



Gabriela Castelo Branco Ribeiro

**Avaliação de tradução automática no mercado de
localização de software: um estudo de caso**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para
obtenção do grau de Mestre pelo programa de Pós-
Graduação em Letras do Departamento de Letras do
Centro de Teologia e Ciências Humanas da PUC-Rio.

Orientadora: Profa. Maria Carmelita Padua Dias

Rio de Janeiro

Abril de 2006



Gabriela Castelo Branco Ribeiro

**Avaliação de tradução automática no mercado de
localização de software: um estudo de caso**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo programa de Pós-Graduação em Letras do Departamento de Letras do Centro de Teologia e Ciências Humanas da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Profa. Maria Carmelita Padua Dias
Orientadora
Departamento de Letras – PUC-Rio

Profa. Marcia do Amaral Peixoto Martins
Departamento de Letras – PUC-Rio

Profa. Laura Sánchez García
Setor de Ciências Exatas – UFPR

Prof. Paulo Fernando Carneiro de Andrade
Coordenador Setorial do Centro de Teologia
e Ciências Humanas – PUC-Rio

Rio de Janeiro, _____ de _____ de _____

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, da autora e da orientadora.

Gabriela Castelo Branco Ribeiro

Graduou-se em tradução em 1998 na PUC-Rio. Desde então, atua como tradutora e revisora especializada em textos técnicos e comerciais, principalmente das áreas de localização e telecomunicações. É professora do módulo de tradução de informática do curso Formação de Tradutores da PUC-Rio.

Ficha Catalográfica

Ribeiro, Gabriela Castelo Branco

Avaliação de tradução automática no mercado de localização de software: um estudo de caso / Gabriela Castelo Branco Ribeiro; orientadora: Maria Carmelita Padua Dias. – Rio de Janeiro: PUC, Departamento de Letras, 2006.

128 f. ; 30 cm

Dissertação (mestrado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Letras.

Inclui referências bibliográficas.

1. Letras – Teses. 2. Tradução automática. 3. Avaliação. 4. Localização de software. 5. Manual do usuário. 6. Interface com o usuário. I. Dias, Maria Carmelita Padua. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Letras. III. Título.

CDD: 400

Ao Beto, meu pai, com saudade eterna.

À Marilene, minha mãe, com a esperança de que encontre um novo caminho.

Computadores são inúteis. Eles só podem dar respostas.
(Pablo Picasso)

- *E a Máquina de Pensar Bonito? - o Rafael perguntou.*
- *A Máquina de Pensar Bonito está logo depois do medo - disse o velho.*
- *O medo estava na nossa idéia - disse Rafael.*
- *E não está mais - completou o Gustavo.*
- *Quer dizer... - ia falando o velho.*
- *QUE A MÁQUINA DE PENSAR BONITO ESTÁ NA NOSSA CABEÇA? - Todos falaram de uma vez, compreendendo o que o velho queria dizer.*
- *Sim. Na nossa cabeça. A Máquina de Pensar Bonito é...*
- *Nossa inteligência? - disse a Gabriela.*
- *A vontade que a gente sente de fazer as coisas direito? - perguntou a Mônica.*
- *A alegria que faz a gente gostar das pessoas?*
- O velho só balançava a cabeça e dizia:*
- *É isso aí. É isso aí.*

(Do livro infantil *A máquina de pensar bonito contra o medo que o medo faz*, de Carlos Alberto Castelo Branco)

Agradecimentos

Ao Cláudio, meu marido, pelo apoio de sempre e pelo prazer de ter seu amor e sua companhia há dez anos. E à minha filha, Vitória, por me motivar a cada dia com a esperança e o encanto da vida.

A Antonio e Vera, meus sogros, pelo apoio generoso com a Vitória, e pelo amor incondicional que faz dela uma criança mais feliz.

À minha irmã Carla e às minhas sobrinhas Juliana e Fernanda, minhas melhores amigas, especialmente pelo carinho que tiveram com a Vitória durante as minhas horas de estudo.

Aos colegas pesquisadores que compartilharam seus conhecimentos e sua bibliografia com tanta generosidade: Belinda Maia, Luís Sarmiento, Lúcia Specia, Carmen Dayrell e Adriana Rieche.

Aos meus professores por tudo que me ajudam a conhecer, em especial Márcia Martins e Maria Paula Frota, que sempre me ensinam e apóiam desde a graduação, e Helena Martins, pelo brilhante curso de semântica que plantou em mim as sementes, ainda acanhadas, da filosofia e do não-representacionismo.

Aos colegas de curso e do Centro de Linguística Computacional (clic), pelo aprendizado conjunto e pelo carinho, em especial Lia Jucá, Fernando Gil, Maria Alice Antunes, Milena Garrão, Cláudia Freitas e Cláudia Oliveira.

À Carolina Alfaro, colega genial e amiga generosa, pela ajuda desde a prova de seleção até a epígrafe, com direito a caronas com bate-papo pra lá de inteligente.

À amiga Ligia Paixão pela ajuda sempre safa e prestativa. Este trabalho não teria sido possível sem você.

À amiga Natália Botelho pelo apoio emocional e com a bibliografia ainda na época da prova de seleção para o mestrado.

Ao engenheiro de localização Ricardo Júnior, por toda a ajuda técnica, neste trabalho e sempre.

A Marcelo Buzzatti e Lilian Alves, por acreditarem na parceria entre o mercado de trabalho e a universidade.

Ao departamento de Letras, especialmente à secretária Francisca Ferreira, pela ajuda sempre competente e atenciosa ao longo do curso.

À CAPES, pelo apoio financeiro.

Em especial, à minha orientadora Maria Carmelita Pádua Dias, escolhida muito antes do próprio mestrado e do tema de pesquisa. Sua orientação generosa e serena e a confiança no meu trabalho só provaram como a minha escolha foi acertada.

Resumo

Ribeiro, Gabriela Castelo Branco; Dias, Maria Carmelita Padua (Orientadora). **Avaliação de tradução automática no mercado de localização de software: um estudo de caso.** Rio de Janeiro, 2006. 128p. Dissertação de Mestrado – Departamento de Letras, Pontifícia Universidade Católica do Rio de Janeiro.

Este estudo foi motivado pela utilização, ainda em caráter experimental, de um tradutor automático por uma empresa multinacional de localização de software. A fim de contribuir para essa iniciativa pioneira no país, propomos uma avaliação do programa, enfocando as implicações da utilização desta tecnologia no processo de localização de software. Empregamos a taxonomia proposta pelo FEMTI (*Framework for the Evaluation of Machine Translation in ISLE*), desenvolvida especialmente para a avaliação de tradução automática, com base nas normas ISO/IEC de qualidade de software. São considerados aspectos operacionais, como a integração do sistema de tradução automática às ferramentas de memória de tradução, bem como questões relacionadas à linguagem. O *corpus* utilizado para a avaliação foi um manual de usuário de um telefone celular. Além dos problemas lingüísticos recorrentes na maioria das ferramentas de tradução automática disponíveis atualmente, são analisados os desvios relacionados à tradução da interface com o usuário, mais especificamente aos menus do telefone celular. Esses desvios são discutidos dentro das categorias pertinentes da taxonomia do FEMTI e, sempre que possível, foram sugeridas soluções. Para complementar a análise lingüística, apresentamos outros três estudos realizados para o português. Nossos resultados indicam que o sistema pode ser bem-sucedido neste mercado em função principalmente da delimitação do domínio e da adoção dos procedimentos impostos pelo processo de localização. Esse sucesso depende da integração do tradutor automático às memórias de tradução e de investimentos relativamente pequenos na atualização dos recursos lingüísticos (regras gramaticais e dicionários) para refletir as características próprias do domínio e do tipo de texto.

Palavras-chave

Tradução automática, avaliação, localização de software, manual do usuário, interface com o usuário

Abstract

Ribeiro, Gabriela Castelo Branco; Dias, Maria Carmelita Padua (Advisor). **Machine translation evaluation for the software localization industry: a case study.** Rio de Janeiro, 2006. 128p. MSc. Dissertation – Departamento de Letras, Pontifícia Universidade Católica do Rio de Janeiro.

This study was motivated by the trial implementation of a machine translation engine by a multinational software localization company. In order to contribute to this innovative experiment in the Brazilian market, we evaluate the engine, focusing on the implications of its implementation in the software localization industry. We use the FEMTI (*Framework for the Evaluation of Machine Translation in ISLE*) taxonomy, which is based on the ISO/IEC guidelines for software evaluation. Operational aspects, such as the engine's integration with translation memory tools, are taken into consideration, as well as language issues. Our evaluation is based on the machine translated version of a mobile phone user guide. In addition to the language problems common to most machine translation engines currently available, we analyze issues related to the user interface, particularly to the phone menus. These problems are discussed as examples of each related FEMTI topic and we suggest solutions whenever possible. To add to our language evaluation, we present three other studies dedicated to Portuguese. Our results indicate the engine can be successful in this industry mainly in terms of domain restriction and localization workflow procedures. Its success depends on its integration to translation memory tools and requires relatively little investment in updating the language resources (rules and dictionaries) to reflect the language characteristics specific to domain and text type.

Keywords

Machine translation, evaluation, software localization, user guide, user interface

Sumário

| | |
|--|-----|
| 1. Introdução | 11 |
| 1.1. Objetivo | 12 |
| 1.2. Estrutura da dissertação | 13 |
| 2. Contextualização | 14 |
| 2.1. Tradução automática | 14 |
| 2.2. Localização | 22 |
| 2.3. Tradução automática no mercado de localização | 34 |
| 3. Avaliação de tradução automática: teoria e iniciativas | 36 |
| 3.1. Avaliação de tradução automática: teoria | 36 |
| 3.2. A proposta de avaliação do FEMTI | 40 |
| 3.3. Três estudos sobre avaliação para o português | 47 |
| 3.3.1. Aspectos gerais dos estudos | 48 |
| 3.3.2. Categorias de erros | 54 |
| 3.3.2.1. As categorias de <i>Dayrell</i> | 54 |
| 3.3.2.2. As categorias de <i>Oliveira et al</i> | 56 |
| 3.3.2.3. As categorias do <i>TrAva</i> | 59 |
| 3.3.3. Resultados dos estudos | 61 |
| 3.4. Comentários finais | 65 |
| 4. Estudo de caso de tradução automática para o mercado de localização | 67 |
| 4.1. Avaliação com base na taxonomia do FEMTI | 68 |
| 4.2. Casos diversos | 105 |
| 4.3. Acertos do programa | 109 |
| 4.4. Comentários finais | 110 |
| 5. Conclusão | 113 |
| 5.1. Possíveis desdobramentos desta pesquisa | 115 |
| 5.2. Perspectivas para o futuro | 116 |
| 6. Referências bibliográficas | 118 |
| Anexos | 123 |

Lista de tabelas

Tabela 1 – Categorias para classificação dos problemas de morfologia e léxico do *TrAva* (Sarmiento, no prelo) 60

Tabela 2 - Categorias para classificação dos problemas de sintaxe do *TrAva* (Sarmiento, no prelo) 61