

Trapezio – Translation Post Editor: um ambiente de pós-edição de traduções automáticas

Felipe Tassario Gomes, Thiago Alexandre Salgueiro Pardo

Núcleo Interinstitucional de Lingüística Computacional (NILC)
Instituto de Ciências Matemáticas e Computação (ICMC)
Universidade de São Paulo (USP) – São Carlos, SP – Brasil

`felipc@grad.icmc.usp.br`, `tasparado@icmc.usp.br`

Abstract. *This paper describes the study and implementation of a translation post-edition tool to help with the edition and revision phase of machine translated texts. This tool offers an interface to assist in the analysis of the texts by human translators, and a memory mechanism capable of learning the corrections made and later suggesting and reproducing these modifications automatically.*

Resumo. *Este artigo descreve o estudo e implementação de uma ferramenta de pós-edição de traduções para auxiliar a fase de edição e revisão de textos traduzidos automaticamente. Esta ferramenta oferece uma interface que auxilia a análise dos textos por tradutores humanos, e um mecanismo de memória capaz de aprender as correções feitas e posteriormente sugerir e reproduzir tais correções automaticamente.*

1. Introdução

A tarefa da Tradução Automática (TA) consiste em produzir automaticamente uma versão de um texto escrito em uma língua natural (língua-fonte) em outra língua (língua-alvo). Esse processo deve preservar o conteúdo do texto na língua-fonte, veiculando-o de forma gramatical na língua-alvo, mantendo-se, preferencialmente, o mesmo estilo de escrita e fluência originais (Specia e Rino, 2002).

A TA é a tarefa mais antiga em Processamento de Línguas Naturais (PLN) e, provavelmente, a primeira aplicação não numérica proposta na área de Computação. Ainda assim, em PLN, a TA é considerada uma das tarefas mais difíceis de serem resolvidas, apresentando, até os dias atuais, resultados insatisfatórios, distantes dos ideais. Isso se deve, principalmente, à quantidade e variedade de conhecimentos necessários para a boa tradução por uma máquina (Arnold et al., 1993). Por outro lado, no contexto de globalização atual, com a ampla disseminação de informações multilíngüe (principalmente na Internet), a TA, mesmo que imperfeita, tem se mostrado essencial para a comunicação e o compartilhamento de dados.

Diante da dificuldade em melhorar a qualidade da TA em análises que envolvem conhecimento humano, a utilização de algumas etapas de revisão manuais pode beneficiar a qualidade da tradução, mesmo mantendo fases da tradução em forma automática. Uma deficiência recorrente da TA ocorre quando a tradução correta de uma frase não pode ser feita sem conhecimento humano implícito sobre a frase, como por

exemplo contexto histórico do texto, ou mesmo conhecimentos humanos explícitos que não podem ser inseridos diretamente no tradutor automático em uso, como por exemplo o uso de um glossário específico ao texto a ser traduzido, problema que ocorre em traduções de textos técnicos que envolvem palavras com sentidos incomuns ou apenas válidos na área de conhecimento do texto em questão.

Tais problemas podem ser corrigidos ou amenizados por etapas de revisão antes ou depois da tarefa de tradução, trazendo a possibilidade de melhorar a qualidade de uma tradução sem envolver a adição de conhecimento lingüístico diretamente ao sistema de tradução. Por exemplo, uma etapa de revisão textual precedente à tradução do texto pode adequar palavras ou expressões do texto em alguma forma previamente conhecida em que já se conhece o resultado que será gerado pelo tradutor automático, e tal resultado é desejado ou adequado para a tradução em questão. Uma etapa de revisão do texto subsequente à tradução automática pode corrigir erros lexicais, gramaticais ou semânticos gerados pelo tradutor, ou realizar substituições de palavras ou expressões por termos técnicos corretos, desconhecidos pelo tradutor automático.

Para a situação de pós-edição apresentada, se tais revisões forem recorrentes, como por exemplo um erro gramatical sempre produzido pelo tradutor automático utilizado, ou um mesmo termo técnico que deve ser substituído várias vezes em um ou mais textos, tal tarefa pode ser automatizada com a ajuda de uma ferramenta. Nesse cenário, a pesquisa proposta aqui se mostra relevante. Diante das imperfeições da tradução automática, da disponibilidade e grande utilidade dos tradutores atualmente, e da possibilidade de melhorar a qualidade da tradução automática, propõe-se um protótipo de uma ferramenta de pós-edição de textos traduzidos automaticamente, chamada Trapézio (*Translation Post-Editor*). Tal ferramenta apresenta ao usuário de um sistema de TA facilidades para revisar e corrigir rápida e eficientemente um texto traduzido automaticamente.

Neste artigo, apresentaremos um estudo de trabalhos relacionados ao desenvolvimento desta ferramenta. Em seguida, apresentaremos a ferramenta desenvolvida, exemplificando suas funcionalidades, seus tipos e estruturas de dados, e levantando suas principais vantagens e limitações. Logo após, serão demonstrados as técnicas de avaliação utilizadas para qualificar a ferramenta, juntamente com seus resultados. Por fim, apresentaremos conclusões tiradas do estudo da área, do desenvolvimento da ferramenta e de sua avaliação, bem como possibilidades de trabalhos futuros relacionados.

2. Trabalhos relacionados

2.1. Memórias de tradução

Uma memória de tradução é um conjunto de pares de expressões entre uma língua fonte e uma língua alvo que representam um segmento de texto na língua fonte e sua correspondente tradução na língua alvo, realizada por um tradutor humano. As memórias de tradução são utilizadas no processo de tradução manual auxiliado por computador, auxiliando os tradutores a traduzir expressões recorrentes mais rápida e precisamente.

As memórias de tradução se mostram especialmente eficazes em trabalhos de traduções que envolvam muitos termos técnicos ou expressões específicas de uma área de conhecimento, onde jargões ou termos importantes costumam repetir-se frequentemente. Além disso, elas são úteis para obter consistência na tradução de um texto longo trabalhado por vários tradutores simultaneamente, garantindo o uso correto de expressões comuns e terminologias. As limitações de memórias de tradução estão ligadas inerentemente ao seu conceito de reutilização de traduções e fragmentos de texto. Por exemplo, a utilização de uma memória de tradução num texto que não apresenta muitas situações de similaridades pode significar um grau de utilização da memória muito baixo, tornando quase todo manual o trabalho de tradução.

Existem vários sistemas de memória de tradução disponíveis comercialmente no mercado. Geralmente, tais sistemas são acoplados a pacotes de soluções no auxílio de tradução, chamados de ferramentas para tradução auxiliada por computadores. Um exemplo de tal software é o Trados¹, disponível em uma licença comercial, ou o OmegaT², ambiente de auxílio à tradução de código aberto. Outras sistemas mais simples não apresentam ambientes de trabalho próprio, e geralmente acoplam-se aos editores de texto tradicionais, estendendo sua funcionalidade ao prover um sistema de memória de tradução. Um exemplo de tal software é o WordFast³, disponível em uma licença grátis e em uma comercial.

No contexto de pós-edição de tradução, como o da ferramenta aqui proposta, as funcionalidades de uma memória de tradução podem ser observadas do ponto de vista das edições feitas através da ferramenta. Em vez de construir um *cópus* bilíngüe entre fragmentos de texto na língua fonte e língua alvo, a ferramenta utiliza um sistema de memória para construir um *cópus* paralelo entre fragmentos de textos incorretos produzidos pelo tradutor automático, e suas correspondentes correções feitas pelo revisor humano. Dessa forma, torna-se possível automatizar o processo de correção de erros recorrentes em um mesmo tradutor automático, ou edições de fragmentos repetidos em um mesmo texto.

2.2. Alinhamento de textos paralelos

O alinhamento de textos paralelos consiste em identificar os segmentos de texto existentes em um texto fonte e relacioná-los com os segmentos de texto no texto alvo que correspondem à sua tradução. O alinhamento de textos pode ser realizado em diversos níveis, como por exemplo no nível lexical, relacionando cada palavra-fonte à sua tradução; no nível sentencial, relacionando as sentenças correspondentes entre o par de textos; ou em níveis de parágrafo, identificando os parágrafos do par de textos relacionados entre si.

Para a pós-edição de um texto já traduzido, é suficiente a presença dos textos original e traduzido em sua forma pura, apenas possibilitando ao revisor visualizar o texto produzido, encontrar erros, e buscar no texto fonte qual os segmentos que ele deverá traduzir corretamente. Entretanto, a realização de um alinhamento entre o par de

¹ <http://www.sdl.com/en/products/products-index/sdl-trados/default.asp>

² <http://omegatplus.sourceforge.net/>

³ <http://www.wordfast.net/>

textos existente é de grande valia, pois esta facilitará o trabalho do tradutor em buscar os segmentos desejados no texto fonte, além de possibilitar à ferramenta armazenar dados sobre quais sentenças estão sendo correta ou incorretamente traduzidas pelo sistema de tradução automática, e extrair informações desses dados. Ultimamente, tais informações podem ser utilizadas pela equipe que produziu o sistema de tradução, para realimentar os dados lingüísticos do tradutor automático e assim melhorar a sua qualidade. Nesta ferramenta, o nível de alinhamento escolhido foi o alinhamento sentencial, por apresentar uma relação relevante entre o par de textos, visto que a maioria dos sistemas de tradução automática realizam suas traduções sentença-a-sentença, e apresentar bons resultados com as técnicas já conhecidas, como analisado e implementado em (Caseli, 2004).

Ao realizar o alinhamento do bi-texto, juntamente com o novo texto obtido, a ferramenta possui agora um alinhamento triplo entre esses fragmentos, como exemplificado na Figura 1, que pode prover mais riqueza de informações tanto para o sistema de memória de tradução utilizado, quanto para o tradutor automático que realizou as traduções.

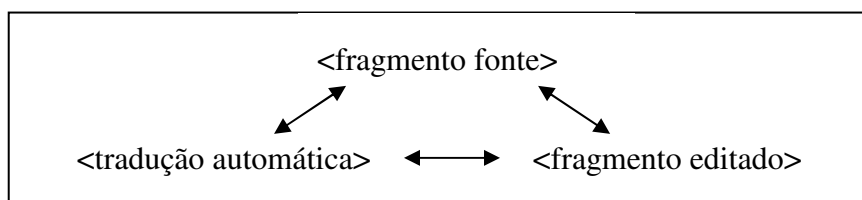


Figura 1. Alinhamento triplo obtido pela ferramenta

2.3. Análise de diferenças textuais

Neste tipo de ferramenta, é necessário detectar as edições feitas livremente pelo revisor, com o objetivo de armazená-las no sistema de memória presente, e de possibilitar a reutilização da mesma edição em fragmentos semelhantes no próprio texto. Para isto, é necessário a utilização de um sistema de análise de diferenças, muito comumente utilizado em contextos de código fonte de sistemas de softwares. Tal sistema possui duas propriedades importantes:

1. dado dois textos, um original e um modificado, o sistema detecta as diferenças entre os dois textos e gere um conjunto de regras de transformações capaz de transformar o texto original no texto modificado, inequivocamente;
2. dado um texto original e um conjunto de regras de transformações gerados a partir dele, o sistema produz novamente o texto modificado.

Mesmo seguindo tais propriedades, para um mesmo par de textos é possível gerar vários conjuntos de regras diferentes que levem do texto original ao texto modificado. Cada sistema de análise de diferença utiliza uma abordagem diferente para a geração dessas regras, de acordo com a forma mais ideal para a aplicação desejada. É possível, por exemplo, gerar transformações de regras em nível de caracteres, que geralmente minimize o número de transformações necessárias. Outra abordagem possível é em nível de linhas do texto, muito utilizada em aplicações que envolvam a visualização de código fonte de software, onde as linhas possuem o significado mais relevante para o texto. Nesta ferramenta, será utilizada uma abordagem de diferenças em

nível mínimo de palavras, estendendo-se a fragmentos de palavras consecutivas, até o limite da sentença. O algoritmo é baseado em (Hunt e McIlroy, 1976) e a implementação possui a capacidade importante de detectar adições, remoções ou modificações de segmentos em uma sentença.

2.4. Ferramenta de correção de traduções

Uma ferramenta de correções de traduções chamada TCTool é apresentada em (Font Llitjós e Carbonell, 2004), com um enfoque de detecção e classificação de erros produzidos pelos tradutores automáticos. Na ferramenta, o usuário edita e realinha as palavras individualmente e, após cada correção ele tenta classificar o tipo de erro que foi corrigido.

Apesar de ter um enfoque diferente, a TCTool é uma das poucas ferramentas de pós-edição encontradas na literatura, e a avaliação da ferramenta aqui apresentada é baseada na avaliação descrita por Font Llitjós, utilizando usuários conhecedores das línguas envolvidas na tradução, porém não conhecedores particulares de lingüística ou lingüística computacional, e obtendo medidas quantitativas baseadas no número de erros e correções detectadas pelos usuários.

3. O ambiente de pós-edição de traduções automáticas

A ferramenta proposta foi desenvolvida utilizando tecnologias web, para ser utilizada a partir de um navegador, sem a necessidade de instalação, com o objetivo de facilitar o alcance dos usuários à ferramenta, e também para possibilitar que vários usuários compartilhem o aprendizado adquirido pela memória da ferramenta.

O sistema consiste em um ambiente que tem como entrada o texto fonte e o texto traduzido automaticamente, ambos em formato de texto puro (sem a necessidade de marcações especiais). Ao entrar com os textos, a ferramenta realizará o alinhamento sentencial e exibirá novamente os dois textos na tela, com todas as funcionalidades de edição disponíveis. A ferramenta possui dois modos de trabalho, o “modo automático”, em que ela sugerirá as correções possíveis que ela detectou no texto e, com a aceitação do usuário, fará as modificações de forma automática, sem a necessidade de intervenção do usuário; e o “modo manual”, em que o usuário pode editar diretamente o texto para corrigir os problemas ainda não detectados, enquanto a ferramenta armazena suas correções para poder passar a detectá-las no futuro.

3.1. Modo automático

No modo automático, a ferramenta faz três operações principais, consultando suas bases de dados e apresentando as possibilidades de modificações ao usuário. Tais operações são:

1. consulta de um sistema de correção ortográfica, para detectar palavras não existentes na língua alvo, e sugerir várias possibilidades de palavras com grafia semelhante;
2. consulta de seu glossário, para detectar a presença de termos técnicos, específicos do texto sendo trabalhado ou termos desconhecidos do tradutor automático, e sugerir a modificação da palavra original pela palavra presente no glossário, como mostrado na Figura 2;

3. consulta de sua memória de tradução, buscando fragmentos do texto alvo presentes no banco de dados da memória, e exibindo ao usuário a possibilidade de modificar o fragmento com a mesma correção realizada anteriormente, quando tal fragmento foi adicionado à memória pela primeira vez.

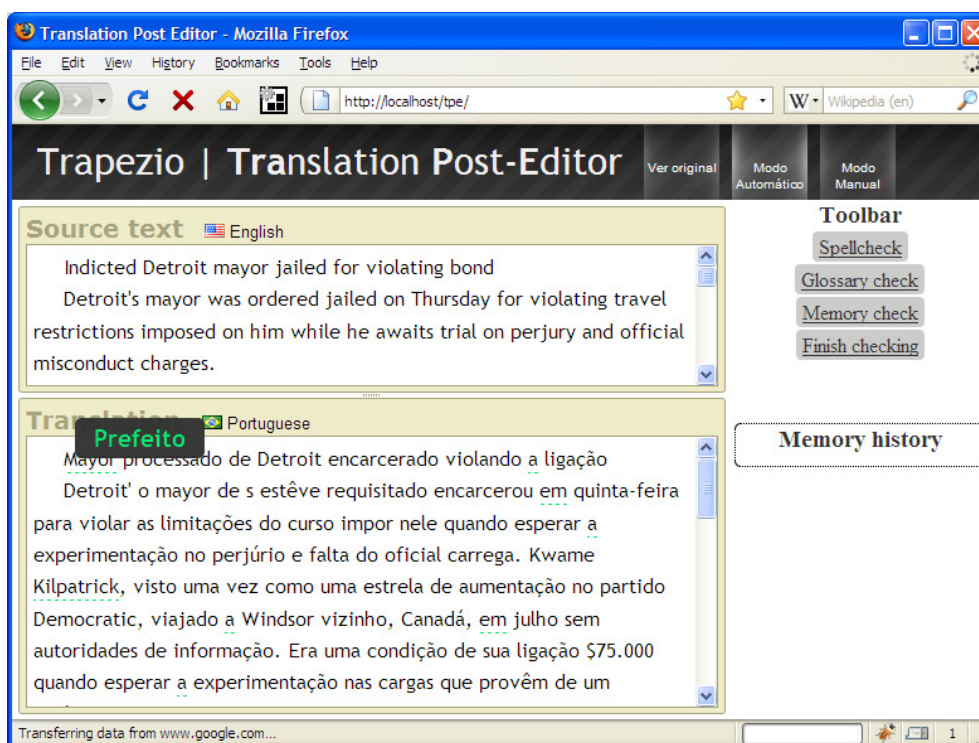


Figura 2. Modo automático, com sugestão do glossário de uma palavra não traduzida

3.2. Modo manual

No modo manual, o usuário pode editar diretamente o texto, fazendo as modificações que forem necessárias para corrigir a tradução, tanto em respeito à glossário, correções ortográficas ou fragmentos incorretos. Após o termino de cada edição, o algoritmo de detecção de diferenças irá analisar quais foram as modificações feitas pelo usuário.

Caso tal modificação seja de uma única palavra, a ferramenta irá perguntar ao usuário se ele deseja adicionar aquela modificação ao glossário ou ao corretor ortográfico, de forma que ela seja corrigida automaticamente da próxima vez que encontrá-la, ou se ele deseja não adicionar tal modificação à base de dados por considerá-la uma modificação pontual. Caso tal modificação seja de um fragmento mais longo de texto, a ferramenta semelhantemente irá perguntar ao usuário se ele deseja adicionar aquela modificação à memória do sistema, ou se ele deseja não adicionar tal modificação à memória. Em ambos os casos, o sistema passará a identificar e reconhecer automaticamente tal modificação, sugerindo a correção presente na memória.

3.3. Aspectos visuais e de interação

A ferramenta proporciona um ambiente de edição de texto familiar ao usuário, proporcionando-lhe liberdade de modificar o texto como de costume no modo manual, ou sugerindo à ele modificações apresentadas visualmente na forma comum dos editores

de texto típicos: apresentando um tracejado embaixo do fragmento sugerido. Tal tracejado pode ser apresentado em três cores, para facilitar a distinção de sugestão de correção ortográfica, glossário ou memória.

No modo manual, a visualização dos textos fonte e alvo são sincronizadas automaticamente a partir das informações do alinhamento sentencial, para facilitar que o usuário encontre as passagens correspondentes no par de textos. Além disso, a edição de texto do usuário é controlada dentro dos limites da sentença, para que ele não edite sentenças de forma a dessincronizar o alinhamento presente e armazenar informações incorretas na memória de tradução. O modo manual está demonstrado na Figura 3.

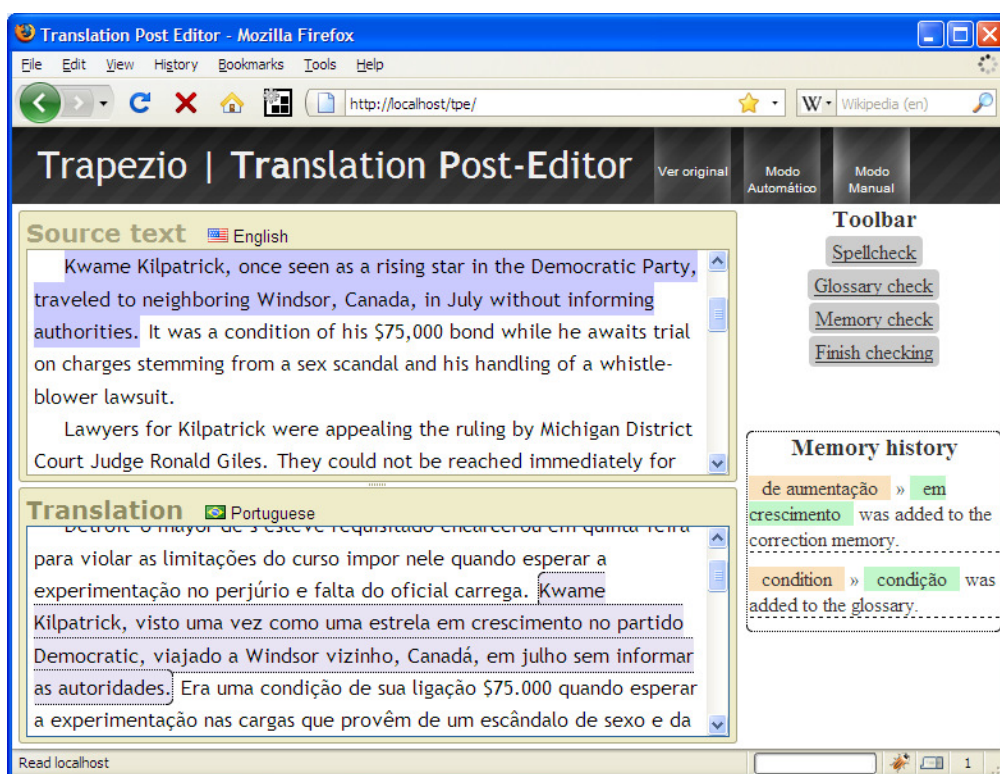


Figura 3. Modo manual, com o cursor de texto editando uma sentença, demonstrando o alinhamento entre os textos e a memória de tradução

Apesar de ser uma ferramenta online, ela possui características de uma aplicação normal, em que todas as interações ocorrem em uma tela específica. Quando é necessário consultar as fontes de dados online do sistema, como o glossário, o corretor ortográfico e a memória, bem como a realização do alinhamento, um indicador de progresso é exibido, de forma que fique claro ao usuário a necessidade de aguardar o termino da análise de seu texto.

As línguas fonte e alvo do par de textos são automaticamente detectadas pela ferramenta, através de um serviço online, eliminando qualquer necessidade do usuário em ter que especificar tal informação. Com isso, um dicionário bilíngüe pode ser acoplado automaticamente e dinamicamente ao sistema, dependendo das línguas envolvidas no par de textos. Além disso, por se tratar de um sistema independente de língua, todos os módulos do sistema já estão preparados para receber caracteres Unicode, e trabalham com a codificação UTF-8, garantindo um funcionamento correto independente da língua utilizada.

3.4. Dados

Todos os dados da memória da ferramenta são armazenados em um banco de dados relacional, em que cada entrada contém informação sobre sua língua, seu tipo de entrada (memória ou glossário), um corpus correspondente (possibilitando à ferramenta trabalhar com vários glossários e memórias diferentes para cada trabalho de tradução), e no caso das memórias, informações sobre o texto perto das proximidades onde o fragmento foi obtido (a sentença completa). Esta última informação não é utilizada diretamente pela ferramenta, mas pode ser de grande valia caso tais dados queiram ser analisados com o intuito de melhorar o sistema de tradução automática que produziu tais resultados.

Sendo armazenados em um banco de dados, os dados podem ser facilmente exportados e importados em diversos formatos, como por exemplo o formato TMX (*Translation Memory Exchange*), padrão mais comum utilizado pelos sistemas de memória de tradução, ou o formato XML:TM, novo padrão que vem surgindo nos novos sistemas de memória de tradução.

3.5. Limitações

Atualmente, o ambiente não apresenta a possibilidade do usuário alterar os alinhamentos gerados, caso ele perceba erro de alinhamento. Este fato, apesar de não atrapalhar o uso da ferramenta para fazer as correções, tanto manuais como automáticas, pode acarretar no registro de alinhamentos errados na memória de tradução da ferramenta, caso uma frase alinhada erroneamente seja corrigida e adicionada à memória pelo usuário.

Além disso, a ferramenta poderia fazer uso de um nível de alinhamento mais sofisticado, como o lexical, abrindo possibilidades maiores de análise e automatização das traduções, utilizando mais informações sobre o texto fonte.

Outra limitação inerente dos princípios da ferramenta está no fato de ela possivelmente exigir muito trabalho manual do revisor no início de seu uso, pois contará com um banco de dados pequenos, ou não relacionado ao texto que ele está trabalhando. Esta característica pode ser em princípio amenizada pelo uso de tradutores automáticos mais sofisticados, que exigiriam um menor número de correções feitas pelo usuário. Esta mudança, entretanto, nem sempre é uma possibilidade ao usuário.

3.6. Disponibilização

O ambiente online para utilização desta ferramenta está disponível para uso da comunidade científica geral através do site do NILC (Núcleo Interinstitucional de Linguística Computacional – <http://www.nilc.icmc.usp.br/nilc/tools/trapezio/>), com o nome de Trapezio (*Translation Post Editor*).

O código fonte da ferramenta, com o código de sua interface e de todos os seus módulos na parte do servidor (corretor ortográfico, glossário, memória e alinhador), está disponível sobre uma licença de software livre, hospedados num site próprio para projetos de código aberto⁴. A interface da ferramenta é produzida em XHTML e JavaScript, e os módulos no servidor possuem partes em PHP, Perl e MySQL.

⁴ <http://code.google.com/p/trapezio>

4. Avaliação da ferramenta

Para avaliar a ferramenta, foram feitos testes com 5 usuários, todos alunos universitários, com o objetivo de avaliar a ferramenta nos quesitos de sua capacidade em auxiliar a tradução de forma automática e manual, e de avaliar suas características de usabilidade. Os testes foram feitos com traduções de textos jornalísticos de inglês para português brasileiro, traduzidos por 2 tradutores automáticos (Babelfish e Google⁵), e todos os usuários testados apresentavam bom domínio de ambos os idiomas.

Para obter medidas quantitativas em relação à capacidade da interface da ferramenta em auxiliar o processo de correção da tradução, em contra-partida à uma correção feita sem auxílio de nenhuma ferramenta específica, pediu-se para que os usuários fizessem a correção de um texto traduzido automaticamente, utilizando-se do mínimo número de correções que ele julgava possível para deixar o texto coerente e gramaticalmente correto na língua alvo. Os textos foram também corrigidos manualmente por um avaliador experiente, e as medidas visavam a comparar o número de correções feitas pelos usuários e o número de correções feitas pelo avaliador experiente, considerando a correção do avaliador como uma correção base. As medidas obtidas foram:

- Precisão: número de erros corrigidos pelo usuário / número de erros corrigidos corretamente;
- Cobertura: número de erros corrigidos pelo usuário / número de erros esperados a serem corrigidos, obtidos pela correção do avaliador;
- Medida-F: $2 * \text{Precisão} * \text{Cobertura} / (\text{Precisão} + \text{Cobertura})$.

A Tabela 1 apresenta a média dessas medidas entre os usuários, analisadas para os dois tradutores automáticos utilizados.

Tabela 1. Resultados da avaliação quantitativa com um texto de 21 sentenças

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

Outra medida quantitativa obtida diz respeito ao desempenho da memória de tradução e do modo automático durante a correção de um texto. A Tabela 2 exhibe os resultados.

Tabela 2. Resultados de medidas quantitativas com um texto de 21 sentenças

Medida	Valor médio
Número de correções sugeridas pelo corretor ortográfico	23
Número de entradas produzidas para a memória de tradução e glossário	40

Após realizada a edição completa de uma tradução de um texto jornalístico, foi inserido na ferramenta um novo texto sobre a mesma notícia, produzido por uma agência de notícias diferentes, de forma a verificar o desempenho da ferramenta com um

⁵ <http://babelfish.yahoo.com> e <http://translate.google.com>

texto semelhante, porém completamente independente do original. O texto foi traduzido com o mesmo tradutor que o texto inicial. Com a memória de apenas um texto traduzido, a ferramenta já foi capaz de sugerir 19 modificações automáticas, relacionadas principalmente à palavras não traduzidas no texto original que foram recorrentes no segundo texto, e à erros gramaticais comuns do tradutor em questão.

Para avaliar a usabilidade da ferramenta, um questionário de usabilidade foi feito para os participantes após a utilização da ferramenta, que avaliaram positivamente o uso da interface, citando que o trabalho da interface tornou mais fácil a comparação das traduções e o trabalho de revisão.

5. 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.

As sugestões oferecidas automaticamente pela ferramenta mostraram-se particularmente eficazes quando utilizadas com textos semelhantes, de forma que a ferramenta ofereceu automaticamente correções para várias palavras não traduzidas pelo tradutor automático utilizada, ou erros recorrentes de tal tradutor, como problemas com gêneros de palavras, construções gramaticais ou preposições. Tal cenário é comum em agências de tradução que se especializam em um tipo de texto (por exemplo, textos jurídicos) ou em situações que um único tradutor automático é utilizado, e seus erros mais frequentes tornam a tarefa de correção mais demorada. Sendo assim, o ambiente de pós-edição pode ajudar o usuário a contornar erros comuns de seu tradutor automático utilizado, como falhas no processo de geração de frases na língua-alvo, ou falta de palavras em seu dicionário interno, em que muitas vezes não é possível para o usuário modificar o glossário de seu tradutor para inserir novas entradas.

Agradecimentos

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

Referências

- Arnold, D.J.; Balkan, L.; Humphreys, R.L.; Meijer, S.; Sadler, L. (1993). *Machine Translation: an Introductory Guide*. Blackwells-NCC, London.
- Caseli, H.M. (2004). *Regras de tradução automática induzidas de textos paralelos envolvendo o português do Brasil*. Monografia de Qualificação. ICMC-USP. São Carlos, Agosto, 67 p.
- Font Llitjos, A.; Carbonell, J. (2004). 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.
- Hunt, J.W, McIlory, M.D. (1976). An algorithm for differential file comparison. In *Computer Science Technical Report*.
- Specia, L. e Rino, L.H.M. (2002). *Introdução aos Métodos e Paradigmas de Tradução Automática*. Série de Relatórios Técnicos do NILC, NILC-TR-02-04. São Carlos, Março, 23p.