LINGÜISTAS E COMPUTADORES: QUE RELAÇÃO É ESSA?

Glaucia da Silva BRITO

Introdução

Explorar as possibilidades oferecidas pelos recursos computacionais ao processamento da linguagem humana tem sido um dos grandes desafios em nossos dias. Desde o advento do computador, esse assunto vem preocupando profissionais atuantes em vários campos do conhecimento, seja na informática ou ciência da computação, seja na ciência da informação ou em áreas de pesquisa específicas como a lingüística computacional, a análise documentária, a inteligência artificial, a tradução automática, a robótica, a engenharia do conhecimento etc.

Entretanto, até hoje não existe uma área integrada, academicamente estabelecida, na qual se concentrem os esforços referentes à implementação da linguagem natural em computador. As equipes de pesquisadores geralmente são formadas por engenheiros especialistas em informática e estes, quando querem construir um sistema no qual intervém a linguagem natural, recorrem quase sempre às vagas lembranças do ensino da gramática que têm de sua própria vida escolar ou consultam os manuais de gramática, onde geralmente os conteúdos são apresentados através de exercícios, numa perspectiva artificial.

Alguns procuram os lingüistas, pois consideram que estes possuem conhecimentos extremamente valiosos relacionados com a utilização da linguagem natural. Entretanto, tais conhecimentos, para serem utilizados com proveito em um programa computacional, devem situar-se em uma perspectiva que coincida com a da

informática, e isso nem sempre acontece, fazendo com que esta relação não dê bons resultados.

O objetivo deste trabalho é descortinar ao leitor um pouco do que vai por trás da implementação da linguagem natural em computador. Pretendemos dar uma pequena noção da vastidão deste campo e mostrar que os lingüistas podem e devem se envolver em projetos dessa natureza.

1. Computador e linguagem natural

A maneira mais comum de as pessoas se comunicarem é através da linguagem natural, seja pela fala, seja pela escrita. Dessas duas formas de comunicação, a mais comum é, sem dúvida, a fala. A escrita é relativamente nova; funcionalmente mais simples, já que menos dependente do contexto da enunciação e, portanto, do sistema de inferência próprio da percepção humana, é, por isso, mais fácil de ser implementada em computadores do que a linguagem falada. As relações entre a informática e a linguagem natural apresentam um panorama complexo, Coulon (1992) distingue duas perspectivas para estas relações:

1. O ser humano que se comunica com uma máquina necessita de uma linguagem. Em certos casos (não em todos), a linguagem natural é a que melhor convém ao usuário.
2. O ser humano, desde tempos remotos, tem refletido sobre sua própria linguagem. Essas reflexões foram absorvidas por certas disciplinas, cujas denominações mudaram conforme a época: gramática, filosofia, lingüística.

Atualmente, a informática vem contribuindo com novas perspectivas capazes de revolucionar o pensamento sobre a linguagem. As pesquisas que visam associar a informática ao processamento da linguagem natural remontam ao próprio advento

do computador. Coulon (1992) distingue quatro fases dessas pesquisas, que resumimos a seguir:

1ª Fase: estatístico-morfológica (1945-1955)

Os primeiros trabalhos (Weaver, 49-55) nos Estados Unidos e (Locke, 55) na Grã-Bretanha mostraram a exeqüibilidade de uma análise morfológica automática. Esses primeiros resultados permitiram prever que, com o auxílio de métodos estatísticos previamente comprovados em criptografia, chegava-se rapidamente a um nível semântico, suficiente para efetuar traduções com auxílio do computador.

2ª Fase: sintática (1955 - 1970)

Essa esperança desfez-se logo, pois tornou evidente a necessidade de considerar a estrutura da linguagem. Essa necessidade veio reforçar aquela corrente que procurava incentivar a introdução de métodos formais em lingüística. Isto resultou num período muito fecundo para ambas as disciplinas, entretanto, perdeu-se um pouco de vista o principal objetivo, pois muitos pesquisadores pensavam que, uma vez conhecida a estrutura sintática de uma sentença, a maior parte do trabalho já estaria realizada.

3ª Fase: lógica (1913 – até hoje)

A interpretação da estrutura sintática foi tentada primeiramente em termos lógicos ou, mais especificamente, pela lógica dos predicados de primeira ordem. Bons resultados surgiram rapidamente em diversos universos bastantes restritos: na interrogação de bancos de dados e no diálogo com robôs. O ponto culminante desta fase foi atingido com Winograd (70-72), que, ao dar um grande avanço ao desempenho qualitativo e quantitativo desses métodos, mostrou também claramente suas limitações. Trabalhos mais recentes com o PROLOG[1] vinculam-se a essa abordagem.

4ª Fase: cognitiva (1974- até hoje)

Apesar de suas inegáveis vantagens, a lógica de primeira ordem não foi bem adaptada ao tratamento de universos complexos. O centro de gravitação das pesquisas desloca-se pouco a pouco para a representação dos conhecimentos, campo que tem recebido uma contribuição maior dos psicólogos do que dos lingüistas.

Em todas estas fases, podemos perceber claramente que a maneira como o processamento da linguagem natural foi tratada (e ainda é) sustenta uma teoria computacional da linguagem ordinária. Tal teoria retrata o uso da linguagem ordinária somente como sujeito a um cálculo de regras que se separam da linguagem em uso real. Quanto a esta observação, poderíamos discorrer muito mais, porém este não é o nosso objetivo.

2. Informática e lingüística

Na visão de um especialista em informática, um único contra-exemplo, em lingüística, basta para fazer cair uma teoria, pois revela um erro na descrição formal da língua. Entretanto, isto não o incomoda.

Na informática, não se é obrigado a respeitar essa exigência: a tarefa relativa à linguagem é original, não sendo necessário justificar os métodos nem no plano formal, nem no da conformidade com os resultados experimentais. O único gabarito segundo o qual o especialista em informática aceita ser julgado é o da eficácia (rapidez, dimensão, portabilidade, capacidade de extensão) dos algoritmos utilizados para resolver problemas concretos. Naturalmente, modelos e hipóteses antecipados por lingüistas são úteis, mas, segundo os especialistas em informática, só podem ser integrados se mantiverem sua independência.

Informática + lingüística = lingüística computacional. A lingüística computacional é a área de pesquisa multidisciplinar que

estuda o processamento sintático, semântico e lógico da linguagem natural. Foi o advento da inteligência artificial que conduziu a lingüística computacional de uma obscuridade relativa nas décadas de 50 e 60, quando realizava principalmente exercícios computacionais com textos, a uma proeminente posição nas disciplinas da ciência da computação e da lingüística.

A inteligência artificial (IA) constitui-se em um conjunto de técnicas de programação para resolver determinados tipos de problemas em informática. Ela procura imitar, através dos programas que comandam máquinas, as formas de resolução de problemas do mesmo modo que o homem o faz. É claro que tais programas são, essencialmente, algoritmos; eles não são inteligentes. O termo IA foi proposto em 1956 por John MacCarthy. Apesar de várias discussões sobre a substituição deste nome por outros (tais como processamento complexo da informação e simulação de processos cognitivos), essas tentativas não tiveram êxito.

Assim como a IA, a lingüística computacional é motivada pelo desejo de fazer uma tecnologia melhor e de entender melhor os processos humanos de comunicação. Allen (1987:01) diz que um objetivo geral da lingüística computacional seria *especificar uma teoria da compreensão e da produção de linguagem num tal nível de minúcia que se possa escrever um programa de computador que entenda e produza linguagem natural.*

3. Processamento da Linguagem Natural (PLN)

O PLN é um campo de extrema complexidade, em que devemos considerar diversos tipos de conhecimentos e agrupar competências variadas e complementares. O objetivo maior da PNL é tratar a língua de maneira automática através de formalismos que explicitem os conhecimentos lingüísticos, tornando-os passíveis de serem tratados por computador. É comum em lingüística considerar-se cinco níveis em relação à linguagem:

1. *nível morfológico*: permite reconhecer as palavras sob as diferentes formas (conjugação, declinação, etc.) que sua função na sentença lhes dá;
2. *nível lexical*: faz corresponder a palavra, uma vez reconhecida, às informações de que se dispõe sobre a mesma;
3. *nível sintático*: trata da disposição das palavras numa sentença;
4. *nível semântico*: estabelece correspondência entre situações do mundo real ou mundo possível e as estruturas reconhecidas ao nível sintático;
5. *nível pragmático*: interpreta essas situações no contexto mais geral de uma troca de informações.

Os dois primeiros níveis não oferecem grandes dificuldades para a informática, onde se criam facilmente instrumentos adequados. Este, porém, não é o caso dos níveis superiores (sintaxe, semântica e pragmática), que utilizam numerosos modelos de processamento. Falaremos, resumidamente, de dois tipos de processamento - o sintático e o semântico, consideraremos a lógica que, em lingüística computacional, é sempre utilizada para representar a significação e exploraremos um pouco a lingüística de *corpus*, que se contrasta fortemente com as abordagens de base lógica e tem sido intensivamente investigada e utilizada da metade dos anos oitenta para cá.

3.1 Processamento sintático

A computação de uma estrutura sintática baseia-se na especificação formal de uma gramática que fornece as estruturas "legais" numa linguagem; e na técnica de análise empregada, que fornece o método para a análise de uma sentença, para resolver a sua estrutura como é apresentada pela gramática. Simplificando, o objetivo do processamento sintático é usar a análise morfológica e léxica de uma sentença para derivar a estrutura sintática; as ferramentas usadas para derivar essa estrutura são um formalismo

gramatical e um analisador (*parser*).

Neste processamento, dois tipos de formalismos gramaticais são utilizados. O primeiro deles é uma gramática independente do contexto (*context-free*), na qual as regras assumem a seguinte forma: "*<símbolo>*$\leftarrow$*<símbolo>*$_1$...*<símbolo>*$_n$". As regras envolvidas, *regras de reescrita*, fornecem as estruturas "legais" da linguagem: como uma sentença pode ser construída por meio da combinação legal de suas subpartes e como essas subpartes podem ser "legalmente" produzidas. Uma representação do *output* de um analisador é uma estrutura em forma de árvore.

O outro formalismo simples usado no processamento da linguagem natural é uma *rede de transição recursiva*. Esta rede representa a estrutura de uma sentença em termos de nós rotulados (*labelled nodes*) conectados por arcos de rótulo (*label arcs*). As redes de transição devem ser cruzadas encontrando-se percursos ou caminhos através delas, e esses caminhos são as seqüências de arcos rotulados. Isso pode ser representado em um diagrama da seguinte forma:

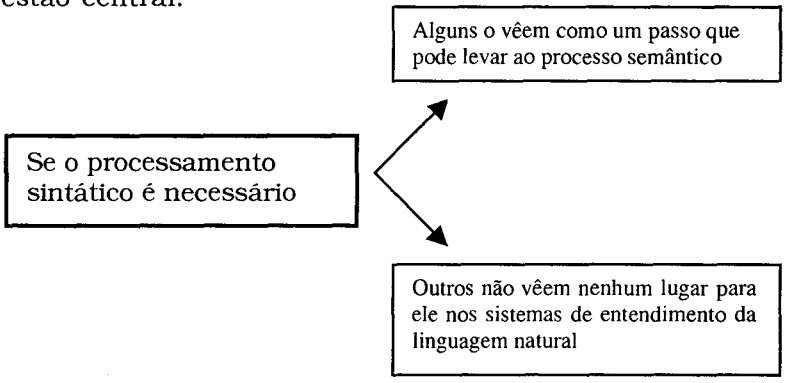
	entrar		SN		SV		sair	
S: (a)	$\rightarrow$	(b)	$\rightarrow$	(c)	$\rightarrow$	(d)	$\rightarrow$	(e)

As gramáticas independentes do contexto e as redes de transição são formalismos simples. No processamento da linguagem natural, as gramáticas apresentam dificuldades inerentes, por exemplo, são incapazes de captar certos fenômenos sintáticos, como a concordância e o fenômeno da ambigüidade também é muito problemático. O processamento sintático é o centro de muitos terminais de linguagem natural desenvolvidos para a integração de bancos de dados. A formulação da linguagem em termos da construção de gramáticas restritas e de suas regras associadas provou ser capaz de produzir ferramentas eficientes, por meio das quais lingüistas computacionais podem

desenvolver programas viáveis. Segundo Graham (1998:212), no processamento sintático a formulação da linguagem não é problemática; é meramente um método computacional para facilitar o desenvolvimento de sistemas de computadores.

3.2 Processamento semântico

Existe um debate muito intenso na lingüística computacional sobre a relação entre o processamento sintático e o processamento semântico (os primeiros a debaterem esta relação foram Sparck - Jones & Wilks, 1985). Embora o processamento sintático tenha sido muito mais estudado, alega-se que o processamento semântico da linguagem natural é que é o real objetivo do entendimento da linguagem natural em computadores. Temos duas visões que giram em torno da seguinte questão central:



Uma das maneiras como é desenvolvida essa representação da estrutura semântica no processamento da linguagem natural dá-se por meio de *gramáticas semânticas*, como as que podem ser encontradas nos sistemas LADDER e SOPHIE (Burton, 1976). As gramáticas semânticas combinam o conhecimento sintático e semântico num único sistema de regras. No entanto, os termos da gramática não são categorias sintáticas, mas sim o que é chamado de "elementos de significação".

As gramáticas semânticas, assim como os analisadores baseados em busca, não visam a representar a atividade humana, mas sim apoiar o processamento de uma máquina, ou seja,

> [...] não visam a ser parte de uma teoria da linguagem natural. Todavia, uma maneira importante como a semântica é tratada na lingüística computacional, a teoria da dependência conceitual, está muitíssimo interessada numa teoria da linguagem, na medida em que pretende representar a maneira como os seres humanos processam a linguagem.
>
> (Graham, 1998:216)

3.3 Lógica e semântica

No processamento semântico utiliza-se muito a teoria da dependência conceitual. Esta recorre à idéia elaborada por Kats & Fodor [2], dos primitivos semânticos (*semantic primitives),* que postula que a significação pode ser representada em termos da configuração de um conjunto de aspectos semânticos. Nesta teoria, é necessário especificar *inferências* que podem ser feitas com relação a uma sentença. Para poder analisar as possibilidades inferenciais de uma sentença, alguns estudiosos do entendimento computacional da linguagem natural recorreram à lógica. Ela também foi mencionada em tentativas de representar o significado de sentenças. No processamento da linguagem natural, a lógica dos predicados de primeira ordem continua sendo o formalismo matemático da representação da significação mais utilizado.

A lógica dos predicados de primeira ordem usa um conjunto de símbolos e regras para combiná-los em termos, em sentenças e em regras de inferência que especificam as possibilidades de extensão das sentenças. Um sistema lógico inicia-se com um axioma (um conjunto inicial de sentenças), e qualquer sentença que possa ser dele derivada usando as regras de inferência pode ser um teorema do sistema. Os símbolos do sistema lógico ganham significação pela

associação dos termos e dos símbolos com objetos tirados de um dado modelo; assim, as sentenças a que o sistema é aplicado se tornam declarações sobre os objetos e de como os objetos se relacionam. Na seqüência, os símbolos do sistema lógico são acoplados a objetos do modelo, e as regras de interpretação determinam como o significado de uma sentença é o produto das condições de verdade que se aplicam aos seus componentes constitutivos: são eles verdadeiros ou falsos? Uma estrutura semântica pode então ser produzida, por exemplo, ligando uma "sentença lógica" a uma dada sentença da linguagem natural.

O significado de uma sentença da linguagem natural é, então, apresentado em termos de um formalismo matemático. São exemplos dos símbolos os "conectivos lógicos": *e* ($\wedge$), *ou* (V), *não* ($\neg$), *implica* ($\rightarrow$); e os "quantificadores": *para todo* ($\forall$), existe ($\exists$). Por exemplo, podemos representar o significado de uma sentença associando uma formulação a ela:

João gosta de Mac. = G(j,m)
João é um lingüista. = L(j)
Mac é um computador. = C(m)
Todos os lingüistas gostam de computador. = $(\forall l)(\forall c)/G(l,c)$

Certamente, a lógica é uma característica importante da computação, sob muitos aspectos. Nada, ou quase nada, é feito em programas de inteligência artificial sem que a lógica seja pelo menos citada. Algumas dessas lógicas são: dos predicados, que especificamos um pouco acima, das proposições, temporal, 3-valores, multivalores, nebulosa ou difusa (*fuzzy*), modal, dinâmica e intuicionista. Excluindo a lógica das proposições e a dos predicados, as outras são, em geral, referidas como lógicas não-padrão; sua aplicação não tem caráter definitivo, sendo, pelo contrário, ainda bastante controvertidas. Uma das razões mais argumentadas para a utilização da lógica, principalmente em programas desenvolvidos com o PROLOG, é o fato

de que ela fornece um modo de especificar as inferências que podem ser feitas em relação às expressões de linguagem natural e, assim, *parece oferecer aos cientistas computacionais a oportunidade de poderem equipar um sistema com o 'conhecimento' necessário exigido por ele para processar uma sentença.*(Graham, 1998:203)

3.4 Lingüística de *corpus*

O campo da PNL agrupa uma série de atividades: processamento da fala; gestão eletrônica de documentos; geração automática de textos; tradução auxiliada por computador; tradução automática; aprendizado de máquina, etc. Estas atividades estão fortemente relacionadas com a lingüística de *corpus*. Este termo, segundo Leech (1991), esteve muito limitado até a publicação de um livro editado por Aarts and Meijs (1984). Para Leech

> Since that time, I have encountered the term "*corpus* linguistics" with growing frequency, but only significantly so in the last one or two years, when it is at last making a bid for being a recognised branch of linguistics, just as (for example) sociolinguistics, psycholinguistics, and text linguistics did in the 60s and 70s. But *is corpus* linguistics really comparable with these other hyphenated branches of linguistics? No, because " *corpus* linguistics " refers not to a domain of study, but rather to a methodological basis for pursuing linguistic research. In principle (and often in practice) *corpus* linguistics combines easily with other branches of linguistics: we can study phonetics, syntax, sociolinguistics, and any other aspect of linguistics by means of corpora, and when we are doing this we can be said to be combining techniques of *corpus* linguistics with the subject - matter of phonetics, syntax, sociolinguistics, and so on. (1991:105)

Portanto, a lingüística de *corpus* se articula com outras áreas da lingüística, pois utilizando suas técnicas podemos estudar fonética, sintaxe, sociolingüística e assim por diante. Essas técnicas com certeza puderam ser melhor elaboradas a partir do advento do

computador, pois este passou a permitir uma manipulação de milhões de palavras. Podemos, assim, trabalhar com modelos estatísticos complexos, tirando conclusões que podem ser a chave para um bom desempenho de um programa, em que o processamento da linguagem natural seja essencial.

Hoje, é possível encontrarmos *corpus* com bilhões de palavras disponíveis em modo eletrônico graças aos esforços da Associação para Lingüística Computacional, com iniciativas como: Coleção de Dados (ACL / DCI), a Iniciativa de *Corpus* Europeu (ECI), ICAME, o *Corpus* Nacional Britânico (BNC), o Consórcio ou Associação de Dados lingüístico (LDC), o Consórcio ou Associação para Pesquisa Léxica (CLR), Pesquisa de Dicionário Eletrônico (CLR), e esforços de padronização como a Iniciativa de Codificação de Textos (TEI) e muitos outros.

A maioria dos projetos que trabalham com o processamento da linguagem natural, aliada a uma técnica de *corpus*, tem como objetivo criar sistemas informáticos que conheçam e reconheçam a língua natural. Isto poderá permitir uma melhoria significativa da comunicação homem - máquina e da comunicação humana com o auxílio do computador.

Considerações finais

Tentamos até aqui descortinar ao leitor um pouco do que vai por trás da implementação da linguagem natural em computador. Tentamos, também, dar uma pequena noção da vastidão deste campo. Ainda não argumentamos, porém, em relação aos lingüistas.

Os especialistas em informática às vezes se consideram como tradutores mais ou menos fiéis do pensamento dos lingüistas, procurando colocar todos os recursos computacionais a serviço de um certo modelo de um nível da língua (por exemplo, da sintaxe); *se investem da missão de descobrir, por conta própria, recursos integrados, com o objetivo de abranger o conjunto dos fenômenos da linguagem natural.*(Coulon,

1992:66). Na biblioteca podemos encontrar trabalhos puramente teóricos ou alguns que têm uma finalidade prática, direcionada a um objetivo específico, mas nas suas equipes, nenhum lingüista, nas suas bibliografias, muitos manuais de gramáticas formais.

Não é difícil escutarmos o seguinte: *Lingüistas na equipe de produção só atrasam o trabalho.* É isto que falam alguns coordenadores de equipes que trabalham com produção de programas de linguagem natural. Mas, que *lingüista* eles não querem na sua equipe? Aquele que hoje, muitas vezes, se dá por satisfeito somente com a educação acadêmica ou profissionalizante que recebeu, que se dedica a estudar a linguagem natural apenas por um foco específico e se recusa a participar de discussões mais acirradas que não sejam da sua área de especialização.

Qual deverá ser o perfil de um lingüista para poder compor uma equipe de produção de programas? Qual seria o seu papel na produção de um programa de linguagem natural? O que um programa de linguagem natural deve retratar?

Comecemos pela última questão. Rabuske (1990:134) diz que:

> As implementações de linguagem natural devem ser escritas em função de seus objetivos. Quando servem como meio de recuperar informações, é importante que não haja interpretação dúbia para as sentenças, pois neste caso a informação de retorno poderia ser pouco confiável ou inútil. Quando tiverem cunho didático, certamente os aspectos gramaticais, a correção e a precisão deverão prevalecer. Quando tiverem um apelo popular, erros cometidos pelo povo certamente poderão ser mais facilmente tolerados.

Partindo da citação de Rabuske, podemos considerar que um lingüista que não prescinda da ciência nem da tecnologia será muito bem vindo numa equipe de produção de programas computacionais. Deverá ser aquele que também saberá agir e não apenas reagir; planejar e não apenas executar; tomar decisões e integrar conhecimentos.

Para poder integrar os conhecimentos linguagem natural + informática não é necessário que este lingüista estude programação, mas que possa visualizar o processamento que tanto conhece na teoria,

rodando numa máquina. É necessário também que o lingüista conheça um pouco das pesquisas e sistemas já em funcionamento nos campos de: geração de textos; documentação automática; tradução automática; processamento de textos; robótica; ensino assistido por computador etc. O número de sistemas desenvolvidos em linguagem natural é muito grande, principalmente na área da inteligência artificial. Outro papel importante do lingüista que resolver se aventurar nesta área deve ser o de tentar correlacionar teorias atuais da linguagem numa perspectiva que coincida com a da informática.

Por exemplo, a maioria dos especialistas em inteligência artificial utiliza um modelo inferencial de comunicação que pode ser assim especificado:

1. O princípio do mínimo esforço: usar, apenas, os recursos necessários para uma comunicação completa e adequada. Simplicidade e ausência de exageros dão uma boa tônica à comunicação escrita ou falada. O assunto ou aspecto que não for relevante não deverá ter vez.
2. O princípio da qualidade: requer que a comunicação seja baseada na segurança e confiabilidade do que é transmitido, tendo sido estabelecidos fundamentos que assegurem isto.

Podemos perceber que esse modelo tem como ponto de partida as inferências desenvolvidas por Grice (1975), que, segundo os especialistas, devem ser preservados convenientemente numa implementação em computador. Outro modelo de comunicação muito utilizado em processamento eletrônico da linguagem natural é o da codificação-decodificação.

Nesta perspectiva caberia ao lingüista perguntar-se: que modelo mais atual de comunicação poderia ser utilizado para desenvolver programas de computadores em inteligência artificial? Será que o modelo da teoria da relevância desenvolvido por Sperber e Wilson (1995) pode ser adaptado numa perspectiva que coincida com a da informática?

Outro exemplo seria o lingüista rever o paradigma chomskiano, que foi muito utilizado na década de 50 e ainda hoje é utilizado por

pesquisadores da área da Inteligência Artificial e de Redes Neurais. Leech, em seu trabalho *Corpora and theories of linguistic performance*, apresenta quatro características da lingüística de *corpus* fazendo um contraste com este paradigma. Seriam elas: enfoque no desempenho lingüístico, em lugar de competência; enfoque na descrição lingüística, em lugar de lingüística universal; enfoque quantitativo, bem como nos modelos qualitativos de linguagem; enfoque num ponto de vista de investigação científica mais empírico, ao invés de racionalista. Caberia a um lingüista explorar de maneira mais efetiva essas características e aproximá-las de estudos na área da IA e de Redes Neurais.

Partindo do princípio que o lingüista sabe como funciona a língua e o especialista em informática sabe como funciona o computador, formou-se, na UFSC, uma equipe interdisciplinar composta por lingüistas e especialistas de informática [3]. Num primeiro momento, a proposta do grupo era desenvolver um pequeno etiquetador. Seria simples se se pegasse um manual gramatical e, a partir dele, as regras fossem criadas e implementadas em uma linguagem de programação qualquer. Problema resolvido? Não. Então surgiu uma nova questão: para que serviria o etiquetador? E uma nova área de discussão foi aliada ao projeto: a informática educativa. Essa área considera que construir uma ferramenta só para descrever um fenômeno sem uma utilização efetiva não serve.

Resumindo, no final de todas essas discussões interdisciplinares, desenvolvemos um protótipo que denominamos *ClassEstar*. Foi implementado com a linguagem Prolog para ser utilizado pelo professor de língua portuguesa seja em sala de aula, num trabalho coletivo, seja na dinâmica do trabalho desenvolvido em um ambiente informatizado. Partimos de um *corpus* lingüístico e não de regras fixas tradicionais para desenvolvimento deste protótipo, procurando manter o papel que a linguagem desempenha na vida social e na conduta prática do ser humano. O *ClassEstar* está na sua versão 2.0. Devido ao potencial deste ambiente e sua aceitação (pois este foi apresentado em diversos momentos para professores do Ensino Fundamental) a equipe

pesquisadora decidiu continuar implementando as sugestões colhidas ao longo do seu desenvolvimento. Este é apenas um pequeno exemplo de um trabalho realizado por uma equipe interdisciplinar, em que as discussões permearam diversas áreas do conhecimento e em que lingüistas e especialistas de informática se integraram e, com certeza, aprenderam muito.

O estudo da linguagem natural com o computador é tão antigo quanto o próprio computador. Despertou muitas esperanças, seguidas de muitas decepções. Está na hora de os lingüistas participarem destas esperanças e decepções; há muito o que resolver e há muitas indagações ainda sem respostas. Cabe aos lingüistas tentar participar de equipes que produzem programas de linguagem natural e assumir o espaço que compete a eles. Não queremos afirmar que, desta forma, a qualidade dos programas será crescente, mas temos que tentar, pois existe um déficit muito grande entre as pretensões programáticas feitas principalmente em favor da inteligência artificial e os sistemas reais que foram produzidos. Um trabalho neste enfoque exigirá, dos profissionais participantes, uma busca constante de novos conhecimentos, uma postura diferente e inovadora.

Referências bibliográficas

ALLEN, J. (1987) *Natural language understanding.* Menlo Park, California: Benjamin, Aummings Publ. Comp.

BRITO, G. Da S.; MENGARDA, E. J. & RODRIGUES FILHO, I. W. (2000) *O processamento da linguagem natural baseado em corpus lingüístico para desenvolvimento de softwares educacionais.* Trabalho apresentado para a Disciplina de Lingüística Computacional do Programa de Pós-Graduação em Lingüística da UFSC.

CLOCKSIN, H. F. & MELLISH, C. S. (1984) *Programming in Prolog.* Berlim: Spring-Verlag.

COULON, D. (1992) *Informática e linguagem natural: uma visão geral dos métodos de interpretação de textos escritos.* Brasília: IBCT; Rio de Janeiro: Senai.

RODRIGUES FILHO, I. W. (1999) *Prolog.* Anotações de aula, UFSC.

GRAHAM BUTTON *et al.* (1998) *Computadores, mentes e conduta.* Roberto Leal Ferreira (trad.). São Paulo: Fundação Editora da UNESP.

LAGE, N. (1999) *Semântica.* Anotações de aula, UFSC.

LEECH, G. (1992) Corpora and theories of linguistic performance. In *Directions in Corpus Linguistcs Proceedings of Novel Symposium* 82 Stockholm, 4-8 august 1991. Berlim: Ed. Jan Svartrik, Norton de Gruyter.

PENNINGTON, M. C. & STEVENS V. (1994) Computers in Appied Linguistics. In *Internacional Perspective.* Clevedon/UK: Multilingual Matters LTD.

RABUSKE, M. A (1990) *Linguagem natural.* Trabalho final do XVII curso avançado em sistemas de computação - CASC 17, IBM do Brasil, Rio de Janeiro.

RABUSKE, R. A. (1995) *Inteligência Artificial.* Florianópolis: Editora da UFSC.

ROCHA, M. (1999) *Lingüística computacional.* Anotações de aula, UFSC.

SINCLAIR, J. M. (1992) The automatic analysis of corpora. In *Directions in Corpus Linguistcs Proceedings of Novel Symposium* 82 Stockholm, 4-8 august 1991. Berlim: Ed. Jan Svartrik, Norton de Gruyter.

SPERBER, D. & WILSON, D. (1995) *Relevance: communication and cognition.* 2ª ed. Cambridge, MA: Harvard University Press.

Notas

[1] Dois pesquisadores, Kowalski e Hayes, foram os que exploraram inicialmente a idéia de utilizar a Lógica com um formalismo executável em computadores. Kowalski foi quem cunhou a expressão "Logic programming" ou seja "Programação em Lógica". Em 1973, foi escrito o primeiro interpretador Prolog, na Universidade de Marseille, pelo grupo de Inteligência Artificial coordenado por Alain Colmerauer.

[2] The Structure of a Semantic Theory. *Language, v. 39,* n.2, 1963.

[3] Fazem parte dessa equipe: Gláucia da Silva Brito, Elias Mengarda e Ilson Wilmar Rodrigues Filho.