

Lingüística Computacional: uma breve introdução

Gabriel de Ávila Othero*
PUCRS



Resumo: Este artigo apresenta a área da Lingüística Computacional, uma área híbrida que envolve pesquisadores da Lingüística e da Informática. A Lingüística Computacional é a parte da ciência lingüística que se preocupa com o tratamento computacional da linguagem, e suas aplicações englobam programas como tradutores automáticos, chatterbots, corretores ortográficos e gramaticais, parsers, entre outros. O conhecimento lingüístico é indispensável para desenvolver tais aplicativos, e somente a interação entre pesquisadores das duas áreas – Lingüística e Informática – pode conquistar bons resultados.

Palavras-chave: Lingüística Computacional. Processamento de Linguagem Natural.

Abstract: This article presents some aspects of Computational Linguistics, a hybrid area that involves both Computational and Linguistic researchers. Computational Linguistics is a part of the linguistic science that is focused on the computational treatment of language. Its applications include programs such as automatic translators, chatterbots, grammar and spelling checkers, parsers, among others. A background in Linguistics is essential for the development of such software and only an interaction between researchers from these two fields – Linguistics and Computer Science – can achieve good results.

Key-words: Computational Linguistics. Natural Language Processing.

* Aluno do Programa de Pós-Graduação em Letras – Doutorado em Lingüística Aplicada –, pela Pontifícia Universidade Católica do Rio Grande do Sul – PUCRS. E-mail: gabnh@terra.com.br.

Introdução

Neste artigo, apresentamos algumas aplicações de trabalhos que envolvem o tratamento computacional da linguagem natural. Queremos mostrar que a área da Lingüística que se ocupa com esse tipo de trabalho – a Lingüística Computacional – envolve (ou deveria envolver) um intercâmbio constante entre as áreas da Lingüística e da Informática. Além disso, falaremos sobre alguns princípios e áreas da Lingüística que servem de base para aplicações em Lingüística Computacional, especialmente em trabalhos de Processamento da Linguagem Natural.

Princípios da Lingüística Computacional

A área responsável pela investigação do tratamento computacional da linguagem e das línguas naturais – e a área em que este artigo se insere – é conhecida como Lingüística Computacional. De acordo com Vieira & Lima (2001, p. 1), a Lingüística Computacional pode ser entendida como “a área de conhecimento que explora as relações entre lingüística e informática, tornando possível a construção de sistemas com capacidade de reconhecer e produzir informação apresentada em linguagem natural”.

A Lingüística Computacional envolve as diferentes áreas de pesquisa tradicionalmente conhecidas em Lingüística Teórica e Aplicada, como a Sintaxe, a Semântica, a Fonética e a Fonologia, a Pragmática, a Análise do Discurso, etc. Todo esse conhecimento é utilizado para tentar processar (leia-se “compreender e produzir”) as línguas naturais e dominar o conhecimento lingüístico envolvido no domínio de uma linguagem natural.

A Lingüística Computacional pode ser didaticamente dividida em duas subáreas: a Lingüística de Corpus e o Processamento de Linguagem Natural (PLN). Essa divisão nem sempre é nítida, uma vez que há muitos trabalhos que envolvem as duas áreas. Vejamos.

A Lingüística de Corpus preocupa-se basicamente com o trabalho a partir de corpora eletrônicos que contenham amostras de linguagem natural. Essas amostras podem ser de diferentes fontes. Por isso, podemos encontrar os mais variados bancos de corpora eletrônicos: há corpora de linguagem falada, corpora de linguagem escrita literária, corpora com textos de jornal, corpora compostos exclusivamente por falas de crianças em estágio de desenvolvimento lingüístico, etc. Os trabalhos envolvendo corpora lingüísticos nem sempre têm como objetivo produzir algum *software* ou

aplicativo. Normalmente, eles estão voltados para o estudo de determinados fenômenos lingüísticos e sua ocorrência em grandes amostras de uma determinada língua (ou de uma variedade, dialeto ou modalidade dela).¹

De acordo com Berber Sardinha (2000b, p. 2),

A Lingüística de Corpus se ocupa da coleta e exploração de corpora, ou conjuntos de dados lingüísticos textuais que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade lingüística. Como tal, dedica-se à exploração da linguagem através de evidências empíricas, extraídas por meio de computador.

A área de Processamento de Linguagem Natural, por outro lado, preocupa-se diretamente com o estudo da linguagem voltado para a construção de *softwares*, aplicativos e sistemas computacionais específicos, como tradutores automáticos, *chatterbots*, *parsers*, reconhecedores automáticos de voz, geradores automáticos de resumos, etc. Cabe à área de PLN justamente a construção de programas capazes de interpretar e/ou gerar informações em linguagem natural. Além disso, de acordo com Vieira 2002, p. 20), “para o processamento da língua natural, vários subsistemas são necessários para dar conta dos diferentes aspectos da língua: sons, palavras, sentenças e discurso nos níveis estruturais, de significado e de uso”.

Algumas Aplicações da Lingüística Computacional

Nesta seção, apresentaremos algumas das aplicações que são frutos da Lingüística Computacional, dando ênfase à parte lingüística envolvida nos trabalhos aqui mostrados. Tentaremos ressaltar a grande importância de uma interação constante entre pesquisadores das áreas da Lingüística e da Informática, uma vez que a cooperação e o diálogo entre essas duas áreas é essencial para o desenvolvimento de programas que lidam com a linguagem natural. Sabemos que, ao lingüista, faltam, muitas vezes, conhecimento prático e teórico sobre linguagens de programação e desenvolvimento de *softwares*. Da mesma maneira, ao engenheiro da computação faltam conhecimentos sólidos e profundos sobre teorias lingüísticas para o tratamento e formalização das linguagens naturais. Acreditamos firmemente que trabalhos em Lingüística Computacional devam envolver pesquisadores provenientes das duas áreas: Lingüística e Informática.

¹ Cf. Garside, Leech & McEnery (1997) e Berber Sardinha (2005).

Como dissemos, pesquisas em Lingüística Computacional podem envolver diversos campos de estudo da Lingüística Teórica e Aplicada. Os primeiros desenvolvimentos da Lingüística Computacional começaram nos anos 1950. Ela teve grande impulso principalmente graças a esforços para o desenvolvimento de programas de tradução automática nas décadas de 1950 e 1960, e sua história está intrinsecamente ligada a desenvolvimentos na área da Inteligência Artificial.²

De acordo com Grisham (1992, p. 1),

o potencial [dos computadores] para o processamento de linguagem natural foi reconhecido bem cedo no desenvolvimento de computadores, e trabalhos em Lingüística Computacional – basicamente para tradução automática – começaram na década de 1950 em diversos centros de pesquisa. O rápido crescimento na área, no entanto, aconteceu principalmente a partir do final dos anos 1970.³

Hoje em dia, a busca por agentes inteligentes, a procura por interfaces computacionais mais amigáveis e por *softwares* voltados para o trabalho com linguagem natural têm motivado muitas pesquisas na área, e diversos são os frutos que envolvem áreas como a Inteligência Artificial e a Lingüística Computacional. Para McDonald & Yazdani (1990, p. 176), “a pesquisa em PLN pode proporcionar *insights* bastante úteis sobre processos e representações da linguagem na mente humana, apontando, assim, para a verdadeira IA”.⁴

Veremos aqui, de maneira bastante sucinta, algumas das aplicações que envolvem essas duas ciências, dando ênfase – devido à nossa formação e ao contexto em que este artigo se enquadra – antes ao domínio lingüístico envolvido nos trabalhos aqui mostrados do que aos formalismos computacionais ou modelos de engenharia da computação que estão por trás deles.⁵

² Cf. Grisham (1992), Russel & Norvig (1995) e Othero & Menuzzi (2005).

³ Trecho original: “The potential for natural language processing was recognized quite early in the development of computers, and work in computational linguistics – primarily for machine translation – began in the 1950s at a number of research centers. The rapid growth in the field, however, has taken place mostly since the late 1970s”.

⁴ Trecho original: “Research into NLP may provide useful insights into processes and representation of language in the human mind, thus heading towards true AI”.

⁵ Evidentemente, não está entre os nossos objetivos abranger aqui, neste artigo, todas as áreas do conhecimento lingüístico que contam com desenvolvimentos e aplicações da Lingüística Computacional. Para uma exemplificação mais completa, remetemos o leitor outras obras, como Cole et al. (1997), Garside, Leech & McEnery (1997), Vieira & Lima (2001), Grisham (2002) e Othero & Menuzzi (2005).

Envolvendo as áreas da Fonética e Fonologia, por exemplo, encontramos muitos aplicativos de PLN. A Fonética e a Fonologia são as áreas da Lingüística preocupadas em estudar os sons das línguas humanas. Grosso modo, podemos dizer que a Fonética ocupa-se dos estudos dos fones, dos sons concretizados na fala. Ela está interessada na parte acústica, articulatória e fisiológica da produção dos sons da fala. A Fonologia, por outro lado, concentra-se em estudar os fonemas e o sistema fonológico subjacente de uma língua. Ela investiga o sistema abstrato que envolve o conhecimento fonológico dos falantes.⁶

De maneira geral, podemos dizer que os desenvolvimentos que envolvem essas duas áreas da Lingüística enquadram-se nas áreas de (i) reconhecimento de fala; (ii) síntese de fala; e (iii) sistemas de diálogos em língua falada.

Entre as aplicações que já foram desenvolvidas, a partir de estudos fonéticos e fonológicos, podemos destacar aqui os aplicativos de *reconhecimento de fala*, que apresentam diversas finalidades: podem servir desde meros reconhecedores de comandos de voz em um aparelho celular (o que torna possível a discagem a partir do reconhecimento da voz do usuário do aparelho, como já acontece, por exemplo, nos aparelhos Motorola V60, Motorola V60i e Motorola C333.⁷ todos já comercializados no Brasil), até programas que reconheçam a fala a ponto de serem capazes de digitar um texto ditado por um usuário, como o *software* Via Voice, desenvolvido pela IBM.⁸

Na área da *síntese de fala*, destacam-se os programas que são capazes de gerar fala a partir de dados. Há, por exemplo, o programa Talk it!,⁹ que pode “ler em voz alta” palavras digitadas pelo usuário. Esse programa pronuncia vocábulos isolados, em inglês e em espanhol, e pode ajudar o usuário a descobrir a pronúncia correta de palavras na língua-alvo. Outros aplicativos que merecem destaque envolvendo a síntese de fala são os programas capazes de ler em voz alta textos maiores e documentos escritos e armazenados no computador, como o programa Virtual Vision,¹⁰ desenvolvido pela MicroPower. Esse programa “lê em voz alta” os textos de documentos do MS Office e de páginas na Internet. Além de facilitar o uso do computador para usuários leigos, eles auxiliam de grande maneira os usuários que apresentam algum tipo de deficiência visual.

⁶ Cf. Callou & Leite (2001), Mori (2001) e Cristófar-Silva (2002).

⁷ www.motorola.com.br.

⁸ www-3.ibm.com/software/speech.

⁹ www.text2speech.com.

¹⁰ www.micropower.com.br/dv/vvision/index.asp.

Ainda nas áreas da Fonética e da Fonologia, podemos mencionar alguns aplicativos ainda mais sofisticados, que envolvem a interação *humano x computador* através de diálogos orais em linguagem natural. Programas que sejam capazes de conversar com o ser humano ainda estão em seus desenvolvimentos iniciais e são provavelmente muito mais comuns em obras de ficção científica (como a precursora máquina inteligente HALL, do filme *2001: uma odisséia no espaço*, de Stanley Kubrick) do que na realidade. Contudo, já sabemos que eles podem ter um potencial enorme para o desenvolvimento de *softwares* de atendimento eletrônico e agentes inteligentes na área de ensino a distância (EAD) e educação através do computador (como programas de ensino de idiomas, ou *CALL – Computer Assisted Language Learning*). Esse tipo de aplicativo envolve um conjunto de conhecimentos complexos. Além de conhecimentos de Fonética e Fonologia, desenvolvimentos nessa área devem envolver também o estudo da Análise da Conversação, da Semântica, da Pragmática (especialmente no que se refere à teoria dos atos de fala), da Análise de Discurso e da Lingüística Textual, entre outros.

Aplicações como essas certamente facilitarão o acesso de computadores pessoais e aparelhos eletrônicos para pessoas com deficiência visual e também para leigos, que poderão interagir com a máquina através da linguagem falada, de uma maneira extremamente natural – ao menos, é o que se espera.

Outras áreas da Lingüística de grande importância para o desenvolvimento de programas de PLN são a Sintaxe e a Semântica. Grosso modo, podemos entender a Sintaxe como a disciplina “que estuda as regras, as condições e os princípios subjacentes à organização estrutural dos constituintes das frases, ou seja, o estudo da ordem dos constituintes das frases” (sic) (Mateus & Xavier, 1992: 1079). A Semântica, também provisoriamente, pode ser definida como a área da Lingüística que se ocupa em estudar o significado das palavras e proposições.

Acreditamos que estudos em Sintaxe Gerativa e Funcional e em Semântica Cognitiva, Computacional e Formal possam trazer bons resultados no desenvolvimento de programas que lidem com a geração automática de sentenças, especialmente no desenvolvimento de programas conhecidos como *chatterbots*.

Os *chatterbots* são programas desenvolvidos para interagir com usuários humanos através de diálogos em linguagem natural, na modalidade escrita.¹¹ Esse nome (*chatterbot*) vem da junção de duas palavras inglesas: *chat* (conversar, bater papo) e *bot* (abreviação de *robot*, robô).

¹¹ Cf. <http://bots.internet.com/search/s-chat.htm> e também <http://cybelle.cjb.net/>.

O primeiro *chatbot* desenvolvido foi a ELIZA,¹² criado pelo pesquisador Joseph Weizenbaum, no MIT, em 1966. O programa ELIZA foi desenvolvido para ser uma espécie de psicólogo, ou conselheiro sentimental virtual. Na verdade, o objetivo principal de Weizenbaum não era criar um “psicólogo virtual”, mas antes um programa de PLN de conversação utilizando um sistema conhecido por *templates*. O curioso da história é que, durante os testes do programa, quando Weizenbaum pedia para que algumas pessoas interagissem com a ELIZA, elas acabavam ficando tão íntimas do programa que chegavam a pedir para que o próprio Weizenbaum saísse da sala para que a conversa pudessem ser mais particular.

De acordo com Covington (1994, p. 14),

ELIZA era tão convincente que as pessoas espontaneamente contavam a ela seus problemas e pediam a Weizenbaum para que saísse da sala para que pudessem conversar privadamente. Os psiquiatras até mesmo pensaram que essa era uma maneira prática de se administrar processos de terapia. Weizenbaum ficou alarmado com esse fato; longe de ser um triunfo da inteligência artificial, ele considerava ELIZA uma evidência de quão fácil era enganar as pessoas com os computadores.¹³

Veja, por exemplo, no Quadro 1, um trecho de uma conversa entre ELIZA e uma usuária humana, apresentado em Weizenbaum (1966):¹⁴

¹² <http://www-ai.ijs.si/eliza/eliza.html>.

¹³ Trecho original: “ELIZA was so convincing that people willingly told it their troubles and asked Weizenbaum to go away so they could converse in private. Psychiatrists even thought it was a practical way of administering therapy. Weizenbaum was alarmed at this; far from a triumph of artificial intelligence, he considered ELIZA to be evidence of how easy it is to deceive people with computers”.

¹⁴ WEIZENBAUM, J. ELIZA – a Computer Program for the Study of Natural Language Communication between Man and Machine. *Communications of the ACM* 9, 1966, apud Othero & Menuzzi (2005, p. 32-33).

Quadro 1: Exemplo de conversação com ELIZA

Usuário: OS HOMENS SÃO TODOS IGUAIS.
ELIZA: EM QUE SENTIDO
Usuário: eles estão sempre nos importunando por algum motivo ou outro.
ELIZA: PODE PENSAR EM UM EXEMPLO ESPECÍFICO
U: Bem, o meu namorado me fez vir até aqui.
ELIZA: SEU NAMORADO FEZ VOCÊ VIR ATÉ AQUI
U: Ele diz que eu fico muito tempo deprimida.
ELIZA: FICO TRISTE EM SABER QUE VOCÊ FICA DEPRIMIDA
U: É verdade. Eu sou infeliz.
ELIZA: VOCÊ ACHA QUE VIR AQUI VAI AJUDAR VOCÊ A NÃO SER INFELIZ
U: Eu preciso de uma ajuda, acho que isso é certo.
ELIZA: O QUE SIGNIFICARIA PARA VOCÊ SE VOCÊ CONSEGUISSSE AJUDA
U: Talvez eu pudesse aprender a conviver com a minha mãe.
ELIZA: FALE-ME MAIS SOBRE A SUA FAMÍLIA
U: Minha mãe cuida de mim.
ELIZA: QUEM MAIS EM SUA FAMÍLIA CUIDA DE VOCÊ
U: Meu pai
ELIZA: SEU PAI
U: De certa forma, você é como meu pai.
ELIZA: QUE SEMELHANÇA VOCÊ VÊ
U: Você não é muito agressiva, mas eu acho que você não quer que eu perceba isso.
ELIZA: O QUE FAZ VOCÊ PENSAR QUE EU NÃO SOU AGRESSIVA
U: Você não discute comigo.
ELIZA: POR QUE VOCÊ ACHA QUE EU NÃO DISCUTO COM VOCÊ
U: Você tem medo de mim.
ELIZA: AGRADA A VOCÊ SABER QUE EU TENHO MEDO DE VOCÊ
U: Meu pai tem medo de todo mundo.
ELIZA: O QUE MAIS VEM À SUA MENTE QUANDO VOCÊ PENSA NO SEU PAI
U: Valentões.
ELIZA: ISSO TEM ALGUMA RELAÇÃO COM O FATO DE O SEU NAMORADO TER FEITO VOCÊ VIR AQUI

Ultimamente, os *chatbots* vêm sendo desenvolvidos especialmente para trabalhar com atendimento virtual e com tutoriais educativos. Um programa desse tipo deve ser capaz de poder manter uma conversa com um humano da forma mais natural possível. Por isso, acreditamos que a próxima geração de *chatbots* deva exigir dos programadores um profundo conhecimento da sintaxe (que permitirá ao programa gerar infinitas sentenças da língua combinando um número finito de regras e elementos lexicais) e da semântica de uma língua (que tornará possível fazer com que o programa seja capaz de interpretar o significado do *input* lingüístico dado a ele pelo usuário humano).

Evidentemente, outras áreas da Lingüística serão necessárias para o desenvolvimento e aperfeiçoamento de *chatterbots*, como a Lingüística do Texto (que dará conta, por exemplo, das relações anafóricas e extra-oracionais do discurso), a Dialetologia (que auxiliará o programa a compreender diferentes dialetos, gírias, regionalismos e jargões) e a Análise da Conversação (que ajudará, por exemplo, a determinar os turnos conversacionais, além de auxiliar na compreensão e uso de marcadores conversacionais), pelo menos.

Conhecimentos em sintaxe e semântica são também fundamentais para outros aplicativos, como tradutores automáticos, *parsers*, geradores automáticos de resumos, corretores ortográficos e gramaticais, classificadores automáticos de documentos digitais, etc.¹⁵

Os tradutores eletrônicos são programas que se encarregam da tradução automática por computador de textos de uma língua para outra, ou outras. Os primeiros tradutores, como vimos, começaram a ser desenvolvidos nos anos 1950 e 1960. Até hoje, no entanto, não há disponível no mercado algum programa que seja realmente eficiente, que consiga traduzir um texto do inglês para o português, por exemplo, de maneira tão exata quanto um tradutor humano conseguiria. Isso se deve à complexidade e riqueza das línguas humanas, que relutam em se entregar à formalização do computador. Na verdade, quando falamos em tradutores automáticos, podemos estar falando de três diferentes tipos de programas: os gerenciadores de terminologia, as ferramentas de tradução automática e os tradutores automáticos que auxiliam a tradução humana.

O primeiro tipo de programa é especificamente planejado para auxiliar na tradução de textos técnicos de determinada área, e seu vocabulário pode ficar restrito aos termos técnicos e jargões dessa área, o que possibilita ao programa alcançar níveis mais satisfatórios de tradução, pois seu domínio é bastante restrito. Já as ferramentas de tradução automática são as mais conhecidas. Elas são programas destinados à tradução automática de textos quaisquer, sem qualquer restrição na escolha do léxico, gênero ou assunto do texto e também sem qualquer interferência ou ajuda de um tradutor humano. Por último, encontramos programas que apenas auxiliam o tradutor humano na tradução de um texto. Esses tradutores não primam pela perfeição ou acabamento do texto final, mas almejam auxiliar o usuário a traduzir grandes quantidades de texto.

¹⁵ Cf. Garside, Leech & McEnery (1997), Vieira (2002) e Othero & Menuzzi (2005).

Outro aplicativo que vem se mostrando extremamente útil no “manejo” das línguas naturais pelo computador são os *parsers*. De acordo com Mateus & Xavier (1992, p. 886), *parsing* é o “processo de atribuição de uma estrutura e de uma interpretação a uma sequência lingüística”. No contexto da Lingüística Computacional, contudo, podemos definir *parsing* como a interpretação automática (ou semi-automática) de sentenças de linguagem natural por programas de computador conhecidos justamente como *parsers*.

Esses aplicativos, os *parsers*, lidam com o tratamento sintático (ou morfossintático, ou ainda sintático-semântico) das línguas naturais. Eles são capazes de classificar morfossintaticamente as palavras e expressões de sentenças em Linguagem Natural e atribuir a elas sua estrutura sintagmática, de acordo com uma gramática pré-estabelecida.¹⁶

Considerações finais

Neste breve artigo, quisemos apenas apresentar uma leve introdução a alguns dos estudos que envolvem a área da Lingüística Computacional. Ela é uma área que vem apresentando um desenvolvimento considerável nos últimos anos, pois desenvolve trabalhos que visam a facilitar cada vez mais a interação *homem x máquina*, o que tem se mostrado uma tendência forte e necessária para a popularização dos computadores.

Ainda assim, a Lingüística Computacional parece-nos pouco aproveitada nos cursos de Letras no Brasil. Esperamos que nosso leitor possa consultar a bibliografia indicada ao longo do texto para aprofundar seu conhecimento nessa área fascinante que é a Lingüística Computacional.

Afinal, defendemos que essa área é uma área “híbrida” por natureza, envolvendo o comprometimento de pesquisadores com formação em Lingüística e em Informática. Acreditamos que a co-operação e a interface entre esses dois campos de estudo é de crucial importância para desenvolvimentos em Lingüística Computacional, porque ela é responsável por desenvolver *aplicativos computacionais* que trabalham com a *linguagem natural*. Como vimos, o objetivo de muitos desses aplicativos é permitir que a interação entre computador e usuário seja cada vez mais simples e amigável para o usuário. Afinal, se temos de trabalhar e lidar com computadores, acreditamos que o ideal seja que eles aprendam a se comunicar através da nossa linguagem e não o contrário.

¹⁶ Para saber mais sobre *parsers*, cf. Garside, Leech & McEnery (1997) e Othero & Menuzzi (2005).

Referências

- BERBER SARDINHA, Tony. *Linguística de corpus: histórico e problemática*. *Revista D.E.L.T.A.*, Vol. 16, N. 2, 2000a. [www.scielo.br/delta].
- BERBER SARDINHA, Tony. *O que é um corpus representativo?*. LAEL PUCSP, 2000b. [lael.pucsp.br/direct].
- BERBER SARDINHA, Tony. *Linguística de corpus*. Barueri: Manole, 2004.
- CALLOU, D.; LEITE, Y. *Iniciação à fonética e fonologia*. Rio de Janeiro: Jorge Zahar, 2001.
- COLE, R. A. et al. (eds.). *Survey of the state of the art in human language technology*. Cambridge: Cambridge University Press / Giardini, 1997. [www.dfki.de/~hansu/HLT-Survey.pdf].
- COVINGTON Michael. A. *Natural language processing for Prolog programmers*. New Jersey: Prentice Hall, 1994.
- CRISTÓFARO-SILVA, T. *Fonética e fonologia do português: roteiro de estudos e guia de exercícios*. São Paulo: Contexto, 2002.
- GARSDIE, Roger; LEECH, Geoffrey; McENERY, Anthony. *Corpus annotation: linguistic information from computer text corpora*. London / New York: Longman, 1997.
- GRISHAM, Ralph. *Computational linguistics: an introduction*. Cambridge: Cambridge University Press, 1992.
- MATEUS, Maria Helena Mira; XAVIER, Maria Francisca (Orgs.). *Dicionário de termos lingüísticos*. Lisboa: Edições Cosmos, 1992.
- McDONALD, Carlton; YAZDANI, Masoud. *Prolog programming: a tutorial introduction*. Oxford: Blackwell Scientific Publications, 1990.
- MORI, A. C. Fonologia. In: MUSSALIM, F.; BENTES, A. C. (Orgs.). *Introdução à lingüística: domínios e fronteiras*. V. 1. São Paulo: Cortez, 2001.
- OTHERO, Gabriel de Ávila. Algumas considerações sobre a importância da continuidade tópica na classificação automática de documentos digitais. *Revista Virtual de Estudos da Linguagem – ReVEL*, ano 2, n. 3, 2004. [www.revelhp.cjb.net].
- OTHERO, Gabriel de Ávila; MENUZZI, Sérgio de Moura. *Linguística Computacional: teoria & prática*. São Paulo: Parábola Editorial, 2005.
- RUSSEL, S.; NORVIG, P. *Artificial intelligence: a modern approach*. New Jersey: Prentice-Hall, 1995.
- VIEIRA, R. *Linguística computacional: fazendo uso do conhecimento da língua*. *Entrelinhas*, ano 2, n. 4, São Leopoldo: UNISINOS, 2002.
- VIEIRA, R.; LIMA, V. L. S. *Linguística computacional: princípios e aplicações*. In: IX Escola de Informática da SBC-Sul. Luciana Nedel (Ed.) Passo Fundo, Maringá, São José. SBC-Sul, 2001.