

Memórias de Tradução: auxiliando o humano a traduzir

1. Introdução

Devido aos avanços tecnológicos que ocorreram nos últimos anos, em especial, à popularização da internet, principalmente com o surgimento de sites e repositórios de informação multilíngue, a tradução se tornou uma necessidade do público em geral. Contudo, a disponibilidade crescente de informação multilíngue se contrapõe à dificuldade dos usuários, espalhados por todo o mundo, em entender esse conteúdo.

Embora existam, atualmente, diversos sistemas para a tradução automática (TA) comerciais ou disponíveis online – como o Systran (<http://www.systransoft.com/>), o Apertium (<http://www.apertium.org/>) e os tradutores do Google (<http://translate.google.com.br/>) – ainda não é possível alcançar as ambiciosas metas dos primórdios da TA: produzir uma tradução de boa qualidade em domínios irrestritos, por meio de sistemas completamente automáticos. Em contraposição, outras frentes de pesquisa surgiram com o intuito de criar ferramentas para auxiliar o humano na tarefa de traduzir (*Machine-Aided Human Translation*) ou de editar a tradução antes, durante ou depois de sua realização (*Human-Aided Machine Translation*). Na primeira frente estão as ferramentas de Memória de Tradução (MT).

Uma MT pode ser entendida como um repositório de segmentos (geralmente frases ou sentenças) originais acompanhados de suas respectivas traduções humanas, armazenados para que possam ser manipulados e reaproveitados [1]. Com o passar do tempo, a MT agrupa um grande conjunto de segmentos que podem ser recuperados pela ferramenta de MT por meio de um casamento (*matching*) completo ou parcial sempre que um segmento igual ou semelhante aparecer novamente. Assim, uma MT pode ser considerada um *corpus* paralelo com segmentos alinhados de textos já traduzidos [2].

Entre as principais vantagens do uso de MT estão: consistência e agilidade. Ao mesmo tempo que as MTs asseguram consistência na tradução de documentos incluindo terminologia, estruturas, expressões e definições comuns; elas também aceleram o processo de tradução já que dividem a tarefa com o tradutor humano recordando o que já foi previamente traduzido e deixando-o focado na tradução de novos trechos.

Este artigo apresenta uma ferramenta de Memória de Tradução online

desenvolvida com o intuito de auxiliar o tradutor humano a superar as barreiras impostas pelas línguas. Para permitir o entendimento da ferramenta apresentada neste artigo, bem como sua comparação com outras similares, a seção 2 traz um breve relato dos trabalhos relacionados. Em seguida, descreve-se a metodologia empregada no desenvolvimento da ferramenta de MT (seção 3) e seus resultados (seção 4). Por fim, na seção 5, este artigo traz uma breve conclusão e propostas de trabalhos futuros.

2. Trabalhos Relacionados

Ferramentas de MT como OmegaT (<http://www.omegat.org/br/omegat.html>), Anaphraseus (<http://anaphraseus.sourceforge.net/>), Wordfast (<http://www.wordfast.net/>), Trados da IBM (<http://www.trados.com>) e o Déjà Vu da Atril (<http://www.atril.com/>) basicamente funcionam como descrito a seguir.

Primeiramente, o texto sendo traduzido é dividido em segmentos (delimitadores de segmentos usualmente empregados são símbolos de pontuação como ponto final ou dois pontos). Em seguida, cada segmento do texto fonte é apresentado ao tradutor humano para ser traduzido. Cada nova tradução produzida pelo especialista é armazenada na MT juntamente com o segmento original. A partir daí, se o texto original for semelhante a um segmento já traduzido armazenado na MT, a ferramenta recupera a tradução anterior e a indica como sugestão. O tradutor pode utilizar a sugestão, editá-la ou desprezá-la realizando uma tradução completamente diferente. Seja qual for sua decisão, a ferramenta de MT armazena o par (segmento fonte e segmento traduzido) resultante, passando para o próximo segmento do texto original [3].

Diversas funcionalidades podem incrementar esse processo e estão presentes em algumas das ferramentas investigadas, como o uso de glossários, casamento de segmentos similares (e não apenas idênticos), edição manual ou semiautomática dos segmentos e análises estatísticas.

3. Metodologia

Para o desenvolvimento da ferramenta de MT apresentada neste artigo utilizou-se: PHP com Zend Framework (no servidor), Javascript com Dojo Toolkit (nas máquinas dos usuários), HTTP Apache (servidor Web) e PostgreSQL (Banco de Dados, BD). O desenvolvimento realizado com essas tecnologias e seguindo o padrão *Model-View-Controller* garante modularidade e facilita a manutenção da ferramenta.

Para permitir o gerenciamento do processo de tradução, bem como armazenar os pares de segmentos original-traduzido, quatro tabelas foram projetadas e implementadas no BD para: (1) armazenar os dados sobre o projeto de tradução, (2) armazenar os segmentos de uma MT, (3) armazenar os dados da própria MT e (4) relacionar o projeto de tradução com uma ou mais MTs em uso.

A ferramenta de MT descrita neste artigo foi criada como parte do projeto [OMITIDO] e está disponível para a utilização via internet por meio do endereço [OMITIDO], podendo ser utilizada com os principais navegadores modernos populares.

4. Resultados

A Figura 1 traz a tela principal da ferramenta a qual pode ser dividida em três partes principais contendo: (1) duas caixas de texto (fonte e alvo); (2) sugestões recuperadas da MT; e (3) informações e ferramentas/recursos.

Na área (1) há dois textos: o texto fonte sendo traduzido (1a) e o texto alvo com os segmentos já traduzidos (1b). No texto fonte, o segmento que está em tradução é destacado e sua tradução deve ser digitada na caixa de edição na área do texto alvo. Botões de setas abaixo da caixa de edição alvo permitem a navegação entre segmentos, enquanto os botões “Cancelar” e “Salvar” ignoram ou adotam, respectivamente, a tradução realizada até então. Vale ressaltar, também, que assim como as demais ferramentas de MT citadas na seção 2, é possível editar a segmentação do texto fonte feita automaticamente por meio da união e da divisão de segmentos.

Em (2) são apresentados os segmentos similares recuperados das MTs em uso. O usuário pode selecionar um segmento e enviá-lo para caixa de edição alvo por meio do botão “Usar sugestão”. A busca por segmentos similares é feita automaticamente pela ferramenta cada vez que um novo segmento fonte é selecionado. Algumas informações (como andamento do processo de tradução) e ferramentas/recursos auxiliares (por exemplo, glossários) são apresentados em (3). No desenvolvimento da ferramenta foram utilizadas as MTs disponíveis na Linguateca (<http://www.linguateca.pt/>) como o texto utilizado na Figura 1.

Os segmentos traduzidos são armazenados automaticamente no BD assim que um novo segmento é selecionado para a tradução. Com isto é possível utilizar os segmentos recém traduzidos ainda durante a tradução do mesmo documento. Outra característica importante dessa ferramenta é a possibilidade de consultar múltiplas MTs

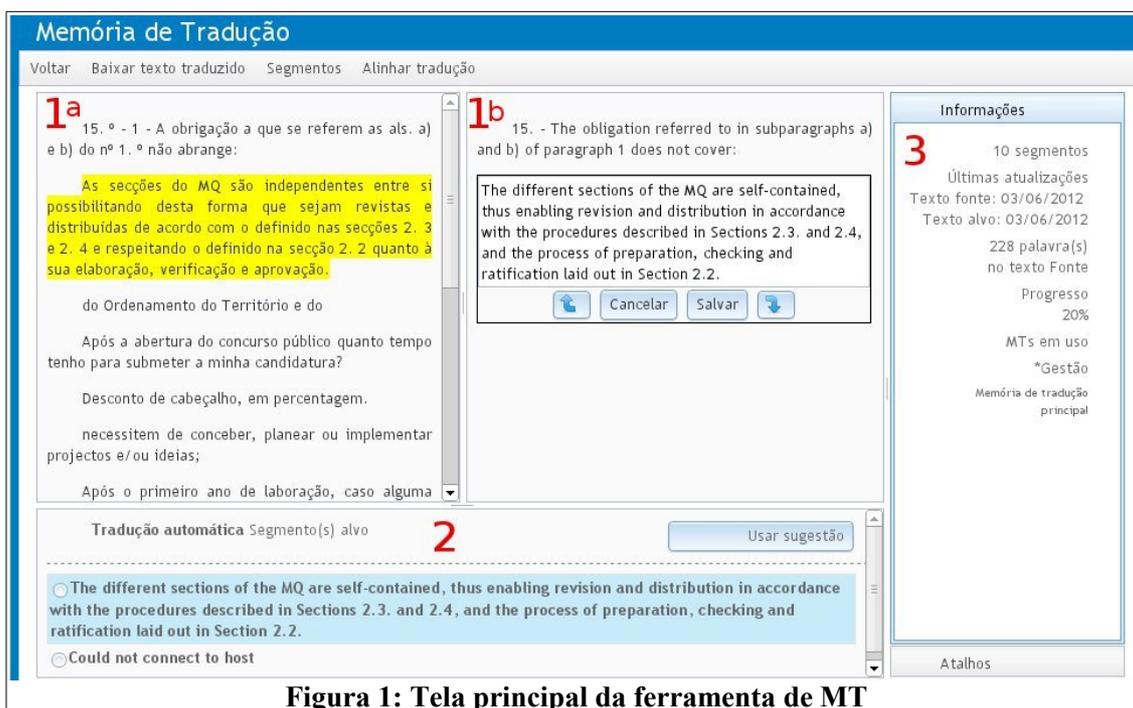


Figura 1: Tela principal da ferramenta de MT

durante a tradução.

Por fim, outra funcionalidade recém implementada é o alinhamento lexical dos textos fonte e traduzido, por meio do alinhador automático GIZA++ [4]. Os pares de palavras alinhadas automaticamente por esta ferramenta podem ser selecionados e incluídos em um glossário bilíngue a ser utilizado em traduções futuras. Essa funcionalidade é um diferencial da ferramenta em relação a outras pesquisadas.

5. Conclusões e Trabalhos Futuros

A ferramenta de MT apresentada neste artigo desempenha as funcionalidades principais esperadas para uma ferramenta deste tipo: segmentação do texto a ser traduzido, recuperação de segmentos similares armazenados na MT e armazenamento dos segmentos traduzidos na MT. Além disso, incorpora uma nova funcionalidade de geração de glossário a partir do alinhamento automático de palavras presentes nos textos original e traduzido.

Esta ferramenta pode ser acessada e utilizada com os navegadores populares e modernos, facilitando seu uso já que não é necessária sua instalação. Além disso, como a ferramenta faz parte de um ambiente maior do projeto [OMITIDO], a mesma pode usufruir de ferramentas presentes neste ambiente como o alinhador lexical citado anteriormente.

Pretende-se, agora, acrescentar novas funcionalidades à ferramenta de MT como o compartilhamento de recursos (memórias de tradução e projetos de tradução) entre usuários e a coleta de informação útil sobre o processo de tradução humana, ou seja, as tarefas de edição que o humano realiza para gerar o segmento traduzido.

Referências

- [1] UNIVERSITY OF LEEDS. School of Modern Languages and Cultures. E-CoLoRe Project. **Ferramentas: Memória de Tradução**. Disponível em: <http://ecolore.leeds.ac.uk/xml/materials/overview/tools/translation_memory.xml?lang=pt>. Acesso em junho 2012.
- [2] NOGUEIRA, DANILO. **Memória de tradução (“CAT”)**. <<http://tradutor-profissional.blogspot.com/2007/09/memria-de-traduo-cat.html>>. Acesso em junho 2012.
- [3] FRACASSI, VAGNER. **Memória de Tradução - Como esses programas podem afetar o seu trabalho**. Disponível em: <<http://www.abrates.com.br/abreartigo.asp?onde=Memoria%20de%20Traducao.abr>>. Acesso em junho 2012.
- [4] OCH, F. J.; NEY, H. A Systematic Comparison of Various Statistical Alignment Models, Computational Linguistics, 2003, número 1, volume 29, p. 19-51.