

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

Processamento de Linguagem Natural

por

DANIEL NEHME MÜLLER

Relatório para a disciplina
CMP187 - Mentis e Máquinas

Prof. Dr. Dante A. C. Barone

Porto Alegre, março de 2003

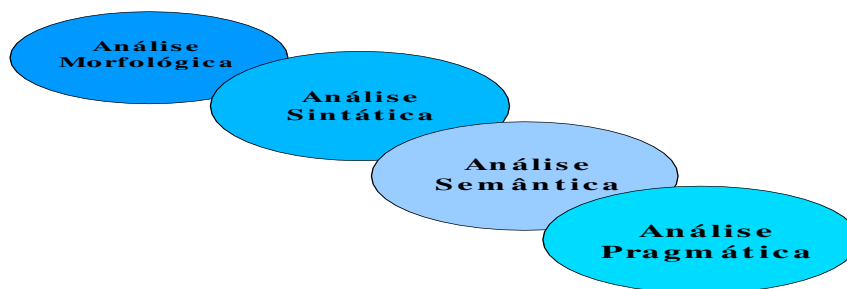


Figura 1: Componentes do Processamento de Linguagem Natural.

1 Introdução

O sonho de construir um computador que se comunicasse com uma pessoa sempre foi buscado pela Inteligência Artificial. Mesmo os mais antigos contos e filmes de ficção científica retratam robôs, computadores e outras máquinas conversando com pessoas. Mas, com o passar dos anos e extensas pesquisas científicas em Inteligência Artificial, foi constatado que um diálogo é extremamente mais complexo do que pode parecer.

Para se ter uma idéia dessa complexidade inerente à linguagem, existe toda uma pesquisa da Ciência Cognitiva que vai do estudo das estruturas neurais do cérebro e das demais redes de percepção de sinais do ambiente até os aspectos psicológicos envolvidos com forma pela qual organizamos nosso conhecimento. Atualmente, estuda-se seriamente o papel das emoções como elemento motor deste sistema de aquisição de conhecimento.

Dentro desse processo de aquisição de conhecimento, a linguagem serve como um facilitador das interações com os objetos de interesse da pessoa. A linguagem falada e escrita, especificamente, serve como facilitador de interações entre pessoas. A idéia, portanto, do campo de pesquisa do Processamento de Linguagem Natural busca que o computador interaja com pessoas como se fosse um outro ser humano.

O problema aí envolvido, como o colega já deve ter inferido, é como modelar o processamento no computador de motivações, crenças, modos de raciocínio, e tantos outros estados mentais necessários ao processo de comunicação através da linguagem falada e/ou escrita. Até o momento, temos apenas aproximações grotescas diante da complexidade real deste processamento, faltando ainda muito o que pesquisar para se chegar a uma modelagem mais próxima ao comportamento humano.

Para dar uma idéia de como o processamento da linguagem escrita é feito hoje, aqui neste relatório da disciplina Mentes e Máquinas, ministrada em forma de seminários pelo prof. Dante Barone, está colocado um registro escrito da apresentação oral deste trabalhos, realizado pelo autor.

2 Processamento de Linguagem Natural

Normalmente, um sistema de processamento de linguagem natural é abordado do ponto de vista da análise do conhecimento morfológico, sintático, semântico e pragmático. A análise morfológica estuda a construção das palavras, com seus radicais e afixos, que correspondem a partes estáticas e variantes das palavras, como as inflexões verbais. A análise sintática diz respeito ao estudo das relações formais entre palavras [JUR 00]. A análise semântica é um processo de mapeamento de sentenças de uma linguagem visando a representação de seu significado, baseado nas construções obtidas na análise sintática [ABR 96]. A pragmática diz respeito ao processamento da forma que a linguagem é utilizada para comunicar, como os significados obtidos na análise

semântica agem sobre as pessoas e seu contexto. Veja figura 1.

Para a implementação de analisadores sintáticos é necessária a construção de um parser para que seja procedida a verificação da posição da palavra na frase. Porém, antes disso, é necessária uma prévia classificação das palavras.

No português, por exemplo, temos 10 classes de palavras: substantivo, verbo, adjetivo, artigo, numeral, pronome, advérbio, preposição, conjunção e interjeição [CUN 82]. Cada uma dessas palavras possuem combinações coerentes com a sintaxe definida pela linguagem. Não é coerente, por exemplo, a colocação de um adjetivo antes de um artigo. A verificação disto é tarefa da análise sintática.

Para a análise semântica são definidos modelos de léxico onde são associados significados a palavras ou expressões. Na implementação de analisadores semânticos são utilizadas diversas técnicas, como redes semânticas, gramáticas semânticas e modelos conexionistas (em geral, redes neurais).

A análise pragmática geralmente é implementada através de quadros (*frames*), casos (*cases*) e roteiros (*scripts*), associando os significados obtidos a um determinado contexto. Normalmente casos de ambigüidade da linguagem são resolvidos neste nível, uma vez que o contexto permite a diferenciação necessária.

Apesar das diversas etapas de processamento da linguagem, há pesquisadores que acreditam ser muito tênue a separação entre reconhecimento sintático, semântico e pragmático. Neal e Shapiro afirmam que alguns aspectos da sintaxe não deveriam ser separados da semântica, uma vez que a mudança de uma palavra que está na frente de outra pode alterar-lhe o sentido. Por exemplo a palavra *areia* na frase *Eu usei areia*, ao colocar um artigo na frente, muda seu significado: *Eu usei a areia*. A primeira frase indica uma areia qualquer, mas a segunda especifica uma determinada areia [NEA 87].

Por outro lado, a semântica também tem íntima relação com a pragmática, uma vez que há palavras que somente encontram seu significado no contexto, como por exemplo a palavra *banco* na frase *Eu tropecei no banco*. Dependendo do contexto, *banco* pode indicar o móvel ou a agência bancária. Este tipo de ambigüidade só pode ser resolvida a nível de análise pragmática.

No decorrer deste relatório serão abordadas todas as fases do processamento da linguagem natural, desde a análise morfológica até a análise de discurso.

3 Análise morfológica

Este tipo de análise é necessário para que o tamanho do dicionário não fique muito extenso, uma vez que é mais simples o armazenamento do radical da palavra e seus afixos. Estes são os componentes que formam uma palavra juntamente com o radical, como prefixos e sufixos. Um exemplo de prefixo é *des* na palavra *desesperança* e de sufixo é *mos* na palavra *calamos*.

O tratamento computacional deste tipo de análise é relativamente simples. Baseia-se em regras que analisam as palavras e as classificam segundo tabelas de afixos. Por exemplo, a entrada *zinho* de uma tabela de sufixos está associada a um diminutivo de um substantivo, portanto, a palavra *bonezinho* é o diminutivo da palavra *boné*, que é seu radical. Desta forma, são reconhecidas as palavras que não estão na sua forma padrão, já adequando-as para a fase posterior de análise sintática. A tabela 1 apresenta o uso de dois sufixos.

Jurafsky e Martin chamam o sistema que realiza a passagem da palavra escrita para a forma classificada de *Conversor de Estados Finitos* (Finit-State Transducer - FST). Este tipo de sistema é o que passa uma palavra de sua forma como é normalmente escrita para uma forma *etiquetada*, ou seja, com identificação de seu radical e afixo [JUR 00]. Na seção 4.1 será apresentado o nível de etiquetagem sintática, complementar à morfológica.

Um exemplo de analisador morfológico para o português é o sistema Palavroso, que realiza a identificação do afixo e pré-classifica as palavras em quatro grupos: verbos, nomes e adjetivos, classes fechadas e advérbios

Tabela 1: Tabela de sufixos.

| sufixo | substantivo (radical) | palavra derivada |
|--------|-----------------------|------------------|
| -zinho | boné | bonezinho |
| | botão | botãozinho |
| -zinha | árvore | árvorezinha |
| | flor | florzinha |

[WIT 98]. A implementação desse sistema tem por base o que foi anteriormente descrito, mas possui ainda outro módulo de dicionário que verifica a existência da palavra construída através das regras morfológicas.

4 Análise sintática

No contexto do processamento da linguagem, as gramáticas utilizadas na análise sintática têm sido chamada de Modelos de Linguagem (Language Models - LM). Esta definição está associada ao amplo universo de frases possíveis de serem modeladas para representação de determinado domínio de análise.

No seguimento, serão abordadas algumas técnicas utilizadas na identificação e etiquetagem dos elementos sintáticos e na organização da construção sintática.

4.1 Etiquetagem

O primeiro processamento que é efetuado na análise sintática é a identificação das classes das palavras (também conhecidas como classes morfológicas, etiquetas lexicais ou partes de fala). Para proceder esta classificação, são implementados parsers que identificam nas frases as classes das palavras que as compõem. Esta classificação de palavras também é conhecida como etiquetagem (tagging).

Por exemplo, a frase *Eu tropecei na pedra*, poderia ser etiquetada da seguinte forma: Eu/PPE tropecei/VP na/PAF pedra/SSF. Esta representação segue a tabela 3, onde são apresentadas as etiquetas, sua descrição e a palavra correspondente na frase apresentada.

Basicamente, a etiquetagem dos atuais parsers sintáticos que vem sendo implementados constituem-se de três tipos: os que usam regras, os que utilizam estatísticas e os híbridos, que utilizam as duas anteriores [PAC 96][JUR 00]. A primeira, que utiliza regras, também é chamada de *lexicalista*, uma vez que ela se preocupa em seguir regras de dicionário e de gramáticas para verificação da consistência da linguagem.

A segunda abordagem chama-se *probabilística*, uma vez que utiliza cadeias de Markov para descobrir qual é a seqüência mais provável de palavras. Dentro deste contexto, a técnica que tem sido mais utilizada em parsers sintáticos probabilísticos é chamada de *n-gramas*, que consiste em estabelecer uma estatística a cada n palavras. Normalmente é implementado $n=3$, sendo que são analisadas 2 palavras para a previsão da terceira. Desta forma, é estabelecida uma estatística de construção de frases, permitindo a descoberta de qual é a palavra de combinação mais provável, permitindo a verificação e correção da construção frasal.

A última abordagem é chamada de *funcionalista* e reúne tanto características referentes ao uso de regras, quanto ao uso de probabilidades. Uma das técnicas mais utilizadas é conhecida como Transformation-Based Learning (TBL), ou Aprendizado Baseado em Transformação, que faz a indução de regras a partir de exemplos de palavras apresentadas. Baseando-se em regras básicas, procura-se inferir a categoria de construção das frases

Tabela 3: Etiquetagem sintática.

| etiqueta | descrição | palavra |
|----------|-------------------------------|----------|
| PPE | Pronome PEssonal | eu |
| VP | Verbo no PAssado | tropecei |
| PAF | Preposição+Artigo Feminino | na |
| SSF | Substantivo Singular Feminino | pedra |

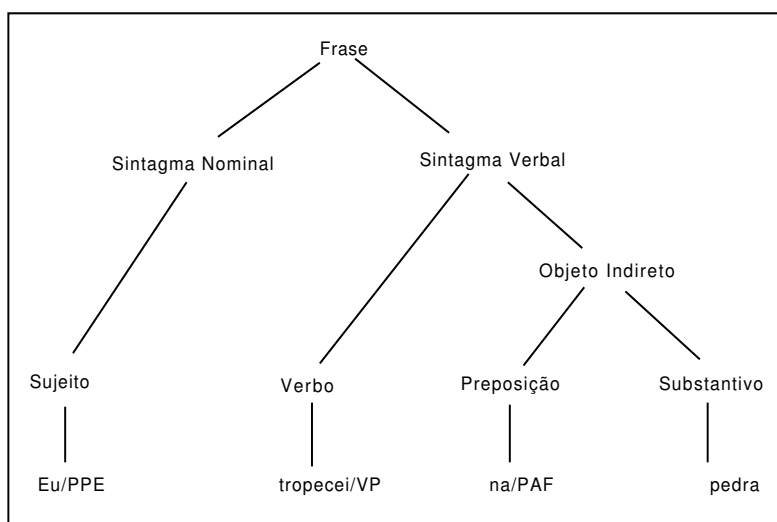


Figura 2: Árvore de parser.

dadas para aprendizado. Sobre este aprendizado, são construídas estatísticas que podem dar origem a novas regras de construção de frases não previstas no modelo básico.

A etiquetagem das palavras, contudo, não basta para a análise sintática. Por vezes ocorre situações ambíguas onde é necessário recorrer-se a mais um nível, uma vez que, dependendo onde se encontra a palavra na frase, ela pode ter a função de advérbio ou de substantivo (ex.: *branco*), de verbo ou conjunção (ex.: *como*), e assim por diante. O próximo passo, portanto, é a análise da construção gramatical da frase para a definição da categoria sintática.

4.2 Construção Gramatical

Após a etiquetagem, passa-se à verificação da construção gramatical da frase. Para tanto, as linguagem é modelada por gramáticas livres de contexto, que divide as frases em árvores de sintagmas nominal e verbal e, a partir daí, verifica que classe enquadra-se em cada sintagma (veja figura 2). Neste sentido, há duas correntes de desenvolvimento: a construção de árvores de parser e de parsers probabilísticos [JUR 00].

4.2.1 Árvores de Parser

Árvores de Parser utilizam técnicas de busca em árvore para determinar a correção da construção frasal, realizando a comparação entre a frase e a estrutura da árvore. Uma técnica bastante difundida que busca resolver os problemas de busca em árvore é o algoritmo de Earley, que baseia-se na transição de estados para

estabelecer qual é a classe mais provável de se seguirá à classe atual [SIK 97][JUR 00]. Este algoritmo possui 3 operadores: *predictor*, *scanner* e *completor*. O primeiro estabelece as transições de estado possíveis, o segundo realiza a comparação entre as transições possíveis e a palavra em análise, e o terceiro módulo realiza o registro final da classe na frase etiquetada. O maior problema deste tipo de parser é a incapacidade de resolver casos de ambigüidade.

Algoritmo de Earley

O algoritmo de Earley realiza uma busca top-down numa gramática estruturada em árvore. Ele começa com a varredura de um vetor chamado *mapa* (chart), contendo tantos elementos quantos forem as palavras da frase em análise. Cada uma destas entradas do vetor contém uma lista de estados gerados pelo parsing até aquela palavra. A lista de estados gerados é composta de uma subárvore que representa uma regra gramatical, de informação sobre o preenchimento da subárvore (regras satisfeitas) e da posição na subárvore de acordo com a palavra do mapa (vetor).

Neste contexto, o operador *predictor* é que cria a subárvore a partir da regra gramatical. Ele aloca as folhas na árvore para estados não-terminais da gramática.

O operador *scanner* verifica qual das ramificações geradas é compatível com a palavra em análise, gera uma nova folha e marca no *mapa* o atual estado da busca.

Por fim, o operador *completer* é necessário quando uma regra terminou de ser avaliada e deve-se realizar o retrocesso (backtracking) para concluir a avaliação das demais ramificações da árvore.

Parsing Probabilístico

O parsing probabilístico utiliza uma gramática cujas entradas possuem ponderações estatísticas (pesos). Estas ponderações são colocadas com base na observação da probabilidade de ocorrência das regras gramaticais no *corpus* (base de palavras). Através destes pesos é possível a resolução de ambigüidades, sendo indicadas as frases de maior probabilidade.

Assim, a probabilidade de uma regra R não terminal ($a \rightarrow b \mid a$) ocorrer numa frase F é a divisão do número de ocorrências de R na avaliação de F pelo número de repetições do símbolo avaliado (a):

$$P(R) = P(a \rightarrow b \mid a) = \frac{\text{cont}(a \rightarrow b)}{\text{cont}(a)}$$

E a probabilidade de uma árvore de parser A para o cálculo da melhor F é o produto de todas as regras envolvidas na construção de A:

$$P(A) = \prod_{n \in A} P(Rn)$$

A partir destas equações é possível saber se as regras usadas para a construção de uma árvore é mais adequada que outra. Veja um exemplo de árvore com probabilidades na figura 3. Isto resolve grande parte dos problemas de ambigüidade nos casos das palavras que podem pertencer a mais de uma classe, como visto na seção 4.1.

Um dos algoritmos empregados para a execução deste tipo de parsing é o Earley probabilístico, que segue o mesmo algoritmo analisado na seção anterior, acrescido dos valores de probabilidades das regras, o que permite a desambigüação.

O maior problema do parser probabilístico é a colocação de pesos iguais a construções frasais diferentes, embora com as mesmas palavras (*ex.: eu tropecei na pedra* poderia ter igual peso de *tropecei na pedra eu*). A

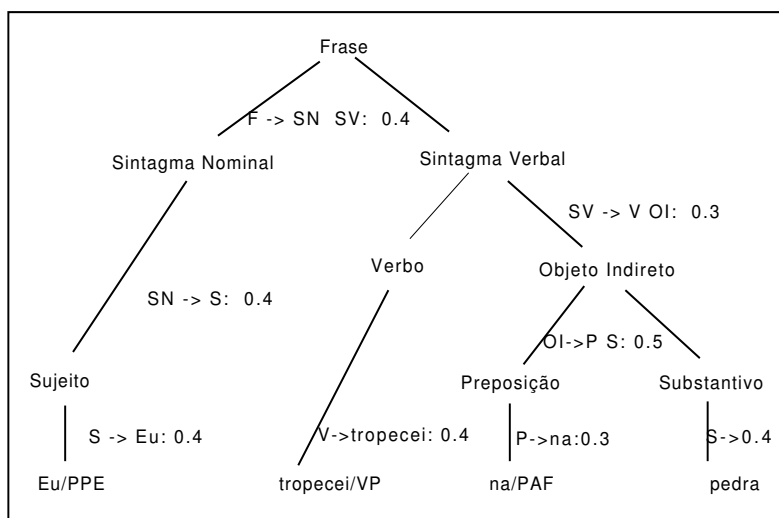


Figura 3: Árvore de parser com indicações de probabilidade.

solução para este caso está na utilização de indicadores lexicais (*lexical heads*), que permite ao algoritmo de busca identificar diferentes ramificações dos sintagmas.

Os indicadores lexicais são palavras que ficam junto das raízes de uma determinada subárvore de parsing. Isso serve para indicar a palavra principal de uma parte da frase (o sujeito num sintagma nominal, por exemplo) e alterar o peso de construções frasais que poderiam ter pesos iguais.

5 Análise Semântica

Apesar do extenso processamento realizado nas análises morfológica e sintática, apenas com estas não é possível distinguir certas categorias de palavras e muito menos precisar o objetivo da frase. Para tanto, são acrescentadas nas árvores de parser os chamados *anexos semânticos* (semantic attachments) [JUR 00]. Este tipo de análise semântica tradicional pode ser construído ainda em tempo de análise sintática, à medida que a árvore de parser vai sendo completada (veja figura 4).

O anexo semântico apresentado na árvore da figura 4 é uma composição de uma rede semântica onde é definido o verbo *tropeçar*. A definição diz que este verbo necessita de um agente (aquele que tropeça) e um paciente (algo em que o agente tropeça).

Por outro lado, há outras formas de análise semântica: gramáticas semânticas (*semantic grammar*), gramáticas baseadas em caso (*case-based grammars*) e outros métodos para casos específicos [CAR 87].

5.1 Gramáticas Semânticas

As gramáticas tradicionais são utilizadas apenas para definir regras de sintaxe. Por outro lado, podem ser definidas regras associadas a frases-padrão, onde variam as palavras envolvidas. É uma forma viável quando existe uma hierarquia de diferentes contextos e necessita-se diferenciá-los através de padrões. Infelizmente, para cada frase analisada, é necessária uma regra distinta, dificultando sua utilização em diferentes domínios do conhecimento [CAR 87][PAC 96][MIN 98] [JUR 00].

A construção de gramática semânticas baseia-se na colocação de variáveis que serão preenchidas de acordo com a afirmação ou questionamento do usuário ao sistema. Um exemplo de regras de uma gramática semântica

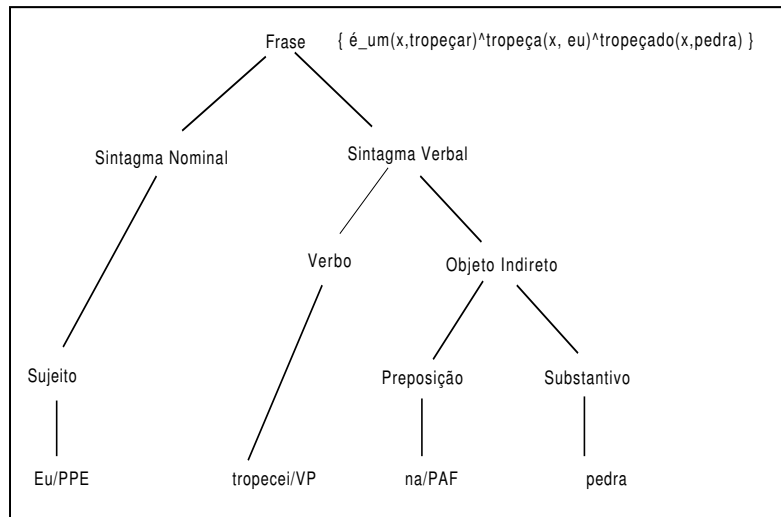


Figura 4: Árvore de parser com anexo semântico.

poderia ser:

QUESTÃO -> *quem tropeçou ONDE*

RESP_QUEST -> *QUEM tropeçou ONDE*

QUEM -> *eu*

ONDE -> *PREPOSIÇÃO LUGAR*

PREPOSIÇÃO -> *na*

LUGAR -> *pedra*

Caso fosse questionado *Quem tropeçou na pedra?*, a resposta seria *eu tropeçou na pedra*, pela associação da expressão *na pedra* com a variável *ONDE* e o pronome *eu* com a variável *QUEM*.

5.2 Gramáticas Baseadas em Casos

Esta forma de análise usa gramáticas semânticas não-terminais para formar padrões. Caso uma dada frase *encaixe* na construção padrão, ela poderá ser reconhecida dentro de um contexto. Este contexto em geral é estruturado na forma de quadros (*frames*), casos (*cases*) [CAR 87][MIN 98][BAR 98].

São construídas gramáticas semânticas para identificar os elementos da frase e indicar o caso (ou quadro) que melhor se aplica a estes elementos. Um exemplo de gramática, para uma aplicação de leitura de correio eletrônico, pode ser o seguinte:

<mensagem> -> <?artigo cabeça-mensagem *caso-mensagem>

<caso-mensagem> -> <%de Pessoa>

<caso-mensagem> -> <%sobre Assunto>

<caso-mensagem> -> <%desde Data>

Onde ? indica um item opcional, * um item passível de repetição e % significa que qualquer palavra da mesma classe daquela indicada a seguir pode ser utilizada; *artigo* indica que pode haver um artigo antecedendo a cabeça da frase; *cabeça-mensagem* é a palavra-chave que indica ser uma mensagem do correio eletrônicos; *caso-mensagem* é a referência para o caso que irá tratar daquela parte da frase; *Pessoa*, *Assunto* e *Data* são variáveis que recebem os dados sobre a mensagem.

Uma frase que poderia se encaixar neste modelo seria *as mensagens de João sobre modelagem desde segunda-feira*. Neste caso, as palavras *João*, *modelagem* e *segunda-feira* servirão de guia para o caso a ser utilizado na resposta do sistema.

A construção de casos para análise semântica antecipa um tipo de processamento geralmente utilizado apenas na análise pragmática, usado para descrever contextos específicos. Na implementação de casos são utilizados *quadros*, que são estruturas utilizadas para caracterizar os elementos semânticos que compõem a frase em análise. *Casos* são, portanto, estruturas contendo quadros utilizados em dado contexto que identificam problemas e ainda outros quadros que identificam soluções para estes problemas. Estas estruturas podem ser usadas na análise semântica para resolver ambigüidades não esclarecidas no processo convencional, baseado em regras.

Um caso para o exemplo citado anteriormente poderia ser o seguinte:

```
[
Nome: Mensagem
Esquema:
  [
  Emissor: &Pessoa
  Assunto: &Assunto
  Data: &Data
  ]
Sintaxe:
  [
  Tipo: Frase-nominal
  Cabeça: (mensagem mensagens mail carta)
  Caso:
    (
    <%de Pessoa>
    <%sobre Assunto>
    <%desde Data>
    )
  ]
]
```

Por outro lado, há quem argumente que a análise pragmática deva ser feita junto com a análise semântica através do uso de casos. A justificativa para isso está na abordagem do estudo das relações existentes na frase de um caso. Dependendo da relação, define-se a função da frase dentro do caso, obtendo-se, ao mesmo tempo, o aspecto semântico (significado) e pragmático (contexto).

6 Análise Pragmática

Este tipo de análise foge à estrutura de apenas uma frase. Ela busca nas demais frases a compreensão do contexto que falta à frase em análise. Em geral, não há estruturas pré-definidas que atendam a uma representação adequada de problemas de referências pronomiais (por exemplo os pronomes *la*, *seu* e *o* no contexto: *João pegou a rosa. Ao pegá-la, seu espinho o espetou*), coerência textual e análise de discurso. Por outro lado, as estruturas mais utilizadas na análise pragmática são os casos, conforme descritos na seção 5.2, que por sua característica já permitem a identificação de contexto.

Para se ter uma noção dos algoritmos utilizados na análise pragmática, pode-se pegar um conhecido algoritmo de resolução de referências pronomiais, proposto por Lappin e Leass em 1994 [JUR 00]. Neste algoritmo, são atribuídos pesos aos pronomes encontrados em frases. Esses pesos são chamados de *valores de saliência* e são atribuídos ao sujeitos encontrados na mesma frase ou nas frases precedentes. Quanto mais próximo do pronome estiver o sujeito encontrado, maior será o valor a ele atribuído como referência ao pronome.

Após ser atribuído pesos aos sujeitos de referência encontrados, retiram-se aqueles que não combinam em gênero e número, e são levadas em conta outras características como o tipo de pronome, se a frase tem objeto indireto ou advérbio, entre outros fatores. Cada uma dessas características podem influir no cálculo do peso da referência, de forma a que o sujeito com maior peso será o selecionado para substituição do pronome. Assim, num contexto como o exemplo citado anteriormente, o pronome *la, seu* e *o* terá como sujeito mais próximo *a rosa*, mas o último pronome (*o*) não combina com o gênero, sendo, então, atribuído ao próximo sujeito, *João*.

Há ainda muitos outros algoritmos citados na literatura acerca de resolução de referências pronomiais, porém a maioria é dependente de uma estrutura sintática previamente determinada ou pouco flexível a diferentes construções frasais. Por esta dificuldade de ser identificada a construção frasal é que tem sido cada vez mais usadas estruturas de preenchimento como *casos* para análise pragmática.

7 À Procura de Novos Modelos

Após esta breve visão do estado-da-arte do Processamento de Linguagem Natural, enfocando a linguagem escrita, percebe-se que as modelagens são extremamente restritas tanto em termos de contexto quanto em capacidade de análise. Urge, pois, que pesquisadores da Ciência Cognitiva busquem novos modelos para uma modelagem mais realista dos processos envolvidos com a linguagem.

Fato é que na última década multiplicaram-se as pesquisas em torno de análises probabilísticas do comportamento da linguagem, uma vez percebido que não era mais possível uma modelagem apenas por processamentos determinísticos. O estudo da linguagem falada também abriu novas frentes de pesquisa para os pesquisadores da língua escrita, uma vez que a fala não possui restrições de ordem gramatical, fugindo quase sempre às regras...

O atual estado da pesquisa em Linguagem Natural, aponta-nos a necessidade de realizar estudos mais aprofundados nos mecanismos fisiológicos, além dos psicológicos, e criar modelos biologicamente plausíveis para o processamento da linguagem. Este é um convite e um desafio a todos os pesquisadores interessados em desvendar os mecanismos da emoção, do pensamento e da linguagem, uma vez que estes sistemas são complementares e necessários para o mútuo desenvolvimento.

Existe uma longa estrada a ser ainda trilhada. Como dizia o poeta: *Caminhante, não há caminho! O caminho fazemos ao caminhar...*

Referências

- [ABR 96] ABRAHÃO, P. R.; LIMA, V. L. S. Um Estudo Preliminar de Metodologias de Análise Semântica da Linguagem Natural. In: **Anais do II Encontro para o Processamento Computacional de Português Escrito e Falado**. XIII SBIA `96. Curitiba: CEFET-PR, 1996.
- [BAR 98] BARKER, K. **Semi-Automatic Recognition of Semantic Relationships in English Technical Texts**. University of Ottawa, 1998. PhD Thesis.

- [BRO 99a] BRØNDSTED, T. The Natural Language Processing Modules in REWARD and IntelliMedia 2000+. In: Kirchmeier-Andersen, S.; Thomsen, T. E. (eds.). **LAMBDA 25**. Copenhagen Business School, Dep. of Computational Linguistics, March 1999, pp. 91-108.
- [CAR 87] CARBONEL, J. G.; HAYES, P. J. Robust Parsing Using Multiple Construction-Specific Strategies. In: BOLC, L. (ed.) **Natural Language Parsing Systems**. Berlin: Springer-Verlag, 1987.
- [CUN 82] CUNHA, C. F. **Gramática da Língua Portuguesa**. Rio de Janeiro: Fename, 1982.
- [JUR 00] JURAFSKY, D.; MARTIN, J. H. **Speech and Language Processing - An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition**. Upper Saddle River: Prentice Hall, 2000.
- [MIN 98] MINKER, W. Stochastic versus rule-based speech understanding for information retrieval. In: **Speech Communication**, [S.l.], 25, Sep. 1998, p. 223-247.
- [NEA 87] NEAL, J. G.; SHAPIRO, S. C. Knowledge-Based Parsing. In: BOLC, L. (ed.) **Natural Language Parsing Systems**. Berlin: Springer-Verlag, 1987.
- [PAC 96] PACHECO, H. C. F.; DILLINGER, M.; CARVALHO, M. L. B. Uma nova abordagem para a análise sintática do português. In: **Anais do II Encontro para o Processamento Computacional do Português Escrito e Falado**. Curitiba: CEFET-PR, 1996.
- [SIK 97] SIKKEL, K. **Parsing Schemata: A Framework for Specification and Analysis of Parsing Algorithms**. Berlin: Springer Berlag, 1997.
- [WIT 98] WITTMANN, L. H.; RIBEIRO, R. D. Recursos Lingüísticos e Processamento Morfológico do Português: o PALAVROSO e o projecto LE-PAROLE. In: Lima, V. L. S. (ed.) **Anais do II Encontro para o Processamento Computacional do Português Escrito e Falado**. Porto Alegre: Tode-schini, 1998.