

PESQUISA INTERDISCIPLINAR COM E SOBRE A LINGUAGEM EM AMBIENTE DIGITAL: ABORDAGENS E TÉCNICAS AUXILIARES NA ANÁLISE E REGISTRO DE TEXTOS

Helvia Pereira Pinto Bastos*

Resumo

Caracterizadas pelo cruzamento e utilização de diferentes enfoques teórico-metodológicos, pesquisas sobre a linguagem em ambiente virtual têm resultado em estudos que abrangem desde a Comunicação Mediada por Computador à Interação Homem-Máquina. O trabalho apresenta abordagens e técnicas utilizadas nesses estudos, assim como iniciativas que contribuem para a construção de acervos e registros da linguagem. .

Palavras-chave

Linguagem. Tecnologias Digitais. Memória.

Abstract

Characterized by theoretical and methodological interfaces, research on language and about language in virtual environments has resulted in studies spanning from Computer Mediated Communication to Human-Machine Interaction. This paper presents approaches and techniques used in such studies, as well as initiatives that contribute for the organization of digital collections and linguistic documentation.

Key Words

Language. Digital technologies. Memory.

Introdução

Este estudo trata das relações interdisciplinares entre os diversos campos da Lingüística e da Ciência da Computação em pesquisas sobre a linguagem textual em meio eletrônico, assim como aplicações desenvolvidas na área da Inteligência Artificial e arquivamento digital de textos.

À abrangência e variedade das áreas envolvidas na investigação sobre a linguagem humana somam-se as constantes inovações produzidas no campo da Informática. Neste estudo, o termo “linguagem” é usado em sentido estrito – meio de expressão e comunicação verbal escrita ou oral; assim, não são contempladas outras expressões semióticas (música, cinema, artes plásticas, por exemplo). O

* Mestre em Linguagem e Cognição (UENF, 2002). Doutoranda em Informática na Educação (UFRGS).

trabalho se propõe, portanto, a apresentar um panorama dos aportes teóricos e técnicos mais recorrentes na pesquisa contemporânea acerca de (i) diferentes aspectos relativos ao discurso em meio eletrônico, (ii) contribuições da Linguística para o desenvolvimento de dispositivos e sistemas computacionais, (iii) apoio das tecnologias na construção de acervos digitais focados no registro da língua e literatura de língua portuguesa no Brasil.

Entre as práticas sociais contemporâneas, destacam-se as possibilitadas pela Comunicação Mediada por Computador (CMC), seja por meio escrito – correio eletrônico, salas de bate-papo, fóruns de discussão, ou com uso de som e imagem – videoconferência, por exemplo. As interações discursivas produzidas por meio dessas ferramentas têm-se constituído um campo fértil para os estudos lingüísticos, em função, especialmente, das particularidades da comunicação baseada em texto no ciberespaço.

A escrita digital é potencialmente maleável, hipertextual, descentralizada e descontínua, permitindo a reconfiguração das formas tradicionais de leitura e produção textual. Destaca-se, nesse aspecto, o advento de ferramentas e sites de escrita coletiva – *wikis*, espaços em que a distinção entre autor e leitor se torna menos nítida. Nesse sentido, verifica-se a expansão de ambientes em que o leitor-autor tem participação ativa na inserção, escolha e edição dos conteúdos, uma característica dos ambientes surgidos com a chamada Web 2.0¹, como blogs e fotoblogs, redes sociais, e ambientes de realidade aumentada. A possibilidade de manipular o texto é exemplificada, também, por programas de criação de mapas conceituais², uma ferramenta bastante útil em projetos educacionais.

A estrutura nodal da internet³ acarreta comportamentos diferentes no ato de leitura. O leitor-navegador prefere parágrafos curtos e objetivos – refletindo a diversidade e velocidade de seu tempo. Esse comportamento é ainda potencializado pela quantidade ilimitada, e inédita na história humana, de informação disponível no espaço cibernético que permite aos sujeitos o acesso e consulta aos mais variados tipos de base de dados – bancos de teses e dissertações, dicionários e enciclopédias, documentação institucional, periódicos eletrônicos, acervos de produção literária, para citar alguns.

¹ Primo e Recuero (2006) explicam que o termo Web 2.0 descreve “o atual período da Rede cuja ênfase passa da publicação (que caracterizou os primeiros dez anos da Web) para a colaboração”.

² Por exemplo, o software livre CMap Tools (<<http://cmap.ihmc.us>>

³ Neste estudo, os termos “internet” e “web” são usados como sinônimos; embora, em sentido estrito, o primeiro se refira à infraestrutura de computadores conectados em rede, e o segundo ao conteúdo que ali circula.

Outras contribuições e inovações surgidas com o desenvolvimento de programas computacionais são discutidas neste trabalho, como a Lingüística Computacional e o desenvolvimento de máquinas que interagem com humanos. A parte final deste estudo trata da contribuição da Ciência da Computação para a pesquisa e trabalhos focados no arquivamento e na preservação de diferentes manifestações da língua.

Interfaces teóricas no estudo da linguagem em ambientes virtuais

Sendo a linguagem a principal mediadora das relações sociais e da própria realidade em que os sujeitos se inserem, os estudos da linguagem têm se desdobrado em diferentes áreas interdisciplinares, por exemplo: Lingüística Textual, Análise⁴ e Teoria do Discurso, Etnolingüística, Sociolingüística, Lingüística Aplicada, Psicolingüística. Essa variedade de campos de estudo, por sua vez, se alia à Informática na pesquisa em diferentes áreas do conhecimento: Antropologia, Sociologia, Psicologia, Ciências Cognitivas, Marketing e Publicidade, entre outras. No campo da Educação, são inúmeras as contribuições para a Educação a Distância, para o ensino e aprendizagem da língua materna, da literatura, e de línguas estrangeiras, entre outros.

Esse leque de possibilidades resulta, em parte, do fato da linguagem não ser mais compreendida, como argumenta Koch (1995, p.9), apenas como um “espelho” (expressão da realidade e do pensamento), ou como um “instrumento” (mero código de comunicação). Segundo a autora, a concepção que melhor contempla os estudos lingüísticos contemporâneos é a que considera a linguagem “como atividade, como forma de ação, ação interindividual finalisticamente orientada; como lugar de interação que possibilita aos membros de uma sociedade a prática dos mais diversos tipos de atos (...)”. Essa noção se fundamenta em diferentes correntes teóricas que apresentam uma perspectiva mais pragmática da linguagem, tais como: a Teoria da Enunciação (BAKTHIN, 2000; BENEVISTE, 1995), a Teoria dos Atos de Fala (SEARLE, 1981; GRICE, 1989), a Teoria dos Jogos de Linguagem (Wittgenstein, 1998). Vale destacar os estudos de natureza sócio-

⁴Caregnato e Mutti (2006) dizem haver “ao menos 57 variedades de análise do discurso, com enfoques variados, a partir de diversas tradições teóricas [...]”, todas comungando a idéia de que a linguagem não é apenas uma ferramenta neutra de reflexão ou descrição da realidade e a certeza de que ela é essencial na construção das identidades sociais.

construtivistas de Piaget (2007) e Vigotski (1998), para quem a linguagem exerce o importante papel mediador nas relações sociais e no desenvolvimento cognitivo dos sujeitos.

Os estudos sobre os gêneros discursivos, as variações de registro, e sobre diferentes aspectos das trocas discursivas em ambiente virtual têm se valido tanto de metodologias de análise tradicionalmente usadas em áreas afins, como de técnicas advindas da Computação. Moraes e Gallazzi (2007, p. 141) preferem considerar as técnicas de análise textual listadas a seguir como metodologias de análise que contêm “em seu bojo todo um conjunto de possíveis técnicas e procedimentos”. Nesse domínio de metodologias de análise *com e sobre* o texto, este trabalho destaca: (i) o estudo dos gêneros textuais; (ii) a análise de conteúdo, e (iii) a análise da conversação.

Trabalhos sobre “gêneros textuais” (também “gêneros discursivos”) se fundamentam, sobretudo, no estudo seminal de Bakhtin (2000, p. 279, grifos no original), em que *gênero* é definido como: “tipos ‘relativamente estáveis de enunciados’ encontrados em todas as esferas de utilização da língua correspondendo aos diferentes enunciados existentes”. Conforme Bakhtin, os gêneros não são estáticos; transformam-se ao assimilar outros em conformidade com as mudanças nas sociedades em que ocorrem.

Nesse sentido, as ferramentas de Comunicação Mediada por Computador (CMC) possibilitaram o surgimento do que Marsusch (2004) denomina de “gêneros emergentes”, entre eles: correio eletrônico, fóruns e blogs (de comunicação assíncrona), chats e vídeoconferência (de comunicação em tempo real). Essas novas formas textuais são caracterizadas por marcas paralinguísticas e de oralidade – estratégias que visam compensar o distanciamento físico entre os interlocutores, situação em que estes não podem se valer das pistas comumente encontradas em encontros face a face (expressão corporal, ritmo e entonação de voz, por exemplo). Assim, a escrita eletrônica é caracterizada por supressão ou uso exacerbado de maiúsculas, supressão ou troca de fonemas, uso repetitivo de pontuação, substituição da acentuação gráfica por vogais, onomatopéias, e utilização de recursos icônicos (BASTOS et al., 2010). Essa reconfiguração da escrita constitui um desafio particularmente importante para os profissionais da educação, uma vez que se vive um tempo de transição em que os alunos incorporam seus novos modos de expressão à prática escolar.

Além da análise textual convencional, duas metodologias semelhantes são frequentemente usadas no estudo dos gêneros textuais digitais: a “análise de

conteúdo” (ACt) e a “análise da conversação” (ACo) A primeira (ACt) é definida por Bardin (2010, p.40) como “ um conjunto de técnicas de análise das comunicações visando obter, por procedimentos, sistemáticos e objetivos de descrição do conteúdo das mensagens, indicadores (quantitativos ou não) que permitam a inferência de conhecimentos relativos às condições de produção/recepção [...] destas mensagens”. A ACt interessa-se pelos aspectos empíricos e pragmáticos do texto, tais como dados não-explícitos e ocorrências de termos. Também uma abordagem de natureza empírica-pragmática, a ACo se concentra no levantamento de unidades de análise, no sistema de trocas de turnos, no mapeamento de marcadores conversacionais, nas estratégias verbais e não-verbais usadas pelos interlocutores (MARCUSCHI, 2003). Essas duas abordagens etnometodológicas realizam o mapeamento de padrões discursivos e o estabelecimento de categorias ou códigos para proceder à análise, razão pela qual são utilizadas como técnicas auxiliares no estudo de transcrições de textos veiculados em ferramentas de CMC existentes em redes sociais, blogs e ambientes virtuais de aprendizagem.

Com o aporte teórico e técnicas de pesquisa comentadas neste trabalho, os estudos etnográficos sobre as práticas discursivas em ambientes virtuais têm se configurado num novo espaço para pesquisa – a netnografia⁵. Este recente campo de investigação começou a se desenvolver com o surgimento e rápida disseminação das ferramentas de CMC e das redes sociais, em geral. Nesse sentido, os métodos e procedimentos usados na exploração das práticas comunicativas no ciberespaço precisam ser “constantemente revisadas para que permaneçam pertinentes e atualizadas” em conformidade com o ritmo e dinamismo tecnológico (AMARAL et al., 2008).

A seção seguinte trata de técnicas e aplicações oriundas do cruzamento de estudos da linguagem com programas computacionais.

Abordagens e aplicações da Linguística Computacional

Os avanços da Ciência da Computação facilitaram e expandiram os estudos lexicográficos, no bojo da Linguística Computacional (LC), termo usado, neste trabalho, para se referir aos campos de pesquisa que utilizam a linguagem natural em sistemas computacionais, seja como objeto ou como meio de investigação.

⁵ Também chamada de “etnografia virtual” por Hine (2000) para se referir ao trabalho de imersão e envolvimento do investigador em páginas da Rede.

Vieira e Lima (2001) complementam: “A lingüística computacional é 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”.

Para “compreender e reproduzir” a linguagem humana, os trabalhos em LC costumam envolver campos da Linguística Geral, como a Sintaxe, a Semântica, a Fonética, a Pragmática e a Análise do Discurso. Com esse aporte, duas subáreas interdependentes e complementares da LC têm concorrido de forma significativa para a compreensão dos fatos lingüísticos: o Processamento da Linguagem Natural (PLN) e a Linguística de Corpus (LCp) – comentada na seção final deste artigo.

Operando de forma paralela e complementar à LCp, o Processamento de Linguagem Natural (PLN) tem dado importantes contribuições para a construção de programas computacionais focados no processamento da língua e na área da Inteligência Artificial. O PLN colabora com a Inteligência Artificial no desenvolvimento de software para tradução automática e geração de resumos, corretores automáticos, reconhecimento de fala, síntese de fala⁶, *parsers* (analisadores sintáticos), para citar os mais comuns.

Um resultado interessante dessa ação conjunta é a criação dos chamados “agentes” – sistemas computacionais que podem ser do tipo não-inteligente ou cognitivo / inteligente; este podendo interagir, mais efetivamente, com o humano. O sistema “tutor inteligente”, utilizado em ambientes virtuais de aprendizagem, atua como um “agente pedagógico animado” – personagem de natureza “social e antropomórfica” que, por meio de recursos de multimídia, “dialoga” com o aluno, auxiliando-o em seu processo de aprendizagem (JAQUES e VICARI, 2005). Exemplos desse tipo de agente para ambientes em língua portuguesa são: “PAT” (*Pedagogical and Affective Tutor*, BOCCA et al., 2003); “Dóris 3D” (FROZZA et al., 2009); e “Clóvis” – agente para interação com portadores de deficiência auditiva (PEREIRA et al., 2007). Esse tipo de agente pode, por exemplo, (i) verificar possíveis dificuldades do aluno; (ii) responder a questionamentos feitos pelo aluno; (iii) enviar mensagens de boas-vindas ou despedida. A imagem seguinte mostra o agente “Dóris 3D” num ambiente de educação a distância.

⁶ “Síntese de fala” é a geração automática de fala pelo computador, ou a “produção artificial do discurso humano”. In CARNEIRO, M. Processamento da Fala Síntese de diálogos anotados em XML. 2009. Disponível em: < <http://subversion.assembla.com/svn/dspkr/relatorio/artigo.pdf>>.



Figura 1: Agente pedagógico animado “Dóris 3D”.⁷

A empresa Microsoft disponibiliza recursos para configuração de agentes animados que agem e reagem conforme o tipo e objetivo do usuário⁸. Esses programas permitem a configuração em diversos idiomas, com adequação de timbre de voz (masculina ou feminina), sotaque, gestos e expressões faciais, entre outros.

Uma modalidade aperfeiçoada de sistemas inteligentes são os agentes conversacionais - *chatbots* – programas capazes de “conversar” com humanos em linguagem natural. Os *chatbots*⁹ são usados, notadamente, no atendimento ao usuário/cliente e em contextos pedagógicos on line. Segundo Othero e Menuzzi (2005), esse tipo de software exige, do programador, conhecimento profundo da sintaxe e da semântica do idioma usado, em função das infinitas possibilidades de combinações léxicogramaticais, e da interpretação de suas ambigüidades. Exemplos de *chatbots* pedagógicos brasileiros são: “Profa. Elektra” desenvolvido no Centro Interdisciplinar de Novas Tecnologias na Educação – CINTED/UFRGS (LEONHARDT et al., 2003); “CyberPoty” do Centro de Educação Tecnológica do Amazonas – CETAM (ALENCAR e NETTO, 2009), e “Ed”¹⁰ implementado pelo projeto CONPET do Ministério das Minas e Energia.

⁷ Fonte: FROZZA et al., 2009.

⁸ URL: <<http://www.lisisoft.com/free-download-version/84580-microsoft-talking-agent-plugin-for-reminder-2007.html>>.

⁹ O termo se origina da junção das palavras “chat” (conversar) e “robot” (robô). O primeiro chatbot – ELIZA – foi criado no Massachusetts Institute of Technology (MIT), em 1966. O primeiro chatbot para a língua portuguesa foi o Cybelle, criado em 1999.

¹⁰ URL: <<http://www.ed.conpet.gov.br/br/quemsoueu.php>>.



Figura 2: CyberPorty (*chatterbot* do CETAM-EAD)

Outro tipo de aplicativo é o que permite a escrita e leitura em editores de texto e em páginas da internet por deficientes visuais. Alguns desses programas foram desenvolvidos para processarem a Língua Portuguesa, entre os quais:

- DOSVOX (Núcleo de Computação / UFRJ)¹¹: primeiro software desse tipo desenvolvido no Brasil, o programa permite que o deficiente use o computador por meio de um sintetizador de voz;
- BR Braille (Faculdade de Engenharia Elétrica e de Computação / UNICAMP)¹²: programa que faz transcrição de caracteres Braille para caracteres alfanuméricos em português;
- Jaws for Windows¹³: leitor de tela mais utilizado no mundo, possui sintetizador de voz que opera em diferentes idiomas, inclusive o português.

Os sistemas computacionais listados nesta seção são evidências de como a Ciência da Computação tem contribuído para o desenvolvimento de programas voltados para resolução de problemas e facilitação de atividades do cotidiano. Nas palavras de Santaella (1996, p. 204):

¹¹ URL: <<http://intervox.nce.ufrj.br>>.

¹² URL: <<http://www.fee.unicamp.br/deb/brbraille>>.

¹³ O Freedom Scientific <<http://www.freedomscientific.com>>.

Cada vez mais a comunicação com a máquina, a princípio abstrata e desprovida de sentido para o usuário, foi substituída por processos de interação intuitivos, metafóricos e sensório-motores em agenciamentos informáticos amáveis, imbricados e integrados aos sistemas de sensibilidade e cognição humana. Enfim, o próprio computador, no seu processo evolutivo, foi gradativamente humanizando-se, perdendo suas feições de máquina, ganhando novas camadas técnicas para as interfaces fluidas e complementares com os sentidos e o cérebro humano até o ponto de podermos hoje falar num processo de coevolução entre o homem e os agenciamentos informáticos, capazes de criar um novo tipo de coletividade não mais estritamente humana, mais híbrida, pós-humana, cujas fronteiras estão em permanente redefinição.

Outro conjunto de técnicas de apoio à Lingüística Computacional é a “Mineração de Texto” (MT), campo de aplicação oriundo da área “Descoberta de Conhecimento em Textos”. Hearst (2003, minha tradução) explica que a MT é “a descoberta por computador de informação nova, previamente desconhecida, pela extração automática de informação de diferentes fontes escritas”¹⁴. A MT é uma técnica bastante útil na identificação e mapeamento de tópicos e termos em textos eletrônicos embutidos em bases de dados e *corpora* linguísticos. Os recursos usados na MT permitem trabalhar com caracteres, palavras isoladas ou sintagmas, e conceitos.

O mapa abaixo ilustra a proximidade dos vários campos referentes ao tratamento computacional de dados textuais.

¹⁴ No original: “Text Mining is the discovery by computer of new, previously unknown information, by automatically extracting information from different written resources”.

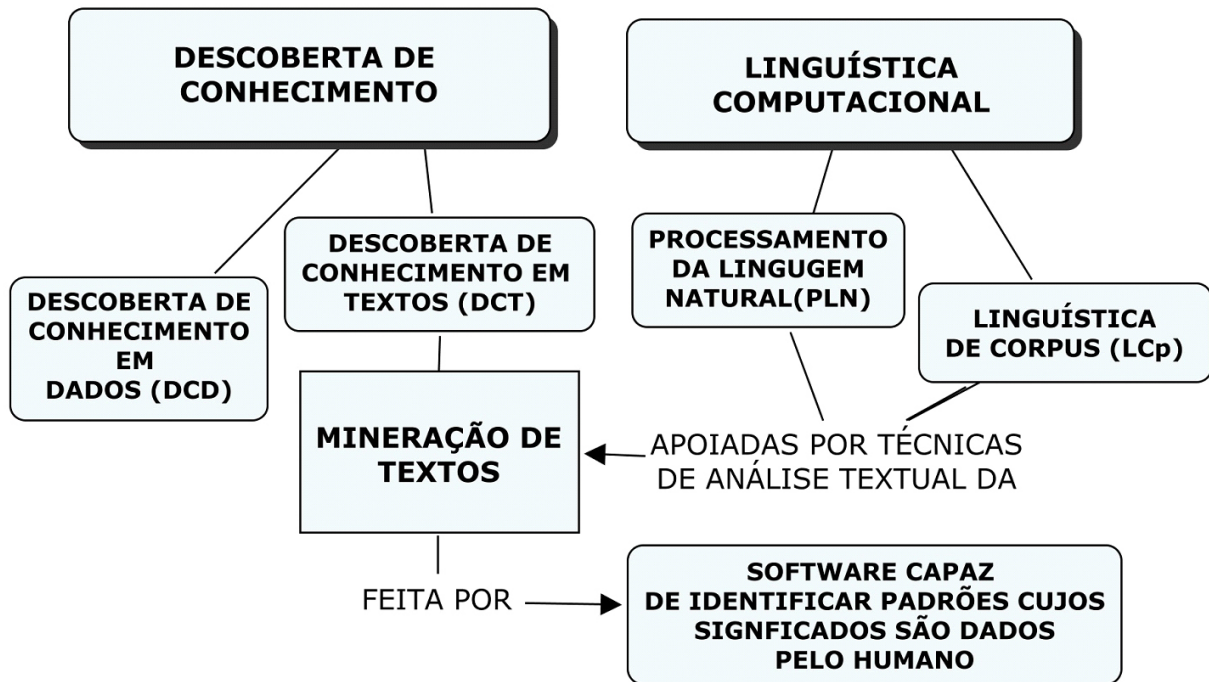


Figura 3: Interfaces no tratamento computacional da linguagem.¹⁵

Com o crescimento do comércio eletrônico e da abertura em larga escala de informações institucionais, as técnicas de MT têm sido utilizadas para melhor conhecer os hábitos e gostos de clientes e usuários, tendência que resultou na sistematização de novas abordagens para análise textual composto por abordagens como: Mineração de Opinião, Análise de Sentimento, Análise da Subjetividade (também conhecida como Análise de Julgamento).

A seção seguinte apresenta projetos voltados para a organização de acervos digitais focados na preservação de diferentes produções textuais.

Corpora lingüísticos e acervos digitais

A seção anterior apontou a possibilidade de se registrar e estudar os fatos lingüísticos em *corpora* digitais. Esses bancos de dados contribuem para a preservação da memória cultural uma vez que constituem registros sincrônicos e diacrônicos da língua falada e escrita.

A Lingüística de Corpus (LCp) permite a verificação e estudos descritivos dos variados gêneros e registros da escrita e da fala. Berber-Sardinha (2004, pg. 3)

¹⁴ Fonte: a autora

explica que a LCp “se ocupa da coleta e exploração de *corpora* ou conjunto de dados linguísticos textuais que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade lingüística”.

A LCp um campo de investigação que tanto se apoia quanto colabora com outras áreas, entre elas: a Fonética, a Informática, as Ciências Cognitivas, a Inteligência Artificial, a Lógica Formal e a Estatística (JURAFSKY e MARTIN, 2000). Com esse suporte interdisciplinar, a LCp costuma ser aplicada na verificação de frequência de palavras, de colocações (em que são verificadas associações entre unidades lexicais e campos semânticos), de concordâncias (padrões de uso e ocorrências), de caracterização de aspectos da fala e escrita de grupos e locais específicos, entre outros.

Vale destacar que esses estudos podem ser investigados em contextos autênticos ou em documentos ficcionais e pré-existentes. À guisa de ilustração, o quadro seguinte apresenta alguns *corpora* construídos em instituições brasileiras.

Quadro 1: Exemplos de Corpora Brasileiros

CORPUS	MODALIDADE	LOCAL
ALERS ¹⁶	Português falado na Região Sul do Brasil	UFRGS / UFSC / UFPR
Banco do Português ¹⁷	Português brasileiro escrito e falado	PUCSP
COMET ¹⁸	Corpus comparativo – textos escritos em português e inglês	USP
CR-LC ¹⁹	Português escrito	USP
NURC ²⁰	Português urbano falado no Brasil	UFRJ
PROFALA ²¹	Português Não-Padrão do Ceará; Português Oral Culto de Fortaleza	UFC
TextQuim ²²	Textos de Química	UFRGS
“Usos do Português”	Português brasileiro escrito	UNESP-ARARAQUARA

¹⁶ ALERS (Atlas Linguístico-Etnográfico da Região Sul): <http://www.ufrgs.br/setordealemao/projetos_pesquisa/cleo_vilson_altenhofen/atlas_linguistico.htm>.

¹⁷ URL: Banco de português: <<http://www2.lael.pucsp.br/corpora>>.

¹⁸ COMET(Corpus Multilíngüe para Ensino e Tradução): <<http://www.fflch.usp.br/dlm/comet>>.

¹⁹ CR-LC (Corpus de Referência Lácio Web): <<http://www.nilc.icmc.usp.br/lacioweb>>.

²⁰ NURC (Projeto de Estudo da Norma Lingüística Urbana Culta): < <http://www.letas.ufrj.br/nurc-rj>>.

²¹ PROFALA: < <http://www.profala.ufc.br/historico.htm>>.

²² TextQuim (Textos de Química): < <http://www6.ufrgs.br/textquim/index.php>>.

Os *corpora* podem ser constituídos de conteúdo lingüístico mais geral ou de terminologia especializada (p.ex. *corpus de termos médicos*). A LCp tem, também, íntima relação com a Lexicografia Computacional para construir dicionários e bases de dados lexicais. Esses bancos visam ao registro das variações lingüísticas de natureza *diatrática* (nas diferentes camadas socioculturais), *diafásica* (nas diversas modalidades e registros), *diatópica* (em diferentes espaços geográficos), *diacrônica* e *sincrônica*. Por sua natureza computacional, os dados encontrados nessas bases podem ser extraídos e manipulados conforme os objetivos do trabalho. Isso permite as atualizações necessárias em função da dinamicidade das línguas.

Outra modalidade de preservação do patrimônio lingüístico são a organização e manutenção de acervos e repositórios digitais com acesso livre a produções de natureza diversa. A partir da possibilidade de digitalização de documentos, as áreas de Arquivologia e Biblioteconomia ganharam novo escopo ao poder disponibilizar e agilizar o acesso a acervos localizados em instituições brasileiras. Este trabalho destaca os seguintes projetos:

- Banco de Dados de Línguas Indígenas do Brasil e de Áreas Adjacentes²³: desenvolvido pelo Laboratório de Línguas Indígenas da Universidade de Brasília, o projeto mantém um banco com material sonoro e escrito.
- Portal Domínio Público do MEC²⁴: criado em 2004, o acervo disponibiliza produções literárias, artísticas e científicas de domínio público ou de divulgação autorizada. A obra completa de Machado de Assis é um destaque desse banco de dados.
- Museu da Língua Portuguesa²⁵: o museu possibilita ao visitante interagir com o acervo por meio de tecnologias de hipermídia (som, imagem, simulações).
- Biblioteca Digital da Fundação Biblioteca Nacional²⁶: acervo público digitalizado de documentação pictórica, sonora e textual, com destaque para originais históricos.
- Rede da Memória Virtual Brasileira²⁷: projeto da Fundação Biblioteca Nacional que disponibiliza acesso aos acervos de instituições nacionais que servem de repositório de patrimônio visual ou textual.
- Biblioteca Digital Brasileira de Teses e Dissertações (BNTD)²⁸ do Ministério

²³ URL: <<http://vsites.unb.br/il/lali/projetos.php>>.

²⁴ URL: <<http://www.dominiopublico.gov.br/Missao/Missao.jsp>>.

²⁵ URL: <<http://biblio.crube.net/?p=954>>.

²⁶ URL: <<http://bndigital.bn.br/apresentacao.htm>>

²⁷ URL: <<http://bndigital.bn.br/redememoria/apresentacao.html>>.

²⁸ URL: <<http://bdtd.ibict.br/>>.

da Ciência e Tecnologia: repositório de produções realizadas nas universidades do país.

- Portal de Periódicos da Capes²⁹: acesso à cerca de 29 mil periódicos nacionais e internacionais e a 130 bases de referência.

Exemplos de iniciativas no setor da imprensa são o acervo digital da revista *Veja*³⁰ e a digitalização do *Jornal do Brasil* pelo site Google com edições de 1890 a janeiro de 1990.

Considerações Finais

O artigo apresentou um panorama das áreas de investigação envolvidas no estudo da linguagem humana, e como esta se presta ao desenvolvimento de programas de utilização prática em diferentes setores e práticas sociais. Com o grande volume de dados em circulação nas redes, a pesquisa *sobre e com* a linguagem e a preservação de documentos torna-se cada vez mais eficaz em função dos sistemas informatizados.

Com o rápido desenvolvimento de aparelhos multimodais portáteis (*smartphones, tablets*), a leitura e escrita apresentam novos questionamentos, incluindo-se, nesse cenário, a difusão do livro digital (*e-book*). Destaca-se, também, o desafio de incorporar as novas modalidades de expressão e interação à prática escolar. Da mesma forma, observam-se os esforços para reproduzir a linguagem natural em sistemas inteligentes – um desafio para a Inteligência Artificial, se considerarmos a complexidade e possibilidade combinatória das estruturas morfossintáticas.

Finalizando, constata-se que da escrita na pedra à tela do computador, a linguagem, como argumenta Perini (1985, p. 15), afirma-se, sempre, como “o mais importante, o mais onipresente dos fenômenos sociais, e um pré-requisito para a existência das sociedades humanas”.

²⁹ URL: < <http://www.periodicos.capes.gov.br/>>.

³⁰ URL: <<http://veja.abril.com.br/acervodigital>>.

Referências

ALENCAR, M.A.S.; NETTO, J.F.M. Cyberpoty: um chatterbot em 3D para interação com usuários de um portal de educação a distância. In: WORKSHOP DE INFORMÁTICA NA ESCOLA, 16., 2010, Belo Horizonte. *Anais...* Belo Horizonte, 2010. p. 1417-1420.

AMARAL, A.; NATAL, G.; VIANA, L. Netnografia como aporte metodológico da pesquisa em comunicação digital. *Revista Sessões do Imaginário*, v. 2, n. 20, 2008.

AUSTIN, J. L. *Quando dizer é fazer*. Porto Alegre: Artes Médicas, 1990.

BAKHTIN, Mikhail. *Os gêneros do discurso*. Estética da Criação Verbal. 3. ed. São Paulo: Martins Fontes, 2000.

BARDIN, L. *Análise de conteúdo*. Lisboa: Edições 70, 1977.

BASTOS, H.P.P.; BERCHT, M. WIVES, L. Análise manual e automática de pistas lexicais de presença social em chat educacional. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 21., 2010, João Pessoa. *Anais...* João Pessoa, PB, 2010.

BENVENISTE, E. *Problemas de lingüística geral I*. 4. ed. Campinas: Pontes, 1995.

BERBER-SARDINHA, T. *Linguística de Corpus*. São Paulo: Manole, 2004.

BOCCA, E.; JAQUES, P.; VICARI, R. Modelagem e implementação da interface para apresentação de comportamentos animados e emotivos de um agente pedagógico animado. *Revista Novas Tecnologias na Educação*, Porto Alegre, v,1, n. 2, 2003.

CAREGNATO, R. C.; MUTTI, R. Pesquisa qualitativa: análise do discurso versus análise de conteúdo. *Texto & Contexto Enfermagem*, p. 15-4, 2006.

FROZZA, R. SILVA, A.K.; LUX, B.; CRUZ, M.K. BORIN, M. Doris 3D: agente pedagógico baseado em emoções. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO – SBIE, 20., 2009, Florianópolis, SC.

GRICE, H.P. *Studies in the Way of Words*. Boston: Harvard University Press, 1989.

HEARST, M. *What is Text Mining?*. 2003. Disponível em:
< <http://people.ischool.berkeley.edu/~hearst/text-mining.html>>. Acesso em: jun.
2010.

HINE, C. *Virtual Ethnography*. Londres: Sage, 2000.

JAQUES, P.A.; VICARI, R.M. Estado da Arte em Ambientes Inteligentes de Aprendizagem que Consideram a Afetividade do Aluno. *Informática na educação*, UFRGS: Porto Alegre, v. 8, n. 1, p. 15-38, 2005.

JURAFSKY, D.; JAMES, H. M. *Speech and language processing: an introduction to natural language processing*. Upper Saddle River, NJ: Prentice-Hall, 2000.

KOCH, I.V. *A interação pela linguagem*. São Paulo: Contexto, 1995.

LEONHARDT, M.D.; CASTRO, D.; DUTRA, R..L.; TAROUCO, L.M.R. Elektra: um chatterbot para uso em ambiente educacional. In: CICLO DE PALESTRAS NOVAS TECNOLOGIAS NA EDUCAÇÃO, 2., 2003, Porto Alegre, RS. *Anais...*

MARCUSCHI, L.A. Gêneros textuais emergentes no contexto da tecnologia digital.

_____; XAVIER, Antônio Carlos (org.). *Hipertexto e gêneros digitais: novas formas de construção de sentido*. Rio de Janeiro: Lucerna, 2004.

_____. *Análise da Conversação*. São Paulo: Ática, 2003.

MORAES, R.; GALIAZZI, M. C. *Análise textual discursiva*. Ijuí: Unijuí, 2007.

OTHERO, G. A.; MENUZZI, S. M. *Linguística Computacional: teoria e prática*. São Paulo: Parábola, 2005.

PANG, B.; LEE, L. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*. v.2, n.1-2, p. 1-135, 2008.

PEREIRA, A.; KNIPHOOFF, M.E.J.; FROZZA, R. Ambiente virtual de aprendizagem

apoiado por um agente pedagógico de acompanhamento para alunos PNEEs Surdos. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 18., 2007, São Paulo. *Anais...*

PERINI, M.A. *Para uma nova gramática do português*. São Paulo: Ática, 1985.

PRIMO, A; RECUERO, R. A terceira geração da hipertextualidade: cooperação e conflito na escrita coletiva de hipertextos com links multidirecionais. *Libero (FACASPER)*, v. 9, p. 89-93, 2006.

SEARLE, J. R. *Os actos de fala*. Coimbra: Almedina, 1981.

SANTAELLA, L. *Matrizes da linguagem e pensamento: sonora, visual, verbal*. São Paulo: Iluminuras, 2001.

SOUZA DE MELLO, J.C.; FROZZA, R. Fred: um agente pedagógico mediador na construção do conhecimento. *Revista Novas Tecnologias na Educação – RENOTE*, Porto Alegre, RS, v.3, n.1, 2005.

SWALES, J.M. *Genre Analysis: English in academic and research settings*. Cambridge: Cambridge University Press, 1990.

VIEIRA, R.; LIMA, V. L. S. *Linguística computacional: princípios e aplicações*". In: NEDEL, L. P. (Org.). *IX Escola de Informática da Sociedade Brasileira de Computação (SBC-Sul)*. Passo Fundo, Maringá, São José. SBC-Sul, 2001.

VIGOTSKI, L.S. *Linguagem e pensamento*. 2. ed. São Paulo: Martins Fontes, 1998.

WITTGENSTEIN, L. *Investigações filosóficas*. São Paulo: Nova Cultural, 1998.