

ReGra - um revisor gramatical para o português do Brasil

por BRUNO BUYS

PARCERIA DA ITAUTEC COM NÚCLEO DE PESQUISA DA USP VIABILIZA CRIAÇÃO DE UM SOFTWARE LINGÜÍSTICO

Há treze anos a Itautec e o Núcleo Interinstitucional de Lingüística Computacional (Nilc) da Universidade de São Paulo mantêm uma parceria de pesquisa, desenvolvimento e mercado na área de processamento computacional da língua portuguesa. Esses anos de trabalho conjunto levaram ao desenvolvimento do ReGra, um programa de revisão ortográfica e gramatical de língua portuguesa, capaz de analisar textos em português, ainda dentro do editor de textos, e apontar incorreções, poupando o usuário do trabalho exaustivo de correção manual.

A parceria começou quando a Itautec buscou ajuda da universidade para a elaboração de uma ferramenta de correção gramatical e ortográfica, a ser embutida em seu processador de textos, chamado Redator. Isso ocorreu em 1993, o que, em termos de desenvolvimento em informática, é muito tempo (só para situar, o MS-Windows estava na sua versão 3.1, a primeira a fazer uso de interfaces gráficas, a internet era um luxo, restrito a usuários corporativos e universitários, e os computadores mais velozes do mercado eram os 486s). Nessa época o mercado para aplicativos de processamento de texto ainda era aberto, uma vez que não havia um produto dominante como o MS-Office de hoje. Dois anos depois, em 1995, o grupo de pesquisa concluiu uma versão do corretor,

disponível para lançamento comercial. Naquela época, a Itautec comercializava o produto com o nome de Redação Língua Portuguesa, um aplicativo completo de revisão. Ele possuía, adicionalmente, uma extensão a ser instalada pelo usuário, que funcionava dentro do Microsoft Word.

A necessidade de um revisor automático de língua portuguesa vem do fato de que não é possível adaptar um revisor programado para outra língua, como o inglês, por exemplo, que tem ferramentas tradicionais de revisão há bastante tempo. Era necessário desenvolver um revisor para o português brasileiro capaz de levar em conta as particularidades da nossa língua.

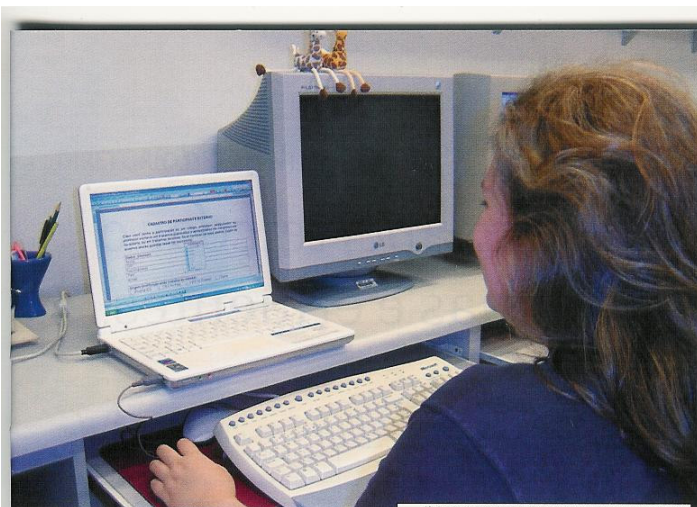
Curiosamente, em relação a esse trabalho, Maria das Graças Volpe Nunes, pesquisadora do Instituto de Ciências Matemáticas e de Computação da USP, e integrante do Nilc, diz que o português não é uma língua particularmente difícil. Contrariando um mito comum de que o português tem um alto grau de dificuldade, Maria das Graças diz que nossa língua apresenta o mesmo grau de complexidade de outras línguas naturais.

FAPESP ENTRA EM CENA

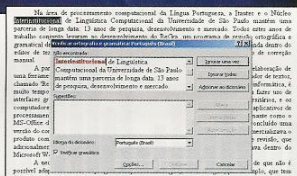
Os anos seguintes seriam de muitos (e bons) acontecimentos na trajetória do ReGra. O primeiro aporte financeiro da

Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp) começou em outubro de 1997, dentro do Programa de Inovação Tecnológica Parceria Universidade-Empresa (Pite). A Fapesp participou com um montante de cerca de R\$ 40 mil reais, para uma contrapartida de R\$ 78 mil da Itautec. Por apostar na qualidade da linha de pesquisa desenvolvida pelo grupo de São Carlos, o financiamento da Fapesp ajudou na sua consolidação dentro da universidade, dando maior visibilidade. Maria das Graças diz que "o nome da Fapesp envolvido nessa parceria deu um peso muito grande, e credibilidade, permitindo a consolidação do grupo de pesquisa. A universidade brasileira ainda está engatinhando na questão da interação com empresas. Hoje temos a Lei da Inovação, mas naquela época isso era uma coisa nebulosa. Quando a Fapesp entrou, a USP passou a ver com outros olhos a nossa parceria", acrescenta a pesquisadora.

Os anos de 1998 e 1999 marcaram uma época de reconhecimento da excelência do produto: o ReGra recebeu os prêmios Peão da Tecnologia, da Fundação Parque de Alta Tecnologia de São Carlos, na edição de 1998, e o Prêmio Inovação Tecnológica, da Alcatel/Brasil, em 1999. Em fins de 1998 a Microsoft iniciou conversações com a Itautec para incorporar o ReGra ao pacote Microsoft



Divulgado



Reprodução

Tela do programa de correção ortográfica

Office, o que viria a acontecer por ocasião do lançamento da versão Office2000. Nesse momento a Itaotec decidiu retirar do mercado seu processador de textos Redator, uma vez que o mercado havia definido o MS-Word como padrão de fato, em processamento de texto. Mas o revisor gramatical ReGra havia atingido um alto grau de excelência em revisão do português, o que atraiu a atenção da multinacional do software, que optou por incluí-lo na versão brasileira do MS-Office2000.

O licenciamento do ReGra pela Microsoft elevou-o à condição de produto hegemônico, sendo vendido embutido e funcionando em conjunto com o Word 2000 e versões posteriores, como solução de processamento de textos em português. Sendo o Word o editor de textos mais usado no Brasil desde então, o ReGra ganhou o mesmo grau de popularidade. A revisão ortográfica e gramatical automática do português brasileiro tornou-se uma prática corriqueira entre usuários domésticos e corporativos.

Um segundo aporte financeiro da Fapesp aconteceu nos anos de 2001 a 2003, com a finalidade de melhorar o desempenho do revisor. Isso envolveria tarefas como reduzir os casos em que a ferramenta intercedia onde não era

necessário (hipercorreção) ou não efetuava uma correção necessária (omissão). Para este segundo projeto a Fapesp contribuiu com R\$ 67 mil, para uma contrapartida de R\$ 216 mil da Itaotec.

A satisfação da Microsoft com o revisor levou às renovações do licenciamento. Segundo Elizabeth Costa, gerente de desenvolvimento de software da Itaotec, "o MS-Office recebe novas versões a cada três anos. Em conjunto com esses lançamentos, uma nova versão do ReGra é lançada, com atualizações e melhorias. Buscamos adaptar nosso produto às novidades que o MS-Office oferece, acompanhando sua evolução".

O FUTURO DO REGRA

O aumento da demanda por software livre tem sido acompanhado de perto pela empresa, que produz e comercializa o Librix, uma solução completa de sistema operacional e programas para desktop, baseada em software livre.

"Nossos planos são continuar com a evolução do ReGra, mantendo a satisfação do nosso grande cliente que é a Microsoft. Estaremos também com o revisor gramatical integrado à suite OpenOffice, para plataformas livres. Eu diria que atualmente o nosso projeto mais arrojado é levar a mesma qualidade que temos com o MS-Office para o OpenOffice, para atender a este segmento de mercado que está demandando plataformas abertas e Linux". No ano que vem, o Librix incluirá o ReGra rodando na suite livre de escritório OpenOffice.org, versão 2.0.

PESQUISA QUE GERA PRODUTO, MAS TAMBÉM ATIVIDADE ACADÊMICA

O Nilc existe desde 1993, tendo iniciado suas atividades com o ReGra, e hoje em dia é uma referência em processamento de linguagem natural (como é chamada a área da computação que dedica-se a estudar linguagens de seres humanos, em oposição às linguagens artificiais, que são as linguagens de programação). Maria das Graças salienta ainda que "por causa do ReGra, dos recursos e da visibilidade obtida, foi possível manter uma equipe de lingüistas e informatas em um grupo de pesquisas que por si só seria muito difícil de manter na universidade brasileira, pela sua própria natureza".

O Núcleo congrega bolsistas, alunos e pesquisadores originários de diversas instituições, em uma área de pesquisa e desenvolvimento que é de alta tecnologia e, inevitavelmente, transdisciplinar. Em seus treze anos de funcionamento, já formou cerca de 30 mestres e 7 doutores em processamento de língua natural. Embora o ReGra seja o projeto mais visível do Núcleo, existem diversos outros projetos, todos orientados a resolver problemas específicos do processamento computacional da língua portuguesa. ●