

ISSN 0101-3335

LETRAS DE HOJE

Nº 82

DEZEMBRO DE 1990

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
Curso de Pós-Graduação em Linguística e Letras
Centro de Estudos de Língua Portuguesa

Letras de Hoje
estudos e debates de
assuntos de lingüística,
literatura e língua
portuguesa

Número monográfico dedicado à
LINGÜÍSTICA COMPUTACIONAL
UM APELO À INTERDISCIPLINARIDADE

organizado pelo
Centro de Pesquisas Lingüísticas

EXPEDIENTE

LETRAS DE HOJE
Fundada em 1967

Administração:
Avenida Ipiranga, 6681
Caixa Postal 1429
90620 Porto Alegre - RS - Brasil

Curso de Pós-Graduação em Lingüística e Letras / Centro de Estudos da Língua Portuguesa em convênio com o Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq e FAPERGS.

Diretor:
Prof. Ir. Elvo Clemente

Assessora Editorial:
Maria Eunice Moreira

Composição e Arte Final:
GRAFLINE
Assessoria Gráf. e Edit. Ltda.

Impressão:
Gráfica EPECÊ

Conselho Editorial:

Para assuntos lingüísticos: Augustinho Staub, José Marcelino Poersch, Leonor Scliar Cabral, Leci Borges Barbisan, Feryal Yavas e Mehmet Yavas.

Para assuntos literários: Gilberto Mendonça Teles, Heda Maciel Caminha, Petrona Dominguez de Rodrigues Pasquês e Regina Zilberman.

Para assuntos interdisciplinares: Ignacio Antônio Neis e Urbano Zilles.

A Revista aceita contribuição de sua especialização.

Os originais enviados à Revista não serão devolvidos, mesmo que não sejam utilizados.

A revista aceita trocas.
On demande l'échange.
We ask exchange

Preço da assinatura:
- 4 números anuais:
Brasil: 10BTN
Exterior: US\$ 10
- Número avulso: 3BTN

Os pagamentos podem ser feitos por cheques bancários ou através de vale postal em favor da EDIPUCRS.

Av. Ipiranga, 6681 - Prédio 33
Caixa Postal 12001
90620 - Porto Alegre - RS - Brasil

EDIPUCRS

Sumário

José Marcelino Poersch - Apresentação	5
Bertilo Frederico Becker - Conceitos de Inteligência Artificial aplicados a Lingüística	15
Vera Lúcia Strube de Lima - Tratamento Automatizado da Língua Natural	41
Gabriela Ardisson Mattos - A Língua Portuguesa no Projeto Eurotra	57
Alejandra Bentolla - O Professor de Línguas confrontado com o Minotauro Informático... Por que pensar necessariamente em Skinner?!	75
Margarete Axt - Um Projeto em Linguagem LOGO para eliciar a Produção de Estruturas Relativas em Portugêses	85
Maria Teresa Rijs da Fonseca Lino - Banco de Neologismos do Português Contemporâneo - Balanço de uma Experiência	103
José Marcelino Poersch - O que a Lingüística tem a ver com o teclado de microcomputadores	117
RESENHAS	171

Apresentação

Lingüística e Computação: um estreitamento de laços

José Marcelino Poersch

Departamento de Lingüística – PUCRS

1. A ESPECIALIZAÇÃO, UM MAL NECESSÁRIO

“É preciso que os homens saibam que (nos tempos modernos) eles não podem nunca parar de aprender. Se você não começar logo em seguida (à obtenção de um diploma, mesmo que seja de uma instituição renomada) e se pôr a par dos novos conhecimentos, ou a aprender a utilizar os novos instrumentos, seu diploma não vale mais nada. É preciso aprender durante toda a vida, porque a revolução científica não vai parar”. Se essas palavras de Servan-Schreiber (1985, pág. 4) são verdadeiras para qualquer ramo da ciência, com muito maior razão o são para a área da ciência lingüística a qual, a cada momento, encontra novos horizontes de investigação e de aplicação, constituindo-se numa das ciências mais produtivas da atualidade. E tanto isto é verdade que basta um afastamento de poucos anos para que o estudioso se sinta desatualizado.

Neste avanço contínuo, o saber tem que tornar-se cada vez mais especializado. Um claro exemplo disso é a atual caminhada que a Lingüística enceta, lado a lado, com a Informática. Haja visto os últimos Institutos Lingüísticos desenvolvidos pela Sociedade Americana de Lingüística (LSA); uma grande parte das disciplinas oferecidas exigem a manipulação do microcomputador, em diversas áreas, como fonologia, morfologia, semântica, sintaxe, psicolingüística, teoria da variação, linguagens de Inteligência Artificial. Considerem-se, entre tantas, as seguintes disciplinas: Semântica Computacional, Estatística

para análise de dados, Experimentos Psicolinguísticos com uso de computador, LISP para lingüistas, Modelos formais e computacionais para aquisição infantil da gramática, Abordagens contextuais e computacionais no desenvolvimento da linguagem, Computadores na tradução, Seminário em morfologia computacional, Modelos computacionais de aquisição da linguagem, Implementação do computador na análise e síntese da voz e PROLOG para lingüistas.

Com base nesse programa, pode-se perceber que a Lingüística Computacional avança com passos de gigante, desbrava novas trilhas, fornece contribuições importantes à Informática.

Dentro dessa realidade, foram estabelecidos contatos interdisciplinares com estudiosos de áreas limítrofes (Computação, Estatística, Didática de Línguas e Lingüística) com os quais planejou-se, em 1986, a criação do Centro Brasileiro de Lingüística Computacional e com os quais elaborou-se um projeto interdisciplinar de pesquisa no campo da Lingüística Computacional. Uma das metas prioritárias deste projeto é oferecer contribuições aos cientistas da computação para a elaboração de programas básicos usados em editores de textos e colaborar no desenvolvimento de estudos da Lingüística Quantitativa para a análise automática de textos, quer se trate de separadores silábicos, de ortografadores ou de indexadores. O projeto global prevê, em suas fases iniciais, o levantamento automático da frequência de caracteres gráficos (isolados ou combinados), o estabelecimento de um separador silábico automático que facilite o emparelhamento dos textos em seu lado direito, e a montagem de um ortografador que corrija automaticamente um texto editado.

Acredita-se que compete ao lingüista conquistar espaços na área da informática para efetuar contribuições ou realizar aplicações. Revistas especializadas apontam para várias destas aplicações, classificadas segundo os diversos níveis da análise lingüística. Assim, a Fonética e Fonologia contribui na elaboração de programas de síntese e de reconhecimento da voz humana, bem como na determinação de regras de separação silábica automática para editoração de textos. À lexicografia compete fornecer listas de frequência de palavras para a organização de dicionários bilingües e a programação de "ortografadores" (spellers). A morfossintaxe e a semântica desempenham papel importante nos programas de tradução automática, no processamento de textos e no tratamento de linguagens naturais. Isso sem falar do papel preponderante do qual os lingüistas foram

incubidos, junto a grupos de investigação das Ciências Cognitivas e da Inteligência Artificial, no desenvolvimento das teorias sobre o processamento da informação lingüística, tanto no aspecto receptivo quanto produtivo.

2. ASPECTOS DA LINGÜÍSTICA COMPUTACIONAL

2.1. Conceituação

Conceituamos Lingüística Computacional (LC), no seu sentido mais amplo, como área interdisciplinar da ciência da linguagem apresentando um tríplice propósito:

- a. servir-se do computador como instrumento de trabalho no intuito de processar e analisar seus dados, editar seus textos e controlar seus experimentos;
- b. fornecer subsídios à informática para a obtenção de software básico;
- c. incentivar uma colaboração mútua com a ciência da computação visando ao progresso dos diversos projetos da Inteligência Artificial.

Da lingüística, a LC explora tanto o aspecto quantitativo quanto o aspecto qualitativo. No aspecto quantitativo, onde pontifica a Lingüística Estatística, usa-se o computador para o processamento e análise dos dados. "A determinação da extensão daquilo a que o falante está preso pelo código lingüístico e, contrariamente, a extensão daquilo que lhe dá liberdade (pelo que ele é livre) e pelo qual ele pode ser original, é a essência daquilo que se chama de Lingüística quantitativa" (Herdan, 1966: 5 e 6). No aspecto qualitativo, algébrico, de formalização da linguagem, a lingüística fornece, ao cientista da computação, elementos que possibilitam a elaboração de linguagens intermediárias, elementos que permitem a segmentação (parsing) tão necessária à análise automática de textos. A linguagem natural é analisada sob o prisma da Lógica Simbólica (Alwood, 1977), em seus ramos da teoria dos conjuntos, do cálculo sentencial, da lógica inferencial (argumentos e inferências) e do cálculo funcional.

2.2. O computador a serviço da lingüística

A LC, num primeiro momento, procura utilizar-se de um computador para processar e analisar estatisticamente os dados quantitativos da língua. Isto realmente constitui a aproximação mais antiga entre a lingüística e a ciência da computação. Com o auxílio do computador, o lingüista consegue:

- a. executar automaticamente o levantamento da freqüência de ocorrência dos diferentes elementos e unidades lingüísticas de um texto (e de acordo com os diversos níveis): letras, sílabas, afixos, palavras, sintagmas, categorias gramaticais e, até, frases, com o objetivo de, por exemplo, organizar o vocabulário fundamental ou efetuar a análise estatística de textos de um determinado autor ou de uma determinada época;
- b. armazenar e recuperar informação lingüística;
- c. extrair dados bibliográficos para um determinado artigo ou determinada obra;
- d. implementar pesquisas psicolingüísticas relacionadas principalmente com os processos de recepção, como Tarefas de Decisão e de Acesso Lexical (*Lexical access and lexical decision tasks*) relatados em diversos números do *Journal of Memory and Language* de 1985 (Hudson and Bergman 1985, Gordon, 1985), com a leiturabilidade de textos e com medidas de maturidade lingüística;
- e. analisar estatisticamente quaisquer dados de lingüística quantitativa, principalmente da variação lingüística, visto já existirem programas gerais de estatística para a lingüística e programas especiais para a variação lingüística (SANKOFF & CEDERGREN, 1981) adaptados a microcomputadores;
- f. usar um editor de textos para a produção de monografias, de relatórios, de obras literárias e de material de ensino (incluindo polígrafos e testes);
- g. programar exercícios para fomentar a rapidez e melhorar o nível de compreensão de textos.

2.3. A lingüística colabora com o cientista computacional

Numa segunda instância, a LC persegue objetivos mais ambiciosos quais sejam os de contribuir diretamente com a ciência da computação na produção de software básico como o design de um teclado ideal para a língua portuguesa baseado em dados de freqüência de ocorrência dos caracteres gráficos e dos encontros ou seqüências destes mesmos caracteres, e na elaboração de um ótimo editor de textos mediante o fornecimento de um algoritmo que possibilite uma separação silábica automática e mediante a programação de ortografadores que permitem uma correção automática de textos. Necessária se faz a contribuição do fonólogo e do foneticista no sentido de descrever criteriosamente todos os sons e fonemas (em seus formantes) de uma língua para obter a síntese da voz e o reconhecimento da fala natural. A lingüística também tem uma significativa contribuição a emprestar na organização das linguagens de programação, principalmente das mais evoluídas ou das mais próximas da linguagem natural como LISP e PROLOG. Enfim, é à lingüística que compete fornecer, aos informáticos, informações sobre o processamento (receptivo e produtivo) de dados da comunicação lingüística para possibilitar o traslado de conhecimentos da mente humana ao computador com o objetivo de torná-lo máquina inteligente. Quanto aos aspectos teóricos da lingüística que devem ser do conhecimento de um cientista da computação, Ulf Gregor Baranow (1983: 23-36) em seu artigo "Perspectivas na contribuição da lingüística e de áreas afins à ciência da informação", faz um estudo bastante minucioso e chega ao ponto de preconizar o estabelecimento de uma disciplina, nos currículos dos cursos de informática, que trate da natureza da linguagem.

2.4. Interatividade

O terceiro aspecto relevante da LC consiste numa interatividade entre lingüistas e informáticos para resolver problemas mais complexos, problemas relacionados com as duas ciências. Faz-se menção principalmente, à análise automática de textos e à tradução automática, atividades que pressupõem a colaboração do lingüista para o estabelecimento de algoritmos que consigam distinguir as diversas categorias gramaticais, discriminar as diversas funções das palavras e, princi-

palmente, desambiguar homografias. Para a indexação automática de um texto, o informático, além dos recursos de um bibliotecário, deve também servir-se dos conhecimentos de um lingüista para saber como identificar os dados que merecem ser levantados (Haller, 1986).

No campo da Inteligência Artificial, onde pesquisadores exploram as decisões básicas da inteligência humana e procuram construir modelos desta inteligência em computadores (Schank & Childers, 1984 29), é ao lingüista que compete fornecer os dados sobre a linguagem mais condizente para o processamento de dados de sistemas especializados (expert systems) que constituem uma das metas prioritárias do importante projeto japonês da 5ª geração (Feigenbaum & McCorduck 1984).

No processamento da linguagem natural, trata-se de desenvolver programas que entendam a linguagem natural falada e escrita, visando à interação com o computador por meio desta mesma linguagem, e não mais em linguagem computacional, o "computés". Para um programa computacional interpretar uma comunicação em linguagem natural, o "conhecimento" necessário envolve a estrutura das sentenças, o significado das palavras, a morfologia das palavras e as regras de conversão, entre outros requisitos. (Visão 1986, p. 35)

2.5. Algumas aplicações

Analisada desta maneira a vasta gama de atividades envolvidas pela LC, ficam ressaltadas as inúmeras aplicações dessa disciplina. Citaremos algumas:

- a. Colaborar, através de pesquisas quantitativas e qualitativas criteriosas, no aperfeiçoamento de editores de textos, dotando-os de todos os recursos possíveis no que tange à separação silábica automática, aos ortografadores que permitem uma correção automática, e a dicionários (thesaurus) que fornecem dados imediatos ao editor/escritor quanto a sinônimos e significados.
- b. Contribuir significativamente na vasta área do Ensino da Linguagem com uso do computador (Costa, 1986) mediante a elaboração de programas instrucionais, tanto no âmbi-

to da leitura e da escrita quanto do vocabulário e da gramática. O levantamento do português fundamental será de valiosa contribuição para a elaboração de material didático graduado e adaptado ao nível de maturidade lingüística do aprendiz, quer se trate do português como primeira língua ou como segunda língua.

- c. Facilitar a localização de dados lingüísticos ou mais diversos, incluindo informação bibliográfica, informação acerca de pesquisas, de cursos, de especialistas, de revistas e de encontros da área, nacionais e internacionais. São os bancos de dados.
- d. Possibilitar a produção, na área da tradução, de dicionários bilíngües, de informação simultânea em várias línguas (projeto EUROTRA) e de programas especiais para munir telefones com interpretação automática instantânea.
- e. Propiciar, em atividades de biblioteconomia, o desenvolvimento de programas de indexação automáticas de livros.

Considerando todas essas aplicações, verifica-se que o lingüista realmente se encontra frente a uma variedade muito significativa de oportunidades. Acontece que raras vezes o cientista da computação "se dá conta da existência do lingüista. Compete a este descobrir estas oportunidades" (Mazzocco, 1979:6).

3. CONTEÚDO DESTES NÚMERO

3.1. Bertilo Becker docente do Instituto de Informática e pesquisador do Centro de Pesquisas Lingüísticas da PUCRS e doutorando em Lingüística Computacional, aborda "Conceitos de inteligência artificial aplicados à lingüística". Inicia conceituando o que são comportamentos inteligentes. Analisa, em seguida, a problemática da busca de solução automatizada para problemas diversos e as várias maneiras de suprir a máquina com os resultados que ela precisa para suas decisões. Ainda são mostrados três enfoques para esse tipo de conhecimento: as redes semânticas, as molduras espaço-temporais e as gramáticas.

3.2. Vera Lúcia Strube de Lima, doutora em Processamento de Linguagens Naturais pela Universidade de Grenoble e docente do Instituto de Informática da PUCRS, em seu artigo "Tratamento automatizado da língua natural - rumo à correção automática?", visa a promover o leitor com uma visão "panorâmica" no que se refere ao tema correção ortográfica automatizada, apresentando um resumo das técnicas e métodos empregados atualmente no tratamento da língua natural, abordando suas vantagens, suas deficiências e sua transposição para o português.

3.3. Patrícia Chittone Ramos, docente do Instituto de Letras e Artes da PUCRS, apresenta o artigo "Tradução automática - um desafio?". O artigo é constituído por rápidas pinceladas sobre o histórico, a metodologia e a atualidade da tradução automática.

3.4. Gabriela Ardisson Matos, docente do Departamento de Linguística da Faculdade de Letras da Universidade de Lisboa e investigadora do Instituto de Linguística Teórica e Computacional (ILTEC) de Portugal, escreve sobre a língua portuguesa no projeto EUROTRA". Entre outros assuntos, são fornecidos dados sobre os objetivos, a organização, a metodologia e as etapas percorridas até o momento na implantação do projeto EUROTRA em Portugal.

3.5. Alejandra Bentolila, mestranda em Linguística na PUCRS em seu trabalho "O professor de línguas confrontado com o minotauro informático", tenta expor e analisar, na medida do possível, as mistificações, medos e crenças amplamente difundidos na área educacional a esse respeito.

3.6. Margarete Axt, pesquisadora do Laboratório de Estudos Cognitivos do Departamento de Psicologia da Universidade Federal do Rio Grande do Sul (UFRGS) e doutoranda em Linguística Aplicada, apresenta "um projeto em linguagem LOGO para elicitar a produção de estruturas relativas em português". A autora utiliza a linguagem LOGO para construir um "kernel", que represente o co-

nhecimento das quatro estruturas básicas do processo de relativização, em língua portuguesa. São elaboradas duas gramáticas de cláusulas definidas para a montagem e análise de orações, a par de um conjunto de procedimentos que interage com o sujeito, fazendo-lhe perguntas que possam completar o aprendizado do programa sobre o conhecimento do mesmo.

3.7. Maria Teresa Riço da Fonseca Lino, docente e investigadora da Universidade Nova de Lisboa, discorre sobre "Banco de neologismo do português contemporâneo - balanço de uma experiência". O artigo constitui-se num relato de uma investigação que pretende desenvolver um banco completo dos neologismo da língua portuguesa.

3.8. José Marcelino Poersch, com pós-doutorado em Linguística Cognitiva pela Universidade da Califórnia-Berkeley, docente e pesquisador do Curso de Doutorado em Linguística da PUCRS, relata, em "O que a linguística tem a ver com o teclado de microcomputadores", os achados de uma pesquisa que sugere um novo "design" para o teclado de microcomputadores tomando em consideração a frequência dos caracteres gráficos em língua portuguesa e a dificuldade de acesso dos dedos às diversas teclas.

REFERÊNCIAS BIBLIOGRÁFICAS

01. ALLWOOD, Jeans et alii. *Logic in linguistics*. Cambridge, Cambridge University Press, 1977.
02. BARANOW, Ulf Gregor. Perspectivas na contribuição da linguística e de áreas afins à Ciência da Informação. *Ciência da informação*. Brasília, CNPq/IBICT, 12(1):23-35, 1983.
03. COSTA, Miriam Solange. O computador no ensino de línguas: retrospecto e perspectivas. *Interação*. São Paulo, Difusão Nacional do Livro, 3(18):17-20, abr. 1986.
04. FEIGENBAUM, Edward and McCORDUCK, Pamela. *The fifth generation: artificial intelligence and Japan's computer challenge to the world*. New York, New American Library, 1984.

05. GORDON, Barry. Subjective frequency and lexical decision latency function: implications for mechanisms of lexical access. *Journal of Memory and Language*, 24(6):631-45, dec. 1985.
06. HALLER, Johann. Análise lingüística e indexação automática de textos. *Veritas* Porto Alegre, PUCRS, 31(123):393-414, 1986.
07. HERDAN, Gustav. *The advanced theory of language as choice and chance*. Heidelberg, Springer-Verlag, 1966.
08. HUDSON, Patrick and BERGMAN, Jarijke. Lexical knowledge in word recognition: word length and word frequency in naming and lexical decision tasks. *Journal of Memory and Language*, 24:(1):46-58, feb. 1985.
09. MAZZOCCO, Alexis (entrevista). Opportunities for linguists in the field of computers. *The Linguistic Reporter*. sep. 1979.
10. POERSCH, José Marcelino. Redistribuição dos caracteres gráficos no teclado de microcomputadores com base na lingüística quantitativa. *Actas do IV Encontro da Associação Portuguesa de Lingüística*, APL, Lisboa, 1988. p. 249-80.
11. POERSCH, José Marcelino. ARDOM, uma nova configuração para teclados de microcomputadores. *Veritas*, 35(138):279-304, jun 1990.
12. SANKOFF, David and CEDERGREN, Henrietta (eds.) *Variation omnibus*. Alberta, Linguistic Research, 1981.
13. SCHANK, Roger & CHILDERS, Peter. *The cognitive computer: on language, learning and artificial intelligence*. Menlo Park, Addison-Wesley Publishing Company, 1984.
14. SCHREIBER, Servan & Jacques, Jean (entrevista). "Informática e Informação", *Veja*, São Paulo, Editora Abril, (900):3-6, a 4 dez. 1985.
15. VISÃO (autor não citado). "Inteligência artificial: o Brasil entra na corrida". *Visão*, pág. 34-38, 22 jan. 1986.