Investigação e Desenvolvimento de uma Ferramenta de Pós-Edição de Traduções Automáticas

Felipe Tassario Gomes, Thiago Alexandre Salgueiro Pardo
Instituto de Ciências Matemáticas e de Computação (ICMC), USP/São Carlos

1. Objetivos

Nesse projeto, visou-se realizar um estudo de ferramentas e técnicas de edição de traduções, com o objetivo de criar uma ferramenta de pós-edição de traduções automáticas. A ferramenta consiste em um ambiente que fornece ao usuário facilidades de edição e verificação de traduções, bem como um sistema capaz de detectar erros de tradução e propor soluções automaticamente.

2. Material e método

Foi realizado um estudo sobre Tradução Automática e sobre as técnicas e motivações [1] relevantes para ferramentas de pré e pós-edição de textos [2]. Foram vistos conceitos como memórias de tradução, alinhamentos de textos paralelos, detecção de modificações textuais e métodos de análise e avaliação de ferramentas de edição. Após a fase de pesquisa, foi realizado um levantamento de requisitos com o objetivo de guiar o desenvolvimento da ferramenta, criando uma lista de funcionalidades desejáveis.

Com base na pesquisa inicial e nos requisitos, foi implementado um sistema de memória de correção capaz de armazenar as correções feitas pelo usuário e reproduzi-las automaticamente em situações recorrentes. Após isto, a interface da ferramenta foi implementada, acoplando a memória de correção e outros módulos como corretor ortográfico e um glossário de traduções.

Por fim, uma avaliação do pós-editor foi feita, seguindo-se metodologias de avaliação semelhantes às já existentes na literatura [3] e valendo-se de questionários para verificar as questões de usabilidade da interface.

3. Resultados

Para avaliar a ferramenta, foram dados pares de texto (texto original e uma tradução automática) para diversos usuários, e se analisou como a ferramenta ajuda a detectar erros e a corrigi-los. As medidas computadas foram a precisão em corrigir os erros detectados, a cobertura em detectar todos os erros presentes, e a medida-F como uma medida que combina unicamente a precisão e a cobertura. Os resultados são apresentados na Tabela 1, com os valores médios obtidos pelos usuários, para dois sistemas de tradução automática diferentes.

Tabela 1: Resultados da avaliação

Tradutor Automático	Precisão	Cobertura	Medida-F
Babelfish	0.814	0.863	0.837
Google	0.887	0.923	0.904

4. Conclusões

O ambiente desenvolvido demonstrou-se como uma boa ferramenta para a fase de correção e pós-edição de um texto traduzido automaticamente, com facilidades para auxiliar um usuário tanto em analisar os textos e fazer as correções que julgue necessárias, bem como oferecer a possibilidade de aprender as alterações feitas pelo usuário para poder sugeri-las automaticamente em uma nova edição.

5. Agradecimentos

Às agências de fomento à pesquisa FAPESP, CAPES e CNPq.

6. Referências bibliográficas

[1] Reeder, F; Loehr, D. Finding the Right Words: An Analysis of Not-Translated Words in Machine Translation. In Machine Translation and Information Soup, pp. 256-263 (1998)

[2] Simard, M.; Goutte, C.; Isabelle, P. Statistical Phrase-Based Post-Editing. In the Proceedings of Human Language Technology Conference, NAACL-HLT (2007)

[3] Font Llitjos, A.; Carbonell, J. The Translation Correction Tool: English-Spanish User Studies. In the Proceedings of REC-2004: Fourth International Conference on Language Resources and Evaluation, pp. 347-350 (2004)