

# O PROCESSAMENTO DA LINGUAGEM NATURAL<sup>1</sup>

João Luís Garcia Rosa

Os usuários dos computadores permitem que esta máquina tenha acesso ao mundo externo, principalmente através do teclado, que é a forma mais simples de entrada de dados, muito embora não seja nada "humano", pois a pessoa tem de aprender a distribuição das teclas (por que?) e a datilografar. Outras formas de entrada são usadas também: *mouse*, disco, etc. Você já devem ter percebido que os computadores estão encolhendo cada vez mais. Mas o teclado (e o vídeo) não podem encolher muito mais do que já encolheram. Imagine um teclado minúsculo, onde para pressionar as teclas é necessário uma pinça (?). Não é tão absurdo, quando se considera que o teclado (e o vídeo, além de unidades de disco, etc.) são limitantes para a miniaturização do computador. Que tal, se ao invés de acionar teclas ou botões do *mouse*, pudéssemos entrar com comandos de voz. Não precisaríamos de teclado, e nem de vídeo, já que o computador também poderia responder através de uma voz artificial.

Já existem alguns "reconhecedores" e "sintetizadores" de voz. Mas, para chegar onde gostaríamos, eliminar por completo outras formas de entrada e saída locais, teríamos que ter, o que a Inteligência Artificial chama de Processamento de Linguagem Natural (PLN) ou Linguística Computacional, ou seja, fazer com que a máquina seja capaz de processar a língua natural que falamos no dia-a-dia, o Português, o Inglês, etc. E podem acreditar, é um trabalho extremamente complexo.

O PLN consiste de uma série de tarefas que a máquina deve executar para analisar um texto. Estas tarefas são, na maioria das vezes, interdependentes. O PLN pode ser definido de formas diferentes. Todas as definições incorporam a noção de armazenamento em computador e manipulação de dados lingüísticos. Entretanto, o ponto de discussão é o grau de sofisticação envolvido, que se traduz em uma porção de estruturas lingüísticas inerentes ao texto original, as quais o sistema pode detectar automaticamente, armazenar e manipular.

Existem várias abordagens para o PLN. E existem muitos trabalhos publicados nas várias abordagens. Porém a maioria dos trabalhos tratam de processamento da língua inglesa. Para se realizar um trabalho sério para a nossa língua há necessidade, portanto, da transposição de procedimentos e idéias para a língua portuguesa, que traz grande dificuldade no tratamento de alguns elementos da teoria da sintaxe como sujeito nulo, flexão do infinitivo, morfologia dos verbos, etc.

O processamento de linguagem natural é geralmente dividido em seis grandes áreas (Obermeier,1987): (1) interfaces em linguagem natural para bases de dados; (2) tradução de máquina - isto é, de uma linguagem natural para outra; (3) investigação minuciosa de texto, ou programas de indexação inteligentes para sumarização de grandes quantidades de textos; (4) geração de texto para produção automática de documentos padrões; (5) sistemas de fala para permitir interação de voz com computadores; (6) ferramentas para desenvolver sistemas de processamento de linguagem natural para aplicações específicas.

O problema central dos sistemas de processamento de linguagem natural é a transformação de uma sentença de entrada, potencialmente ambígua, em uma forma não ambígua que possa ser usada internamente por um sistema de computador. Estas representações internas variam, é claro, de uma aplicação para outra. A transposição de uma frase potencialmente ambígua para uma representação interna é conhecida como parsing (análise).

Convém salientar que o PLN é uma área multidisciplinar, envolvendo principalmente a Psicologia Cognitiva, a Computação e a Lingüística.

---

<sup>1</sup> Artigo publicado no *Caderno de Informática* do jornal *Diário do Povo* de Campinas, SP, em 14/09/95.

*Referência:*

Obermeier, K. K. (1987). "Natural Language Processing". Byte, pages 225-232, December.

João Luís Garcia Rosa  
Professor do Instituto de Informática – PUC-Campinas  
Mestre em Engenharia Elétrica - FEEC-Unicamp  
Doutor em Linguística Computacional - IEL-Unicamp