

Idéia or ideia?

Tools compatible with current versions of the Office Suite are licensed by Techno to Microsoft

Marili Ribeiro

The last breakup occurred in 2009. The introduction of spelling reform left the written Portuguese language without the umlaut, for example. But this had already been done by a private company, far from the university campus. Consolidated and quite successful, the *Automatic Grammar Checker for Portuguese* – developed through a partnership between the Interinstitutional Center for Computational Linguistics of the University of São Paulo (USP) on the São Carlos campus, and Itautec-Philco S.A., with support from FAPESP – is today operated by Techno Software.

“The agreement with Itautec ended in 2008 and since then we have not invested in the software,” says Maria das Graças Volpe Nunes, who has been the project coordinator since its inception in 1993, and who is also a professor in the Department of Science and Computational Linguistics of USP in São Carlos. “I don’t believe the product has evolved since then, because it had already been stabilized within an optimum operating standard, after proofreading adjustments made by the team at Techno Software (an Itautec partner),” she added.

Now the research group at the USP Center, which had an active role in the proj-

ect, continues to invest in various studies related to the computational processing of the Portuguese language, according to the professor. “The Grammar Checking project was undoubtedly a precursor to all the development of this area in Brazil.”

The proofing tools (spelling, grammar, hyphenation, thesaurus) are owned by Itautec and were worked on in conjunction with USP. Upon termination of the agreement with USP and at the request of Itautec, Techno Software took over all product development. “The latest implementations were to adapt them to the spelling reform, and this was done in late 2009. Since then we have provided maintenance on the tools, as well as adding to and correcting the lexicon when necessary,” said Carlos Henrique Ferreira of Techno Software in an e-mail. “Because these tools are licensed to Microsoft, we also make adjustments for purposes of compatibility with current versions of the Office Suite,” he added. The rights to the software produced and marketed by Itautec were acquired by Microsoft, which incorporated the software into the Office 2000 program.

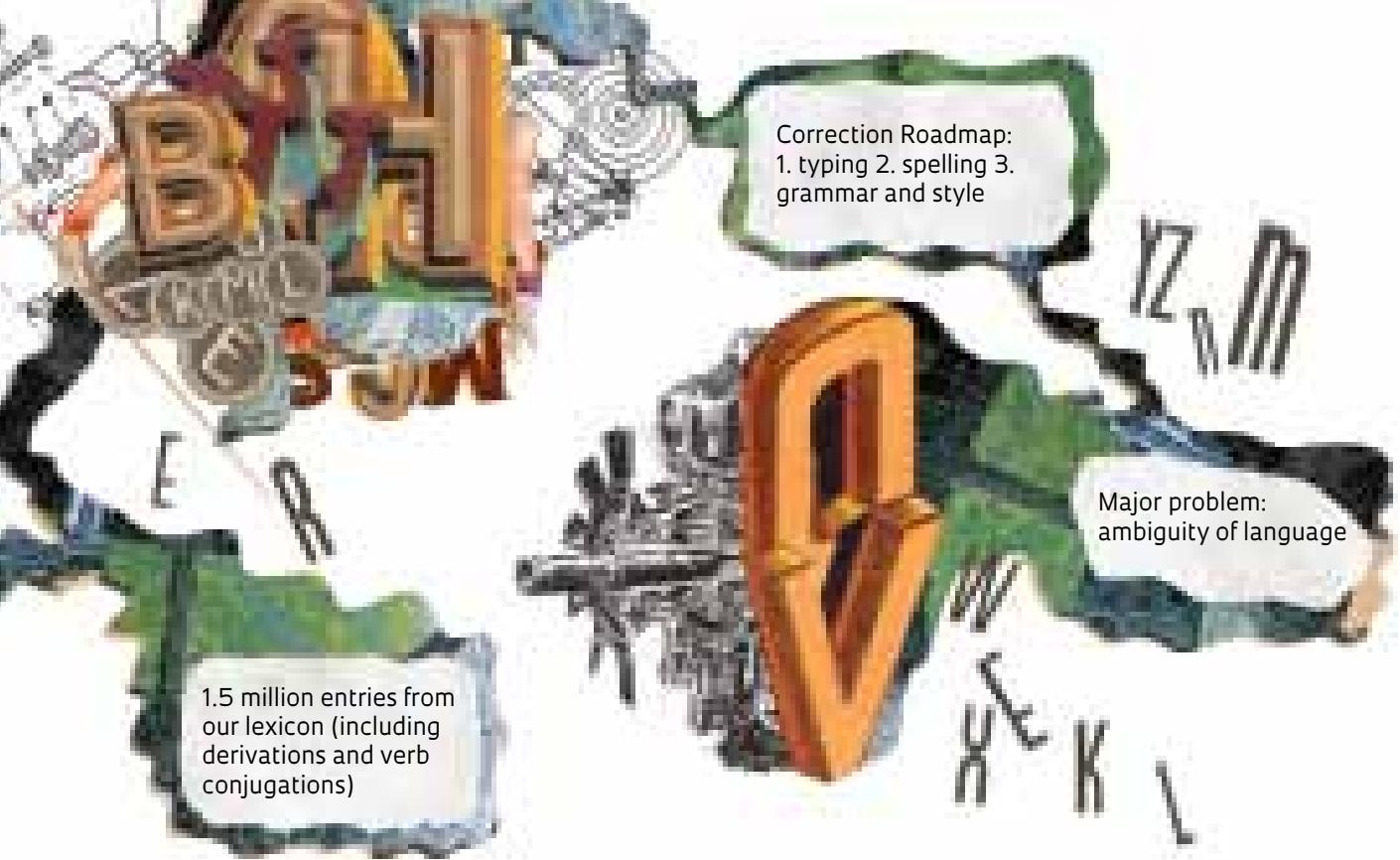
THE BEGINNING

The grammar checking software for Portuguese was born as part of the research

projects in which FAPESP finances from 20% to 70% of the initiative, within the framework of the Support to Research Program in the Partnership for Technological Innovation (PITE), always with an interested counterpart company. The greater the risk of the project, the bigger the role of FAPESP. This line of investment results from a growing awareness that the pace of global technological innovation has accelerated so much that Brazil needs to boost its ability to act like other developed countries. There is no doubt that those who do not have such a capacity to innovate will be marginalized.

In line with this demand, Brazilian companies are testing the effects of partnerships with universities and research institutes. New products and industrial production processes are emerging from this cooperation, with a guaranteed financial return for those betting on this interaction. One way to jump-start the process, which is expanding in different economic sectors, is thanks to the support of PITE.

The grammar-checking project was approved by PITE in 1996, and from then on it gained momentum. Prior to that, the proposal for a grammar checker had been included in the research conducted since 1993 through an agreement be-



1.5 million entries from our lexicon (including derivations and verb conjugations)

tween Itautec-Philco and the Foundation for Support to Physics and Chemistry of São Carlos. The work was performed by a multidisciplinary team of linguists and professionals from the computation area, with the participation of instructors from the Department of Computational Science and Statistics of the Institute of Mathematical and Computational Sciences and the Institute of Physics of São Carlos, and was coordinated on an ongoing basis by Volpe Nunes.

The entry of FAPESP helped to broaden the scope of the research, which has included the collaboration of Professors Claudio Lucchesi, Tomas Kowaltowski and Jorge Stolfi, Computational Institute of Unicamp. In São Carlos, and under the coordination of Volpe Nunes, algorithms were designed and the data bank of words constructed, and in Campinas, the system was compressed and the program's response time decreased.

THE PRODUCT

Itautec began selling the first version of the grammar checker in 1997 under its own name. The product attracted attention on retail shelves, and by the end of that same year, drew the interest of the giant Microsoft, which sought out the company to incorporate the checker

into its Office program, the most widely sold program in Brazil and around the world. For the Portuguese spoken in Brazil, Microsoft was using the old grammar checker created in Portugal, which had 200,000 words. Itautec's already had 1.5 million words. The grammar checker was incorporated into Office 2000, with Itautec licensing the product for a period of three years at a price of US\$421,000.00. New arrangements were subsequently made when the license was renewed.

To develop the product, Itautec spent R\$78,000.00, while FAPESP invested R\$17,900.00, in addition to US\$9,200.00, which was used to purchase machines and equipment for USP. Among the academics involved with the project, none would have predicted how big it would become, especially under the contract with Microsoft.

The grammar checker detects a large number of common mistakes made by users at a second grade level. It is able to spot – and suggest alternatives to – spelling and mechanical errors such as improper placement of punctuation, or absence of it at the end of a sentence. Grammatical errors related to usage of the crasis, syntax, verb and subject agreement, pronoun placement and lexical and other shortcomings are also indicated.

The software does not interfere with style, or standardize the structure of the text, but goes through it sentence by sentence to check its syntactical structure and offer grammatically correct options for construction. The program "guesses" the function according to the proximity of a verb or adjective and suggests corrections for the different syntactical structures arising from each different meaning. ■

PROJECT

Design and implementation of an automatic grammar checker for Portuguese – No. 1997/02608-1 (1997-1998)

COORDINATOR

Maria das Gracas Volpe Nunes – Institute of Mathematical and Computational Sciences (USP)

GRANT MECHANISM

Support to Research Program in the Partnership for Technological Innovation (PITE)

INVESTMENT

R\$17,900.00 and US\$9,200.00 (FAPESP) and R\$78,000.00 (Itautec Philco S.A.)

FROM OUR ARCHIVES

The benefits of a partnership
Issue No. 58 – October 2000

The new directions of technological research
Issue No. 47 – October 1999

A new version of the automatic grammar checker
Issue No. 35 – September 1998