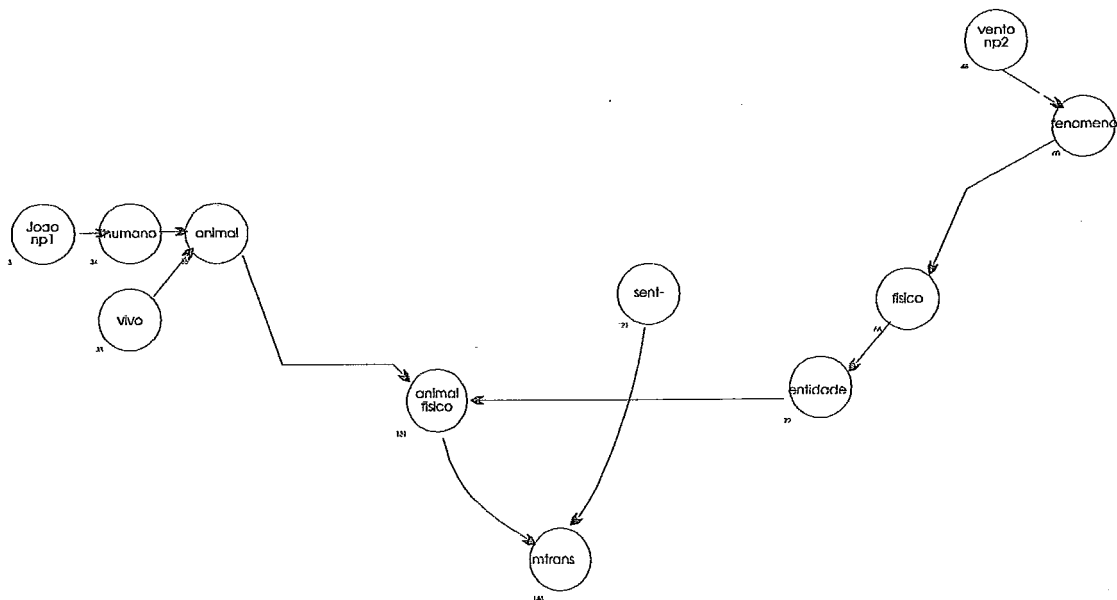


Fig. 21 Trecho da rede semântica

Para verificar o paralelismo das sub-redes foram realizados também testes com a rede inteira. Além de verificar a ativação do diagnóstico semântico adequado, determinando tanto a aceção correta do verbo como os papéis desempenhados pelos termos nominais, devemos ressaltar a ativação das características dos termos nominais resolvendo possíveis ambigüidades como no exemplo a seguir:

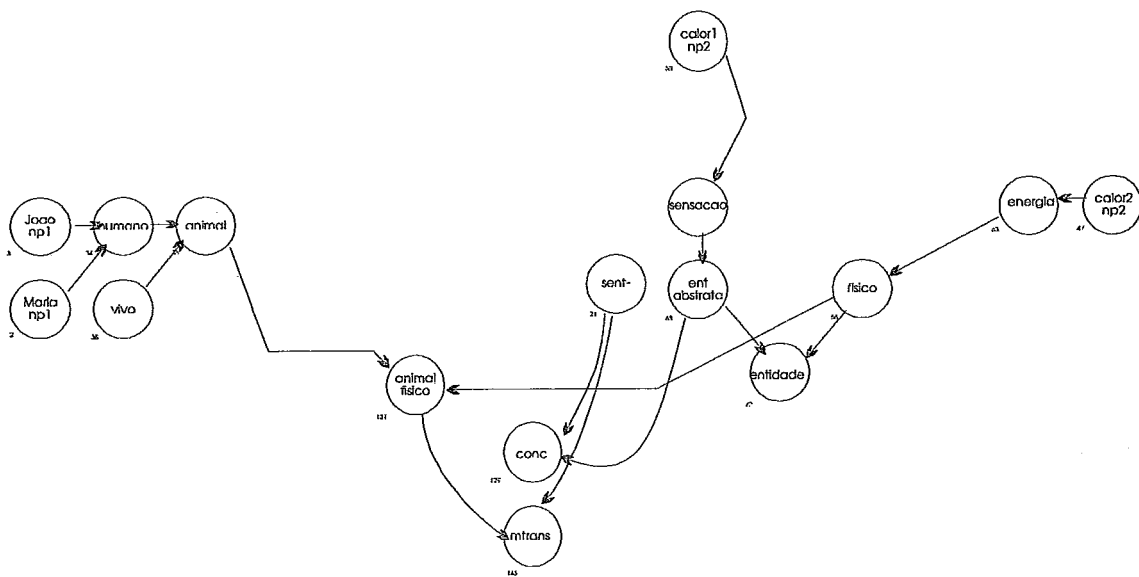


Fig. 22 Trecho da rede semântica

Na figura acima aparecem as unidades relevantes à simulação *João sente calor* onde *calor* possui duas acepções. As duas acepções são possíveis, a rede estabilizou na que requeria objeto mais específico, inibindo a que requeria o mais geral. A simulação indicou sensação como a acepção correta para o item léxico *calor*. Trata-se de um caso de herança múltipla. Caso quiséssemos a outra acepção fosse selecionada teríamos que modificar as probabilidades condicionais de modo a tornar aquela mais provável. Já no caso de *João pensa numa data* como não há conflito a rede exibiu os dois diagnósticos como prováveis, ambigüidade que poderia ser removida se houvessem mais participantes ou informação do contexto.

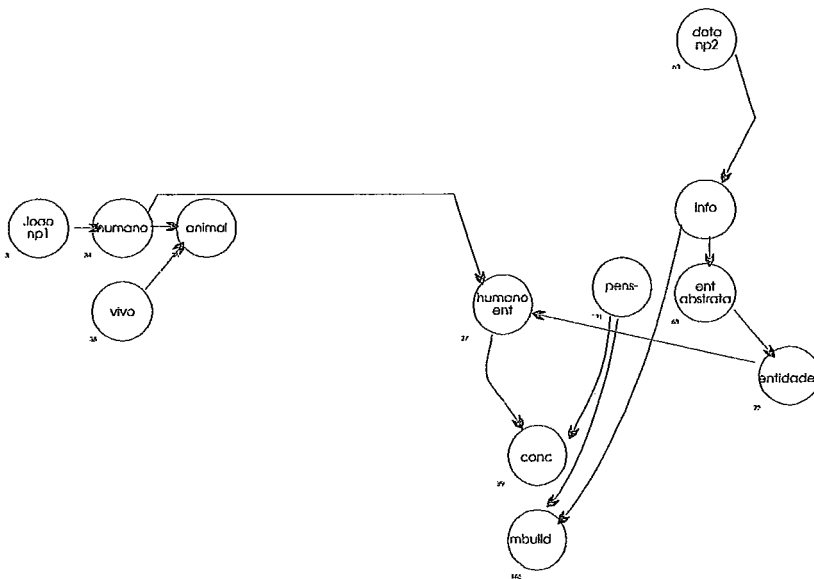


Fig. 23 Trecho da rede semântica

Outro comportamento interessante da rede durante a simulação é o tratamento de exceções. No caso dos verbos *atrans* o objeto requerido é do tipo *posse* que é qualquer *ent.física* exceto *corpo*. Assim atribuímos à probabilidade condicional $p(\textit{posse} / \textit{ent.física} \ \& \ \sim\textit{corpo})$ um valor alto e para $p(\textit{posse} / \textit{ent.física} \ \& \ \textit{corpo})$ um valor baixo.

Em vista dos resultados, verificamos que a escolha do modelo Bayesiano possibilitou o tratamento de informações incompletas, herança

múltipla e exceções em uma representação uniforme e com uma semântica bem definida.

O modelo, como já havia sido verificado em Eizirik [1991], é adequado a aplicação de regras. A definição dos conceitos em função de sua utilização na rede embora por vezes incompleta foi suficiente para determinar as características relevantes dos itens participantes possibilitando a resolução da ambigüidade. Os problemas de herança múltipla que porventura ocorreram quando da entrada de frases com ambigüidade estrutural ou de contexto, podem ser tratados com atribuição de probabilidades. O tratamento probabilístico se revelou adequado também ao tratamento de exceções. Apesar dos arcos serem indiferenciados, a localização acrescenta as informações necessárias à construção da rede.

A identificação dos termos nominais com categorias dispostas em uma hierarquia serviu para identificar as características relevantes dos participantes da frase. O mecanismo de herança possibilitou a compactação da informação semântica numa hierarquia não linear.

São necessários testes com relações compostas definidas a partir de combinações de relações simples para verificar se a herança das Descrições Papel/Participante se dá conforme sugere Brachman [1978]. Seriam necessários operadores semelhantes a Chain, Restrict e operadores numéricos para a geração de tais relações compostas. Assim poderemos representar verbos complexos que conjugam as ações representadas por outros como por exemplo *andar* pode ser analisado como uma combinação de *move* (mover o corpo) e *ptrans* (deslocamento físico).

Os operadores quantificacionais não foram incluídos em nossa rede por motivo de simplificação, devendo sua inclusão ser feita como atributo das relações. Do mesmo modo, a diferenciação estrutural / assertivo não foi explicitada pois o propósito é apenas resolução da ambigüidade. Esta distinção pode ser introduzida juntamente com os tags quantificacionais.

O tratamento de problema de referência e attachment requer a inclusão de relações complexas e seriam necessárias tantas hierarquias taxonômicas

quantos fossem os elementos referenciados na definição do conceito, caso contrário não poder-se-ia identificar os participantes do conceito. As relações complexas requerem a inclusão dinâmica de nós, que correspondem aos papéis a serem preenchidos.

A saída do sistema pode servir de entrada para um classificador nos moldes de Woods [1991] adequado à representação do conhecimento.

O sistema carece ainda de uma interface amigável para especificação e teste das redes, bem como um gerenciador para as simulações, que aceitasse entradas em linguagem natural e fornecesse os diagnósticos correspondentes. Um simulador paralelo tem seu algoritmo descrito em Eizirik [1991]. Uma análise detalhada da complexidade dos algoritmos de simulação é necessária antes de tentar-se qualquer implementação mais ambiciosa.

APÊNDICE A

MÉTODOS DE SIMULAÇÃO

O método da simulação estocástica consiste em calcular as probabilidades dividindo o número de ocorrências pelo número total de simulações. Com a simulação estocástica é possível resolver o problema da complexidade da determinação das probabilidades de ocorrência de cada unidade. Além disso é possível uma implementação paralela pois existem unidades cujas vizinhanças não se sobrepõem e podem ser atualizadas simultaneamente.

Na presente tese implementamos um simulador estocástico com annealing (Barbosa [1993]) objetivando acelerar a convergência da rede. A simulação sem annealing, por começar por um estágio aleatório, utiliza um número elevado de passos para anular o impacto do estágio inicial sobre a rede. Já na simulação com annealing, as probabilidades são calculadas a partir da seguinte forma:

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

A introdução do parâmetro T permite a inicialização da rede num estágio global de excitação médio e a aproximação em passos geométricos do estado final.

Analisando o comportamento da probabilidade p em função de T vemos que quando T é elevado a probabilidade tende a

$$p = \frac{1}{1 + e^{\ln(b/a)/\infty}} = \frac{1}{1 + e^0} = \frac{1}{2}$$

Onde a é a = p(1) e b = p(0) Por outro quando T se aproxima de 1

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

Por outro lado quando T se aproxima de 0, se $b > a$:

$$p = \frac{1}{1 + e^{\ln(b/a)/0}} = \frac{1}{1 + e^{\infty}} = \frac{1}{1 + \infty} = 0$$

e se $b < a$:

$$p = \frac{1}{1 + e^{\ln(b/a)/0}} = \frac{1}{1 + e^{\infty}} = \frac{1}{1 + \infty} = 0$$

De acordo com Barbosa [1993] o comportamento da rede quando T se aproxima de 1 o valor das unidades se aproxima do valor médio , enquanto quando T se aproxima de zero a rede se aproxima do estado global mais freqüente. No entanto como veremos nos resultados experimentais a nossa rede se estabiliza com valores de T da ordem de 10^{-1} .

Os valores associados às probabilidades condicionais foram 0.995 (próximo de 1) e 0.005 (próximo de zero). Caso houvesse unidades onde ocorre herança múltipla, como no caso de ambigüidade , podemos ajustar estes valores de modo a regular o impacto da ocorrência das unidades conflitantes sobre o maior generalizador comum , forçando a ocorrência de uma das unidades. Esta unidade seria aquela de cuja ocorrência o maior generalizador comum dependesse com a probabilidade mais alta.

O algoritmo utilizado foi o seguinte:

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

Saída: O estado global mais freqüente

Método:

1) Para $i = 1, \dots, n$ faça:

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

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

1.1.2) $d^1 := \prod_{\substack{Y \in W \\ Y\text{-filho-de-}X}} P(Y = y \mid X = 1, \text{valores_dos_outros_pais_de_}Y)$

1.1.3) $d^0 := \prod_{\substack{Y \in W \\ Y\text{-filho-de-}X}} P(Y = y \mid X = 0, \text{valores_dos_outros_pais_de_}Y)$

1.1.4) $a := c d^1$

1.1.5) $b := (1 - c) d^0$;

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

1.1.7) atribua a X o valor 1 como probabilidade de p .

2) Pare, os valores de X correspondem ao estado global mais freqüente.

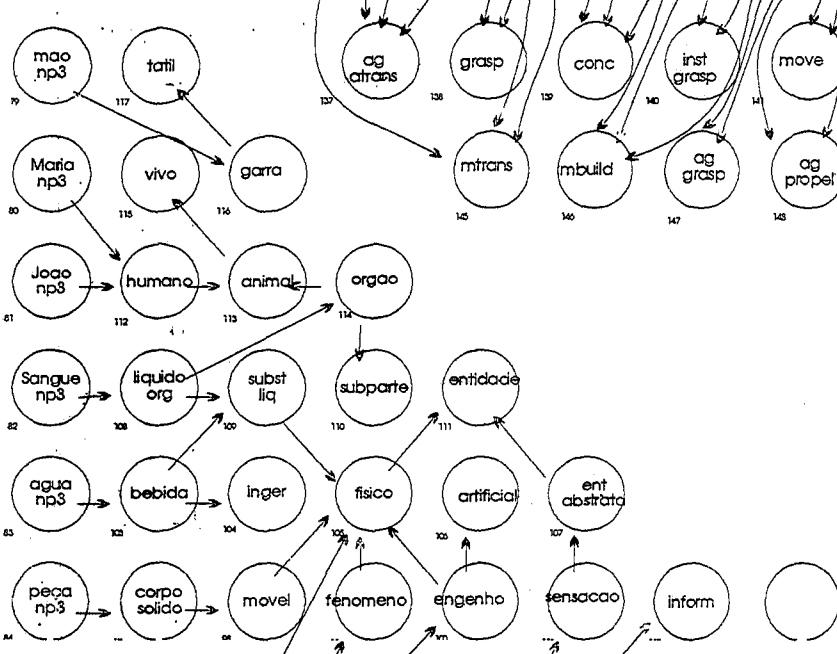
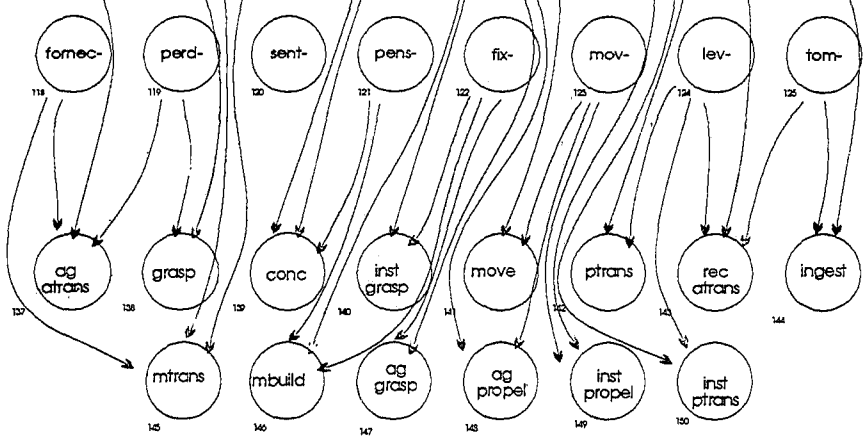
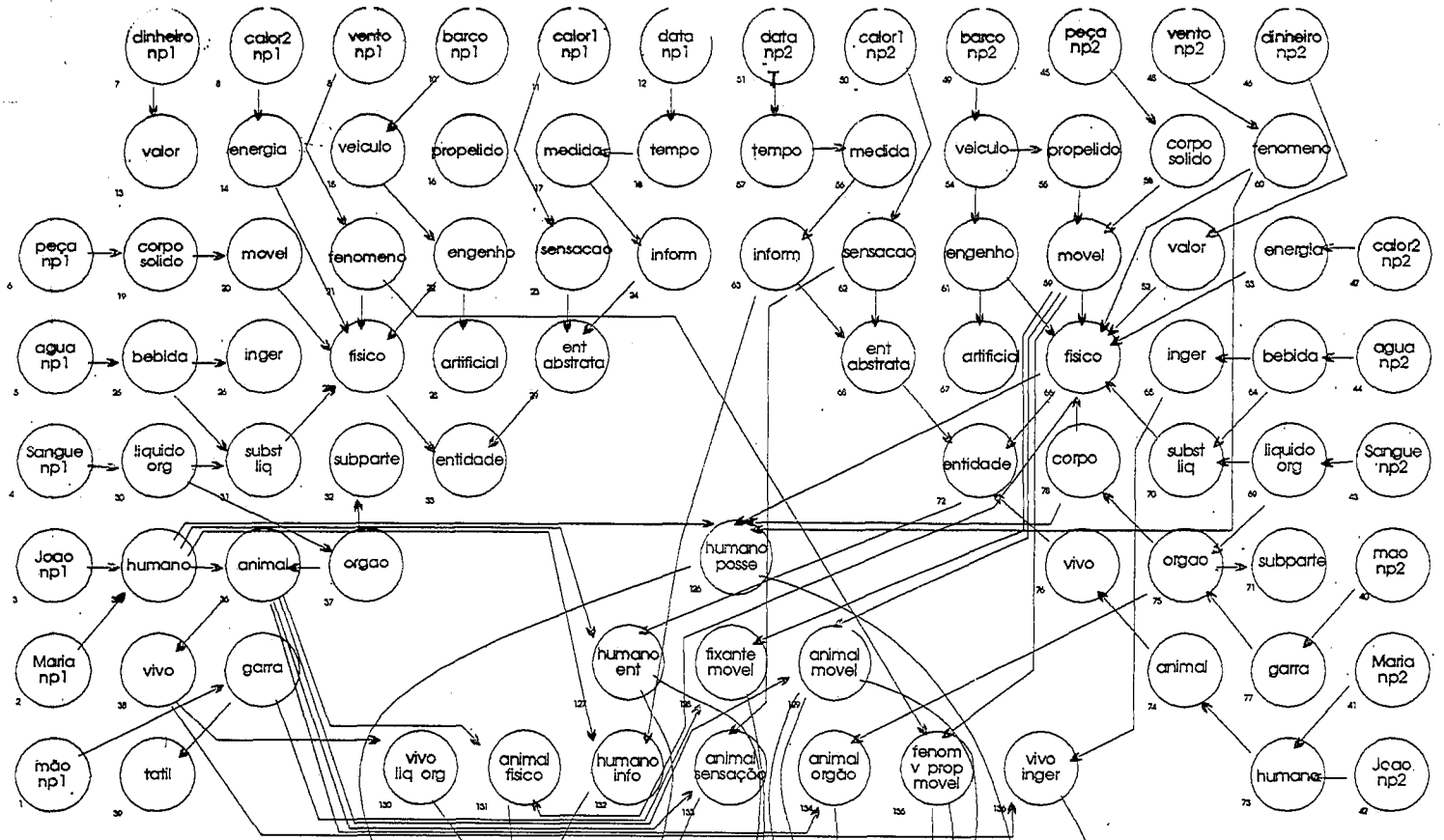
No primeiro passo faz-se a inicialização

No passo dois atualizam-se os valores para todas as unidades: Primeiro calcula-se a probabilidade p de X dados os estados dos vizinhos. Em seguida escolhe-se um valor para X favorecendo o estado 1 com probabilidade p .

O número de passos n , a temperatura inicial T_i e final T_f variam de acordo com o tamanho da rede e a complexidade da mesma. No presente trabalho, a rede se estabilizou com temperatura final $T_f < 10^{-1}$. Os parâmetros foram regulados de modo a permitir um número suficiente de iterações antes de estabilizar. Assim garantimos que a estabilização não se deu num estado aleatório.

O algoritmo usando simulação estocástica com annealing apresentado aqui converge para a configuração conjunta mais provável de acordo com

Hrycej (1990). Outros resultados importantes a respeito, bem como um algoritmo paralelo aparecem em Barbosa (1993).



BIBLIOGRAFIA

BARBOSA, V. C. (1993) *Massively Parallel Models of Computation*, London, Ellis Harwood.

BRACHMAN, R. J. (1979), *On the Epistemological Status of Semantic Networks*, in Findler, N. V. (ed.), *Associative Networks: Representation and Use of Knowledge by Computers*, New York, Academic Press.

BRACHMAN, R. J., FILKES, R. E., and LEVESQUE, H. J. (1983), *KRYPTON: A functional Approach to Knowledge Representation*, *IEEE Computer*, 16(10): 65 - 73.

BRACHMAN, R. J.. (1979), *On the Epistemological Status of Semantic Networks*, in *Associative Networks: Representation and Use of Knowledge by Computers*. Findler, N. V (ed.), New York, Academic Press.

EIZIRIK, L. M. R. (1990), *Uma Abordagem Baseada em Redes Semânticas para Solução da Ambigüidade Léxica*, Tese de Doutorado, Engenharia de Sistemas e Computação, COPPE/UFRJ, Rio de Janeiro.

EIZIRIK, L. M. R., BARBOSA, V. C., MENDES, S. B. T.(1993) *Bayesian Network Approach to Lexical Disambiguation Problem*, in *Cognitiva Science*, 17 (2), Ablex Publishing Corporation, Norwood, NJ.

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.

HAYE, P. J. (1977), *Some Association- Based Techniques for Lexical Disambiguation by Machine*, TR 25, Computer Science Department, The University of Rochester, Rochester, N. Y.

HENDIX, G. G. (1979) Encoding Knowledge in Partitioned Networks, in Findler, N. V. Associative Networks: Representation and Use of Knowledge by Computers, New York, Academic Press.

HRYCEJ, T. (1991) Gibbs Sampling in Bayesian Networks, in Artificial Intelligence 49 pp 351-363.

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

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

SCHANK, R.C. (1973) Identification of Conceptualizations Underlying Natural Language, in Schank, R. C., and Colby, K. C. (eds.) Computer Models of Thought and Language, Freeman, San Francisco, California.

SCHUBERT, L. K., GOEBEL, R. G., CERCONE, N. J. (1979), in Findler, N. V. (ed.), Associative Networks: Representation and Use of Knowledge by Computers, New York, Academic Press.

SZOLOVIT, P., HAWKINSON, L. B., and MARTIN W. A. (1977) An Overview over OWL, a Language for Knowledge Representation, MIT/LCS/TM 86 Laboratory for Computer Science, MIT, Cambridge, Massachusetts.

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.

WOODS, W. A. (1991), "Understanding Subsumption and Taxonomy: A Framework for Progress", in Sowa, J., Principles of Semantic Networks, Morgan Kaufman, San Mateo, CA.