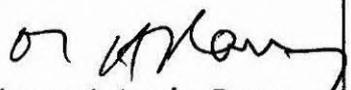


1. Publicação nº <i>INPE-3834-PRE/906</i>	2. Versão	3. Data <i>Março, 1986</i>	5. Distribuição <input type="checkbox"/> Interna <input checked="" type="checkbox"/> Externa <input type="checkbox"/> Restrita
4. Origem <i>DIN/DCS</i>	Programa <i>INTEL</i>		
6. Palavras chaves - selecionadas pelo(s) autor(es) <i>SISTEMA-APRENDIZ. LINGUAGEM NATURAL</i> <i>RELAÇÕES SEMÂNTICAS DOMÍNIOS DE CONHECIMENTO</i>			
7. C.D.U.: <i>681.3,019</i>			
8. Título  <i>LINGUAGEM NATURAL: UM SISTEMA DE RELAÇÕES</i>		10. Páginas: <i>08</i>	
		11. Última página: <i>06</i>	
9. Autoria <i>Carlos Alberto de Oliveira</i>		12. Revisada por  <i>Paltonio Daun Fraga</i>	
Assinatura responsável 		13. Autorizada por  <i>Marco Antonio Raupp</i> Diretor Geral	
14. Resumo/Notas  <p><i>Este artigo pretende propor uma abordagem para a linguagem natural (LN), abordando esta baseada numa estrutura linguística que pode ser considerada como um sistema de relações entre seus componentes (por exemplo, sons, palavras, frases, textos). É proposto neste artigo um sistema aprendiz que divide domínios de conhecimento em relações semânticas concernentes a mundos específicos para otimizar o processamento da LN. Tal sistema aplica-se especialmente a interfaces para sistemas especialistas.</i></p>			
15. Observações <i>Trabalho submetido a apresentação no VI Congresso da Sociedade Brasileira de Computação, 19-25 de julho de 1986 em Recife-Pe.</i>			

## LINGUAGEM NATURAL: UM SISTEMA DE RELAÇÕES

Prof. CARLOS ALBERTO DE OLIVEIRA\*

### SUMÁRIO

Este artigo pretende propor uma abordagem para a linguagem natural (LN), abordagem esta baseada numa estrutura lingüística que pode ser considerada como um sistema de relações entre seus componentes (por exemplo, sons, palavras, frase, textos). É proposto neste artigo um sistema aprendiz que divida domínios de conhecimento em relações semânticas concernentes a mundos específicos para otimizar o processamento da LN. Tal sistema aplica-se especialmente a interfaces para sistemas especialistas.

### ABSTRACT

This paper intends to propose a treatment for natural language (LN) based on a linguistic structure which can be considered as a relation system between its components (i.e. sound, words, phrases, texts). An apprentice system to divide knowledge domains in semantic relations concerned with specific worlds is proposed in order to improve the NL processing. Such a proposition applies itself specially to interfaces to expert systems.

\* Licenciado em Letras (UERJ, 1976), Mestre em Lingüística (UnB, 1983), Doutorando de Computação Aplicada (INPE, 1986); pesquisador auxiliar, projeto Inteligência Artificial, processamento de linguagem natural; INPE, Departamento de Informática, Divisão de Computação e Sistemas. Caixa Postal 515 - 12201 - São José dos Campos - SP.

## INTRODUÇÃO

A grande maioria das linguagens (senão todas) podem ser expressas por uma linguagem natural (LN) ou traduzidas para ela. Porém, a recíproca não é verdadeira. Logo, por esta constatação e a título de início de discussão, a LN pode ser vista como uma espécie de amálgama de todas as linguagens por ser a mais rica das formas de expressão. Dessa forma, seus elementos componentes (palavras, frases, etc.) não podem ser tratados como representantes unívocos de algum referente em particular.

À primeira vista parece existir um obstáculo difícil de ser transposto no que concerne ao tratamento de textos em LN. Tal impressão é reforçada pelos ensinamentos pouco didáticos que na escola são passados aos alunos<sup>(1)</sup>. Dessa forma, a LN torna-se um instrumento de suplício para quem tenta analisá-la por métodos tradicionais ou computacionais, visto possuir um grande número de "exceções" e um outro número de significados, além da "falta de lógica" que geralmente dá a impressão de existir em seus enunciados.

No entanto, basta um pouco mais de reflexão para notar que não é desta forma que deve ser conduzida a análise das linguagens naturais. A gama de informações que um enunciado em LN transporta é de tal monta que se torna necessário situar o contexto da comunicação dentro dos limites do "assunto" abordado; este é o tema desta comunicação.

## AS RELAÇÕES SEMÂNTICAS<sup>(2)</sup>

As "palavras" não contêm um grande número de significados agregados a elas como fazem supor dos diversos "sentidos" que elas potencialmente possam vir a ter. Para cada "domínio" da comunicação há um e um só significado apontando para tal "palavra",

<sup>1</sup> Lembrar-se de exemplos clássicos, tais como manga, onça, dentre outros.

<sup>2</sup> "Para que uma análise automática da LN atinja os objetivos que o seu possível uso na Informática exige, é mister tomar os "inputs" como realmente eles o são: um texto. Se for considerado que o usuário potencial do código verbal é um sistema aberto e está em relação com os códigos e os usuários do macrossistema cultural, então está contido em cada evento verbal ("token") todo o universo cultural em que se insere esse utente. Se for considerado que a Língua é um sistema de signos verbais inserto no macrossistema dos códigos culturais e, como tal acha-se em relação a esses códigos e seus usuários, ter-se-á que cada "token" é um universo de relações. Relações estas, grosso modo, sintáticas, semânticas, pragmáticas e culturais, um truísmo. Ademais, a Língua é um instrumento de comunicação por excelência e, por isto, cada "token" apresenta, pelas múltiplas relações a que está sujeito, aspectos comunicativos que são válidos no contexto no qual ambos (interlocutor e "token") estão imersos". (OLIVEIRA, 1985)

conforme as relações lingüísticas geradas no interior desse ambiente comunicativo venham a determinar. Dessa forma - e aqui aparece como nova evidência - a definição de "domínios limitados" para a consecução de determinados sistemas demonstra que dadas relações sô se tornam consistentes e/ou transparentes dentro de limites prē-fixados.

O mesmo raciocínio se aplica ao processamento de linguagens naturais: quando há a criação de um vínculo comunicativo entre dois elementos, o primeiro elemento a ser determinado é o "contexto" em que foi tal evento se efetivarã. A partir daí, os significados das palavras já estão *a priori* definidos. Cada "domínio" gera um certo número de relações e estas fixam os significados que devem ser pesquisados quando se faz análise semântica de tal texto.

Cumprе salientar que tais relações não devem ser inseridas pelo pesquisador (ou analista), mas sim apreendida no diálogo com o usuário final. Isso decorre da necessidade de armazenar conhecimento lingüístico-semântico oriundo da vivência do usuário no trato com sua língua materna. Nesse afazer, ele invoca padrões de comportamento lingüístico que inconscientemente usa na expressão de seu conhecimento sobre o assunto abordado.

É imaginável que, para a manutenção de um processo comunicativo qualquer, os usuários de dada língua não tragam como vocabulário efetivo toda a gama de possibilidades combinatórias que a estrutura lingüística lhe permite<sup>(3)</sup>. Neste contexto serão abordados dois tipos de relações semânticas.

#### RELAÇÕES FORTES E RELAÇÕES FRACAS

As relações anteriormente citadas, além de outros tipos, podem ser descritas neste artigo, para efeito exemplificativo, como:

- (a) de ação                    (TOMAR, João, café);
- (b) de estado                (ESTAR, João, triste);
- (c) de definição            (SER, João, humano).

<sup>3</sup> " a) o Léxico é um sistema de armazenamento de informações relacionais;  
b) somente através dessas relações é que se poderá atribuir significados aos lexemas;  
c) cada lexema, instanciado por uma realização verbal, é em si mesmo um texto e pode traduzir o universo cultural que o gerou pelas relações que permite reconstruir;  
d) a lexia verbal, pelas categorias gramaticais que relaciona e congrega, é um núcleo de relações que podem reconstruir o estado inicial de um dado processo de comunicação." (OLIVEIRA, 1985).

quecerão em detrimento das outras. Por exemplo, se o assunto for *café da manhã*, serão fortalecidas as relações de PASSAR como UM PEDAÇO<sup>(6)</sup> e MANTEIGA, bem como as demais palavras que se relacionam com tal evento: pão, bolo, biscoito, etc.

Se, porém, ainda a título de exemplo, o assunto for *conservação e higiene do vestuário*, as relações com ROUPA e UM PEDAÇO terão maior grau de certeza de ser as verdadeiras, bem como com as palavras designativas de vestuário, ou de parte dele, comumente sujeitas a tal atividade.

Se, ainda, o assunto for *atividades escolares*, DE ANO e UM PEDAÇO serão as primeiras palavras a serem pesquisadas e "um pedaço" poderia ser "sofrimento" ou relacionar-se com "teste, prova, matéria, ano letivo, etc."

Assim, é desejável que tais relações sejam construídas através do diálogo com os usuários finais por um sistema-aprendiz, registrando-se qual o "domínio" no qual se insere o processo dialógico. Entre o usuário e o "domínio" também seriam estabelecidas as relações.

Em tal processo, numa visão sistêmica e após certo número de interações com usuários diferentes, seria abstraído das relações criadas o que é particular ao "domínio", o que fortaleceria tais relações em detrimento das demais. Da mesma forma, segundo o raciocínio análogo, relações entre "domínios" seriam tratadas para que se determinassem quais deles são subconjuntos de outros ou em que eles se interseccionam, criando-se novas relações.

Quanto ao fortalecimento ou enfraquecimento dessas relações, podem ser dispostos valores que, analogamente registrados por domínio-assunto, possam atribuir a cada uma delas, conforme já discutido, ênfase no grau de certeza com que atuarão na análise semântica de dado texto.

Um exemplo dessa abordagem dada a LN é o sistema-experimental TRADAUT ILNES, escrito em PROLOG, que objetiva interfacear sistemas especialistas, gerando regras de decisão para estes. O especialista, quando participa na elaboração de regras, nem sempre as redige com a eficiência desejada ou deixa escapar pormenores que podem ser de vital importância para o funcionamento geral do sistema: ele domina bem sua especialidade,

<sup>6</sup> A "palavra" PEDAÇO traduz a partição de um todo. No entanto, em:

- (h) eu quero um pedaço,
- (i) eu só fiz um pedaço,

percebe-se que está subentendido "pedaço de alguma coisa". O "domínio" onde transita o diálogo é que determina qual vai ser a coisa da qual se substraiu uma parte. Logo, tal palavra tem uma relação forçosa com alguma outra para que seu sentido seja completo.

porém pode não dominar bem as regras de expressão de sua língua materna. E é um fato que a qualidade da informação decresce à mesma proporção em que se dá maior atenção à escrita. No que concerne ao usuário leigo desse sistema especialista, o problema é o mesmo.

Nesse passo, o vocabulário de ambos, bem como as regras de estruturação sintática, são aceitos como verdadeiros através das relações que suscitam dentro da estrutura frasal. A partir daí, torna-se mais fácil apreender as regras que se quer operacionalizar ou armazenar, pois ambos (usuário e especialista) participaram inconscientemente da elaboração de relações lingüísticas que aquele "domínio" do conhecimento humano permite construir. A "semântica" será a mesma para ambos, embora variem entre eles o estilo, o conhecimento o grau de perícia no trato com a língua.

## CONCLUSÃO

São muitas as dificuldades para a implementação de um processador automático de textos em LN, visto que a orientação geralmente adotada para tal tipo de análise baseia-se em modelos formais que não levam em conta o conhecimento prévio e intuitivo que o usuário tem da língua que usa.

Daí decorrem afirmações que posicionam o usuário como um desconhecedor das regras de comunicação de sua própria língua. Na verdade o que ocorre com os erros e desobediência às regras que dado sistema adota e com o uso dado à língua é que:

1. as regras do sistema-processador são, na verdade, regras usadas pelo seu construtor e não podem representar a atuação geral de usuários concretos e diferenciados;
2. as possibilidades combinatórias que a estrutura de qualquer língua oferece são por demais amplas para que se possa determinar com precisão qual a construção sintática a ser usada e/ou qual o "sentido" que determinada frase ou texto estará veiculando.

Em Schank e Abelson (1977), faz-se menção ao caráter do conhecimento quanto à sua forma e ao seu conteúdo. Percebe-se que tal conhecimento está relacionado também com a individualidade do usuário e com o particionamento de domínios.

A staggering amount of knowledge about the world is available to human beings individually and collectively. Before we set out on a theory of knowledge systems, we ought to ask ourselves: knowledge about what? We must be wary of the possibility that knowledge in one domain may be organized according to principles different from knowledge in another. (p. 3)

Nesse contexto, propõe-se que quem realmente vai dialogar com a máquina deve supri-la com as regras que ela deve usar para poder haver um entendimento mais perto do desejável e, conseqüentemente, um aumento do grau de operacionalidade do sistema como

um todo. E um sistema-aprendiz poderá apreender as relações particulares de cada usuário, abstrair generalizações e construir partições que diminuam o alto grau de incerteza e de pesquisa quando se faz uma análise semântica.

#### BIBLIOGRAFIA

- OLIVEIRA, C.A. *Uma proposta de (re)classificação dos verbos com unidades geradoras de texto*. São José dos Campos, INPE, 1985. (INPE-3722-PRE-856).
- SCHANK, R.C.; ABELSON, R.P. *Scripts, plans, goals and understanding*. Hillsdale, NJ, Lawrence Erlbaum, 1977.