

Universidade de São Paulo - USP
Universidade Federal de São Carlos - UFSCar
Universidade Estadual Paulista - UNESP

Anotação manual de erros de tradução automática em textos traduzidos de inglês para português do Brasil

Débora Beatriz de Jesus Martins
Helena de Medeiros Caseli

NILC-TR-13-02

Junho, 2013

Série de Relatórios do Núcleo Interinstitucional de Linguística
Computacional
NILC - ICMC-USP, Caixa Postal 668, 13560-970 São Carlos, SP, Brasil

Resumo

Este relatório detalha o processo de anotação manual de erros de tradução automática (TA) em textos traduzidos de inglês para português do Brasil (en-pt). Para tanto, foram definidas categorias (tipologia) de erros e regras de anotação utilizadas por dois anotadores nativos do português do Brasil, com bons conhecimentos em inglês, para anotar cerca de 1300 sentenças traduzidas automaticamente. As sentenças anotadas formam o *corpus* de treinamento utilizado na construção de uma ferramenta de identificação e pós-edição automáticas de erros TA. As pesquisas em TA, utilizando desde abordagens linguísticas até modelos estatísticos, têm avançado muito desde seu início na década de 1950. Entretanto, os textos traduzidos automaticamente, exceto quando utilizados apenas para um entendimento geral do assunto, ainda apresentam muitos erros e, portanto, precisam passar por pós-edição para que se tornem bem escritos na língua alvo. Atualmente, a forma mais comum de pós-edição é a executada por tradutores humanos, sejam eles profissionais ou os próprios usuários dos sistemas de TA. Esse pós-processamento manual gera custo e demanda tempo. Como uma tentativa de avançar no estado-da-arte das pesquisas em TA envolvendo o português do Brasil, foram criadas categorias de erros e regras de anotação a fim de gerar manualmente uma base para ser aplicada na construção de um sistema de pós-edição automática (*Automated Post-Editing* ou APE) que deverá: (i) identificar automaticamente o que é um erro de TA e (ii) realizar a correção automática do erro previamente identificado.

Índice

1	INTRODUÇÃO.....	1
2	PÓS-EDIÇÃO DA TRADUÇÃO AUTOMÁTICA.....	4
2.1	IDENTIFICAÇÃO DE ERROS NA TRADUÇÃO AUTOMÁTICA.....	6
2.1.1	<i>Ferramenta Blast.....</i>	8
2.2	PÓS-EDIÇÃO DE ERROS DE TRADUÇÃO AUTOMÁTICA.....	10
3	ANOTAÇÃO MANUAL DE ERROS DE TRADUÇÃO AUTOMÁTICA.....	12
3.1	TIPOLOGIA DE ERROS.....	12
3.2	REGRAS DE ANOTAÇÃO.....	16
3.3	PROCESSO DE ANOTAÇÃO E RESULTADOS.....	16
4	CONCLUSÃO.....	19
5	REFERÊNCIAS BIBLIOGRÁFICAS.....	21

Anotação manual de erros de tradução automática em textos traduzidos de inglês para português do Brasil¹

1 Introdução

Este relatório detalha o processo de anotação manual de erros de TA em um *corpus* paralelo inglês-português, bem como as regras de anotação e as categorias (tipologia) de erros definidas para esse processo.

A tradução é o processo que envolve a compreensão de um texto em uma língua fonte e a elaboração de sua representação correspondente na língua alvo desejada. A tradução pode ser executada por humanos, automaticamente, ou uma combinação de ambos. A Tradução Automática (TA) ou *Machine Translation* (MT), é a tarefa de traduzir textos em língua natural automaticamente.

As investigações sobre TA tiveram início na década de 1950, onde a sistematização computacional das classes de palavras descritas nos manuais de gramática tradicional e a identificação computacional de poucos tipos de constituintes oracionais eram os objetos de estudo (SILVA, 2010). Na atualidade, apesar do desenvolvimento de sistemas complexos integrando conhecimento linguístico e extralinguístico, além de abordagens estatísticas avançadas e manipulação de imensos *corpora*² (SILVA, 2010) ainda há vários pontos a serem refinados no estudo da TA, sendo que em alguns idiomas falta avançar mais quando comparados a outros. Entre os idiomas que ainda carecem de mais pesquisas está o português do Brasil.

Os tipos de TA mais conhecidos são:

- SMT (*Statistical* MT ou TA Estatística): utiliza medidas estatísticas a fim de determinar a tradução na língua alvo mais provável de ser a correspondente na língua fonte. Sistemas de SMT têm como base os modelos propostos em (BROWN et al., 1990) e podem ser construídos por *toolkits* como o Moses³ (KOEHN et al., 2007);
- EBMT (*Example-Based* MT ou TA Baseada em Exemplos): usa reconhecimento de padrões para realizar a tradução de parte de cada sentença da língua fonte em

¹ Este trabalho foi desenvolvido com o apoio da FAPESP.

² Um *corpus* (cujo plural é *corpora*) é um conjunto de textos escritos em uma língua. Essa coletânea serve como base de exemplos para aprendizado em diversas áreas de Processamento de Língua Natural (PLN). Quando a coletânea é formada por pares de textos em idiomas diferentes sendo um a tradução do outro, ela recebe o nome de *corpus* paralelo.

³ Disponível em: <<http://www.statmt.org/moses>>. Acesso em: 27 jun. 2013.

seu correspondente na língua alvo. Um exemplo de sistema de EBMT é o Pangloss (BROWN, 1996);

- RBMT (*Rule-Based MT* ou TA baseada em Regras): emprega conhecimento linguístico. De acordo com a definição encontrada em (LAGARDA et al., 2009), os sistemas RBMT possuem dois componentes principais, cujos dados são gerados por linguistas: as regras sintáticas e o léxico, sendo o último composto por informação morfológica, sintática e semântica. Um exemplo de um sistema RBMT é o Apertium⁴ (ARMENTANO-OLLER et al., 2006), disponível para vários pares de idiomas, entre eles o português-espanhol.

Os sistemas de TA podem ser classificados, ainda, segundo a metodologia (ou abordagem) de tradução sendo utilizada: TA direta, TA por transferência ou TA por interlíngua. De acordo com a definição encontrada em (SILVA, 2010):

- TA direta: usa correspondência direta entre unidades lexicais da língua fonte e da língua alvo;
- TA por transferência: realiza a análise sintática da frase na língua fonte e, aplicando regras de transferência sintática, monta a representação sintática na língua alvo;
- TA por interlíngua: utiliza uma “língua” intermediária denominada interlíngua para representar a língua fonte e a partir desta representação efetua a tradução para a língua alvo.

Além das abordagens de tradução completamente automática, outras frentes de pesquisa se desenvolveram com o intuito de criar ferramentas e recursos para auxiliar o humano na tarefa de traduzir (*Machine-Aided Human Translation*, MAHT) ou para editar a tradução antes, durante ou depois de sua realização (*Human-Aided Machine Translation*, HAMT). Nesse sentido podem ser citados os produtos da Trados⁵, da IBM⁶ e o Déjà Vu da Atril⁷. Um dos componentes da MAHT que podem ser aplicados para auxiliar a correção de erros é a memória de tradução (*Translation Memory*, TM). Uma memória de tradução contém partes de textos em seu formato original, ou seja, na língua fonte, e seu correspondente traduzido na língua alvo. Por meio de uma ferramenta de TM, ao realizar uma nova tradução, a sentença de entrada é comparada com a sentença correspondente na língua fonte armazenada na TM juntamente com seu par na língua alvo, a fim de encontrar a tradução.

⁴ Disponível em: <<http://www.apertium.org/>>. Acesso em: 15 out. 2012.

⁵ Disponível em: <<http://www.trados.com/>>. Acesso em: 15 out. 2012.

⁶ Disponível em: <[http://www-01.ibm.com/software/pervasive/ws translation server/](http://www-01.ibm.com/software/pervasive/ws%20translation%20server/)>. Acesso em: 15 out. 2012.

⁷ Disponível em: <<http://www.atril.com/>>. Acesso em: 15 out. 2012.

Recentemente, a pesquisa em TA vivenciou uma mudança de paradigma passando do linguístico (baseado em teorias linguísticas bem definidas que especificam restrições sintáticas, lexicais ou semânticas) para o empírico (o qual utiliza pouca ou nenhuma teoria linguística no processo de tradução). Segundo Hutchins (2005), o paradigma de TA baseada em regras (RBMT) dominava o cenário da TA até a década de 1980 quando as técnicas baseadas em *corpus* ganharam força. Essa mudança pode ser explicada, em parte, pelos avanços de hardware necessários para as abordagens do paradigma empírico e que não estavam disponíveis há alguns anos. Além desse fator, outro de grande relevância para tal mudança de paradigmas é a disponibilidade crescente de recursos como dicionários e *corpora* paralelos envolvendo várias línguas. Assim, sistemas puramente fundamentais como os RBMT têm dado espaço para EBMT e, com grande ênfase nos últimos anos, sistemas SMT.

Em mais de 60 anos de pesquisas em TA foram produzidos diversos sistemas e ferramentas comerciais ou disponíveis on-line, de código aberto ou não, como o Systran⁸, o Apertium⁹ e os tradutores do Google¹⁰. Tais sistemas foram desenvolvidos seguindo várias abordagens como a TA direta ou por transferência, e vários paradigmas como a TA baseada em regras, TA estatística e a TA baseada em exemplos. Porém, mesmo com esses esforços, ainda não foi possível alcançar as ambiciosas metas impostas no surgimento da TA: produzir TA de boa qualidade em domínios irrestritos, por meio de sistemas completamente automáticos. Assim, a tarefa de produzir textos de alta qualidade traduzidos automaticamente ainda é um grande desafio para a área de PLN (Processamento de Língua Natural).

Nesse sentido, este relatório descreve o processo de anotação manual de erros gerados por um tradutor automático estatístico baseado em frases (*Phrase-Based Statistical Machine Translation*, PB-SMT) na tradução de textos escritos em inglês para o português do Brasil. A motivação para esta anotação é a criação de um *corpus* a ser utilizado no treinamento de modelos de aprendizado de máquina capazes de identificar automaticamente erros de TA. Essa identificação automática é o primeiro passo na criação de um APE, um sistema automático de pós-edição da saída tradução. O processo de pós-edição é explicado no Capítulo 2 e os detalhes na anotação do *corpus* – tipologia de erros e regras de anotação – são apresentados no Capítulo 3. Por fim, o Capítulo 4 apresenta algumas conclusões e descreve os passos futuros desta pesquisa.

⁸ Disponível em: <<http://www.systransoft.com/>>. Acesso em: 15 out. 2012.

⁹ Disponível em: <<http://www.apertium.org/>>. Acesso em: 15 out. 2012.

¹⁰ Disponível em: <<http://translate.google.com.br/>>. Acesso em: 15 out. 2012.

2 Pós-edição da Tradução Automática

Apesar de mais de 60 anos de esforços na produção de tradutores automáticos, a qualidade da TA como produto final ainda não atingiu os patamares desejados. Em uma análise de sistemas de TA existentes para o português do Brasil, idioma principal sob investigação nesta proposta, apresentada em (CASELI, 2007), constatou-se que as sentenças incorretamente traduzidas ultrapassavam 50% do total. Os números encontrados em um estudo linguístico recente, detalhado na Seção 3.3, apontam que atualmente persiste o cenário onde ainda há muito a ser melhorado. O estudo linguístico utilizou um tradutor automático que segue a abordagem considerada o estado-da-arte da TA, a PB-SMT. Nesse estudo, cerca de 67% das sentenças traduzidas do inglês para o português do Brasil apresentaram um ou mais erros.

Na atualidade, os textos traduzidos automaticamente, exceto quando utilizados apenas para um entendimento geral do assunto, precisam passar por um processo de pós-edição para que se tornem mais inteligíveis e bem escritos na língua alvo.

O processo de pós-edição da TA é normalmente executado por um tradutor humano que faz modificações no texto de saída da TA com o intuito de transformá-lo em um texto aceitável para o propósito desejado (KRINGS, 2001). Para executar tal tarefa, o tradutor humano faz uma comparação do texto fonte (original) com a saída do tradutor automático (KRINGS, 2001). Assim, enquanto o processo de tradução geralmente tem como entrada apenas um texto escrito na língua fonte a ser transformado na língua alvo, o de pós-edição tem como entrada o texto no idioma original e a saída da TA. Texto original e traduzido automaticamente são, ambos, utilizados pelo tradutor para gerar o texto final que, dependendo do seu uso, precisará ser produzido em um padrão publicável ou apenas deverá estar em um padrão que permita um entendimento geral (O'BRIEN, 2002). Além dessa diferença entre a tradução e a pós-edição da tradução, existem outras. O'Brien (2002), por exemplo, defende que seja incluída a disciplina de pós-edição da TA nos cursos de tradução tradicionais, pois é uma tarefa que tem várias peculiaridades próprias que a diferem da tradução pura e simples.

Deve ficar claro, portanto, que a pós-edição efetuada por humanos é diferente da revisão ou tradução tradicionais. Ao revisar ou traduzir um texto o tradutor trabalha para que o texto produzido fique o mais próximo possível do que é esperado na língua alvo. Já na pós-edição basta que o texto final esteja de acordo com o significado e as regras básicas da língua alvo, ainda que siga mais fielmente a estrutura do texto fonte.

Os erros encontrados por um pós-editor também diferem dos erros tradicionais das traduções geradas manualmente, já que inconsistências nos textos produzidos por um tradutor

humano e um tradutor automático geralmente são diferentes na frequência, repetitividade e tipos. Por exemplo, um tradutor humano pode traduzir mal uma palavra ou estrutura no texto apenas uma vez, enquanto é bem provável que um tradutor automático erre repetidas vezes ao traduzir a mesma palavra ou tipo de estrutura (KRINGS, 2001).

Embora ainda seja uma tarefa tradicionalmente executada por humanos, há várias pesquisas em andamento e ferramentas já criadas para auxiliar e/ou automatizar a tarefa de pós-edição da TA. Essas ferramentas lidam com determinados erros produzidos por diferentes paradigmas de TA.

Assim como outras ferramentas similares propostas na literatura (BÉCHARA et al., 2011, POTET et al., 2011, LAGARDA et al., 2009, SIMARD et al., 2007, GEORGE; JAPKOWICZ, 2005, ELMING, 2006, LLITJÓS, 2007), a ferramenta de APE que se pretende construir a partir de erros anotados manualmente neste trabalho surge como uma alternativa para a pós-edição completamente manual da TA. A ferramenta será construída com base na investigação de técnicas de aprendizado automático (Aprendizado de Máquina, AM) que, aplicadas ao *corpus* anotado com as categorias de erros e seguindo as regras definidas neste documento, dará origem a: (i) modelos capazes de identificar automaticamente um erro e (ii) regras para pós-edição. Desse modo, pretende-se aplicar o conhecimento de “como corrigir a TA” na pós-edição automática ou semiautomática de textos traduzidos automaticamente. Modelos e regras aprendidos a partir do *corpus* anotado darão origem a um sistema de APE inglês-português do Brasil. O tradutor automático usado como *baseline* nesta proposta é o PB-SMT treinado usando o *toolkit* de TA Moses (KOEHN et al., 2007) e o corpus FAPESP-v2 (AZIZ; SPECIA, 2011), doravante denominado TAEIP. Os modelos de tradução e de língua usados no TAEIP podem ser obtidos no Portal de TA, PorTA¹¹.

A ferramenta de APE a ser implementada será modular, resultante de técnicas de AM e, apesar de ser desenvolvida com o foco na correção de erros da TA estatística, poderá ser acoplada a diferentes tipos de sistemas de TA, com o propósito de melhorar a qualidade dos textos traduzidos por eles e reduzindo, assim, o esforço de pós-edição manual. A seguir, as próximas seções descrevem as duas etapas principais do APE: identificação de erros na TA (Seção 2.1) e pós-edição de erros de TA (seção 2.2).

¹¹ Modelos para en-pt disponíveis em Downloads. Disponível em: <<http://www.lalic.dc.ufscar.br/portal/>>. Acesso em: 15 out. 2012.

2.1 Identificação de Erros na Tradução Automática

Uma parte importante da pós-edição é a fase de análise da qualidade da TA para se determinar quais trechos precisam de correção. Essa análise, quando executada antes da pós-edição, evita que uma determinada saída da TA, de boa qualidade ou que não apresente erros do tipo que se deseja atacar, seja desnecessariamente pós-editada. A análise da qualidade da TA pode ser feita, por exemplo, dando-se notas a segmentos traduzidos ou por meio da identificação de erros específicos da TA baseados em categorias pré-definidas.

A identificação de erros em sentenças traduzidas automaticamente pode ser feita manualmente (por humanos) ou automaticamente. Os trabalhos que realizam a identificação automática dos erros podem ser divididos de acordo com a abordagem em:

- Identificação automática de erros com base no alinhamento da saída da TA com a tradução de referência: (POPOVIC, 2011), testado em (POPOVIC; BURCHARDT, 2011), e (ZEMAN et al., 2011);
- Identificação automática de erros com uso de algoritmos de AM: (FISHEL et al., 2012) e (FELICE; SPECIA, 2012).

Este relatório foca a identificação manual de erros de TA realizada com o intuito de criar um *corpus* de treinamento a ser empregado na geração de modelos capazes de identificar automaticamente erros de TA com base em AM, como (FISHEL et al., 2012, FELICE; SPECIA, 2012). Por isso, a seguir são descritos apenas os principais trabalhos de anotação manual dos erros da TA utilizados como base para a proposta de uma tipologia de erros e de regras de anotação empregados neste trabalho.

Em (VILAR et al., 2006) os autores organizam as categorias de erros em uma tipologia composta por vários níveis. O nível mais alto da tipologia contém cinco classes de erro:

1. palavra ausente,
2. ordem de palavra,
3. palavra incorreta,
4. palavra não conhecida e
5. pontuação.

Com exceção da classe pontuação, todas as classes contêm vários subníveis. Para Vilar et al. (2006), a criação das categorias foi motivada pela necessidade de identificar com mais precisão os problemas de um tradutor automático. Para tais autores, os valores gerados pelas medidas mais comumente usadas na avaliação da TA – WER (*Word Error Rate*) (NIESSEN et al., 2000), PER (*Position-independent word error rate*) (TILLMANN et al., 1997), BLEU

(PAPINENI et al., 2002) e NIST (DODDINGTON, 2002) –, não são facilmente relacionados com os erros da saída da TA. As categorias propostas por Vilar et al. (2006) foram adaptadas por (POPOVIC; BURCHARDT, 2011) que, por sua vez, serviram como base para as categorias utilizadas no trabalho aqui descrito.

Outro estudo de avaliação por humanos pode ser encontrado em (CALLISON-BURCH et al., 2007) onde são verificadas escalas distintas para adequação¹² com valores que variam de 1 (nenhuma) a 5 (total). Outro tipo de avaliação proposto realiza o alinhamento automático dos constituintes sintáticos das sentenças fonte com os mesmos constituintes na sentença traduzida usando GIZA++ (OCH; NEY, 2003). Assim, a tradução de cada constituinte é avaliada separadamente com notas de 1 (pior) a 5 (melhor) com relação a outros sistemas de TA. As avaliações feitas por humanos foram posteriormente comparadas a 11 métricas automáticas de avaliação da TA. A comparação foi feita a partir da avaliação humana medindo sua correlação com métricas automáticas usando para isso o coeficiente de Spearman¹³ (SPEARMAN, 1904) e descobriu-se que três métricas menos utilizadas que BLEU obtiveram uma correlação maior quando comparadas às avaliações humanas nesse estudo. Elas são: *semantic role overlap* (GIMÉNEZ; MÁRQUEZ, 2007), ParaEval-recall (ZHOU et al., 2006) e METEOR (BANERJEE; LAVIE, 2005).

Em (FARRÚS et al., 2010), os autores propõem categorias de erros linguisticamente motivadas para que sejam usadas como complemento às avaliações automáticas. As categorias propostas são divididas em níveis linguísticos:

- ortográfico (por exemplo, pontuação e acento),
- morfológico (por exemplo, concordância nominal e verbal),
- lexical (palavras incorretamente traduzidas, não traduzidas, entre outros),
- semântico (polissemia, homonímia, expressões incorretas) e
- sintático (por exemplo, artigo ausente ou extra, erros em preposições).

Concluiu-se, no referido estudo, que a avaliação humana dos níveis lexical e semântico são aparentemente consistentes com as medidas BLEU e TER (*Translation Edit Rate*) (SNOVER et al., 2006) para as quais foi feita essa comparação.

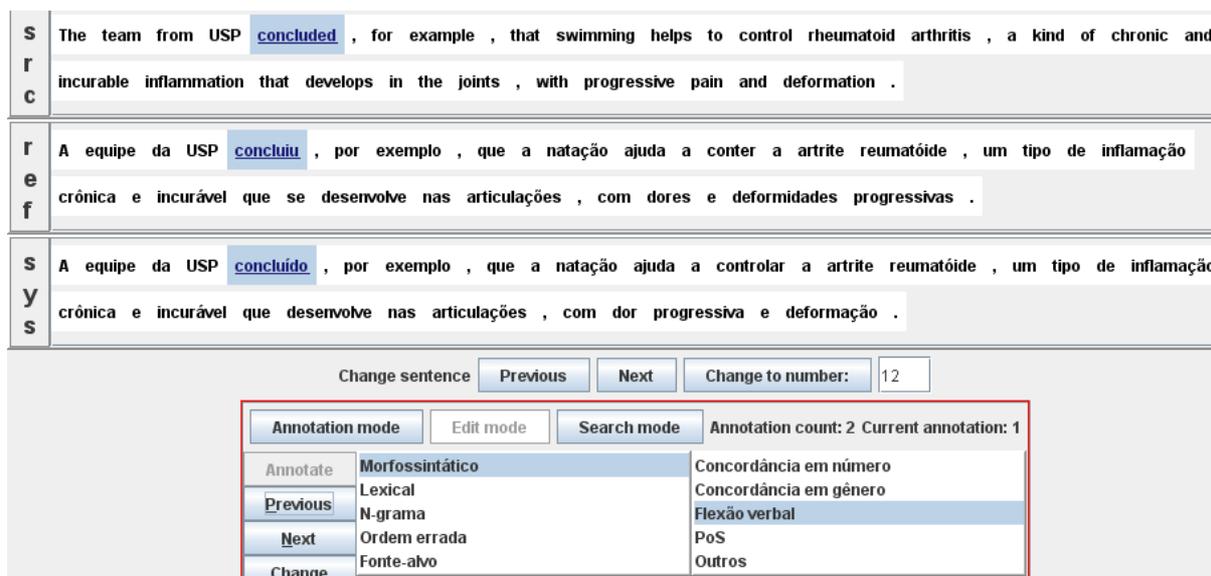
¹²A adequação mede quanto a sentença traduzida automaticamente consegue expressar o que está na referência.

¹³ Spearman é uma das medidas que mede o nível de similaridade entre dois rankings diferentes, calculando assim sua dependência estatística.

2.1.1 Ferramenta Blast

Além de categorias ou notas, existem ainda ferramentas para auxiliar o humano na tarefa de anotação de erro, entre elas a Blast¹⁴ (STYMNE, 2011a), usada no processo de anotação aqui descrito. A Figura 1 a seguir apresenta a tela da ferramenta Blast sendo usada na anotação de um erro de tradução.

Figura 1 – Tela da ferramenta Blast sendo usada na anotação de um erro de flexão verbal.



Para utilizar a Blast são necessários 3 arquivos de entrada contendo:

- sentença(s) fonte (src)
- sentença(s) de referência (ref)
- sentença(s) traduzida(s) automaticamente (sys)

Além de um arquivo contendo as categorias de erros que serão usadas na anotação manual. Essas categorias são carregadas na parte inferior da interface da ferramenta, como apresentado na Figura 1.

Ao final do processo de anotação, um arquivo de saída é gerado contendo blocos de erros anotados com as seguintes informações:

- sentença fonte
- sentença de referência
- sentença traduzida
- posição(ões) da(s) palavra(s) com erro na sentença fonte
- posição(ões) da(s) palavra(s) com erro na sentença traduzida
- posição(ões) da(s) palavra(s) com erro na sentença de referência
- código do erro

¹⁴ Disponível em: <<http://www.ida.liu.se/sarst/blast/>>. Acesso em: 15 out. 2012.

As Tabelas 1, 2 e 3 mostram exemplos de sentenças contendo erros anotados e algumas categorias (todas as categorias são detalhada na Seção 3.1). A primeira linha da tabela representa o src, a segunda o ref, a terceira o sys. Na quarta linha estão as posições das palavras anotadas em src, sys e ref (separadas por vírgula quando há mais de uma palavra), além do código da categoria de erro, todos separados pelo caractere #. Onde não há anotação, a posição da palavra é gravada usando o número -1. Essa é a forma como as anotações ficam armazenadas no arquivo de saída da ferramenta Blast.

Tabela 1 – Exemplo de sentença anotada com erros de TA: flexão verbal, palavra incorretamente traduzida e concordância em gênero.

" It may be that they run faster for having half the weight of the well nourished ones , or for managing to store more energy , since they have spent 23 hours resting and one hour in activity " , says Rosa .
" Pode ser que corram mais por terem metade do peso que os bem nutridos ou por conseguirem armazenar mais energia , já que passam 23 horas descansando e uma hora em atividade " , diz Rosa .
" Pode ser que elas correm mais depressa por terem metade do peso do poço nutriam , ou por conseguirem armazenar mais energia , pois eles passaram 23 horas descansando e uma hora em atividade " , diz Rosa .
6#5#4#morph-verbFlex 15#14#13#lex-incTrWord 16#15#14#morph-verbFlex 28#25#-1#morph-genderConc

Tabela 2 – Exemplo de sentença anotada com erros de TA: palavra incorretamente traduzida e ordem errada.

He swapped the boat for a racing bike , but he still wakes up at half past four in the morning to pedal from 80 to 120 kilometers .
Trocou o barco e o remo por uma bicicleta de corrida e ainda acorda todo dia às 4h30 da manhã para pedalar de 80 a 120 quilômetros .
Ele trocou o barco para uma corrida de bicicleta , mas ele ainda acorda às quatro e meia da manhã para pedalar de 80 a 120 quilômetros .
4#4#6#lex-incTrWord 6,7#6,7,8#8,9,10#ord

Tabela 3 – Exemplo de sentença anotada com erros de TA: palavra ausente e n-grama incorretamente traduzido.

" Little by little , we are revealing the strategies for survival that allow the parasite to do so well in the cells where it shelters " , Célia explains .
" Pouco a pouco estamos desvendando as estratégias de sobrevivência que permitem ao parasita se dar tão bem nas células em que se hospeda " , afirma Célia .
" Aos poucos , estamos desvendando as estratégias de sobrevivência que permitem ao parasita para isso bem nas células onde ela abriga " , afirma Célia .
16,17,18,19#14,15,16#14,15,16,17#grm-incTrGram -1#-1#22#lex-abstWord 24#20#-1#morph-genderConc

Uma das vantagens do uso da Blast é o fato de possuir algumas tipologias de erros facilmente adaptáveis. Outros pontos fortes da ferramenta são sua agilidade na anotação de erros e o fato de permitir a geração de estatísticas referentes aos erros anotados. Outra característica da ferramenta é a gravação dos arquivos anotados em um padrão que pode ser facilmente convertido para tabelas permitindo seu uso posteriormente como parâmetros de entrada em experimentos.

Além da Blast, outras ferramentas foram desenvolvidas com o propósito de pós-editar ou avaliar a saída da TA como a Trapézio (GOMES; PARDO, 2008), a ferramenta de Memória de Tradução do PorTAI (KAWAMORITA; CASELI, 2012) e a PET (AZIZ et al., 2012), entre outras.

2.2 Pós-edição de Erros de Tradução Automática

Os sistemas de pós-edição da TA (APE) podem seguir diferentes abordagens das quais as mais usualmente encontradas são:

- APE baseada em SMT – Nesse caso os sistemas de pós-edição da TA são treinados seguindo os mesmos passos empregados no treinamento de um tradutor estatístico (SMT), só que ao invés de se utilizar duas línguas distintas para fonte e alvo, os textos fonte são resultantes da TA e os textos alvos são as traduções corretas (referência). Exemplos de trabalhos dessa abordagem são: (BÉCHARA et al., 2011), (POTET et al., 2011), (LAGARDA et al., 2009) e (SIMARD et al., 2007). Nos dois primeiros, a TA sendo pós-processada é do tipo SMT e nos dois últimos, é do tipo RBMT.
- APE auxiliada por memórias de tradução – Como já mencionado anteriormente, as memórias de tradução podem ser usadas em um sistema de APE para agilizar o trabalho dos tradutores humanos na revisão de textos traduzidos automaticamente. Exemplos de trabalhos dessa abordagem são: (KRANIAS; SAMIOTOU, 2004) e (GOMES; PARDO, 2008).
- APE usando verificadores gramaticais – Nesse caso a APE é realizada como um processo de correção de gramática da língua alvo. Exemplos de trabalhos dessa abordagem são: (DOYON et al., 2008) e (STYMNE, 2011b).
- APE com AM – Algoritmos (Naive Bayes, árvores de decisão, etc.) e técnicas (como *Transformation Based Learning*, TBL) de AM foram empregadas na resolução de diversos problemas, entre eles a pós-edição de TA. Exemplos de

trabalhos dessa abordagem são: (GEORGE; JAPKOWICZ, 2005), (ELMING, 2006) e (LLITJÓS, 2007).

Para mais detalhes sobre esses diferentes tipos de APE consulte (MARTINS, 2012).

3 Anotação Manual de Erros de Tradução Automática

Com o intuito de criar um *corpus* de treinamento para a geração de um APE en-pt, os textos do conjunto teste-a descrito em (AZIZ; SPECIA, 2011) foram traduzidos de inglês (en) para português do Brasil (pt) usando o TAEIP. Em seguida, a ferramenta Blast foi utilizada para auxiliar no processo de anotação manual de erros de TA.

A anotação de erros foi realizada por dois nativos do português com conhecimento satisfatório do inglês. Por isso, além da especificação da tipologia de erros também foi necessário definir um conjunto de regras de anotação. As categorias de erros e as regras de anotação resultantes serão descritas nas seções a seguir.

3.1 Tipologia de erros

As categorias de erro selecionadas para este estudo foram baseadas em (POPOVIC; BURCHARDT, 2011) que, por sua vez, baseia-se em algumas categorias de (VILAR et al., 2006). Essas categorias também são descritas em (MARTINS et al., 2013). São elas:

A. Erros morfossintáticos (*Inflectional Errors* de Popovic e Burchardt (2011))

Escopo de marcação: engloba apenas uma palavra.

Descrição: a palavra na qual o erro ocorre tem a forma base (lema) correta (comparando-se a TA com a Referência), mas a forma superficial está errada.

Forma de anotação: marcar a palavra incorreta na sentença Fonte, TA e Referência. Na dúvida, não marcar na Fonte ou Referência.

A.1. Concordância em número: erros de concordância de singular e plural.

Exemplo:

Fonte: *The girls went to school.*

Referência: *As garotas foram à escola.*

TA: *A garotas foi à escola.*

A.2. Concordância em gênero: erros de concordância para masculino e feminino. Exemplo:

Fonte: *The girl went to school.*

Referência: *A garota foi à escola.*

TA: *Q garota foi à escola.*

A.3. Flexão verbal: verbos que apresentam conjugação ou forma verbal incorreta.

Exemplo:

Fonte: *The girls went to school*

Referência: *As garotas foram à escola*

TA: *As garotas iam à escola*

A.4. **PoS** (*Part of Speech*): mudança de categoria, por exemplo, “sonhar” traduzido como “sonho”, em que houve mudança de verbo para substantivo.

A.5. **Outros**: erros em traços morfológicos que não se enquadram em nenhuma das subcategorias acima e respeitam a definição dessa categoria: a palavra na qual o erro ocorre tem a forma base correta, mas a forma superficial está errada.

B. Erros lexicais

Escopo de marcação: engloba apenas uma palavra.

Descrição: a palavra na qual o erro ocorre não compartilha a forma base (lema) com nenhuma palavra na Referência.

Forma de anotação: marcar a palavra incorreta na sentença Fonte, TA e Referência. Na dúvida, não marcar na Fonte ou Referência.

B.1. **Palavra extra** (*extra words* de Popovic e Burchardt (2011)): uma tradução na saída da TA que não tem nenhuma correspondência na sentença original (Fonte). Exemplo:

Fonte: *The girl went to school*

Referência: *A garota foi à escola*

TA: *A garota foi a sua escola*

Neste caso é importante sempre olhar para a sentença fonte para determinar se houve inserção de alguma palavra como no exemplo a seguir no qual a TA está correta:

Fonte: *The girl went to her school*

Referência: *A garota foi à escola*

TA: *A garota foi a sua escola*

(TA correta)

B.2. **Palavra ausente** (*missing words* de Popovic e Burchardt (2011)): uma palavra presente na sentença original (Fonte) para a qual a tradução não está presente na saída da TA. Exemplo:

Fonte: *The girl went to school*

Referência: *A garota foi à escola*

TA: *A garota foi escola*

Neste caso é importante sempre olhar para a sentença fonte para determinar se houve remoção de alguma palavra como no exemplo a seguir no qual a TA está correta:

Fonte: *The girl went to school*

Referência: *A garota foi a sua escola*

TA: *A garota foi à escola*

(TA correta)

B.3. Palavra não traduzida (*incorrect lexical choice* de Popovic e Burchardt (2011)): uma palavra presente na sentença original (Fonte) que é mantida igual na saída da TA, ou seja, a palavra não foi traduzida. Exemplo:

Fonte: *The girl went to school*

Referência: *A garota foi à escola*

TA: *A garota foi to escola*

B.4. Palavra incorretamente traduzida (*incorrect lexical choice* de Popovic e Burchardt (2011)): uma tradução correspondente ocorreu, mas está incorreta.

Exemplo:

Fonte: *The girl went to school*

Referência: *A garota foi à escola*

TA: *A garota foi de escola*

Neste caso é importante notar que mesmo diferente da referência, se a saída da TA estiver correta então não se deve marcar o erro, como ocorre no exemplo a seguir:

Fonte: *The girl went to school*

Referência: *A garota foi à escola*

TA: *A garota foi para escola*

(TA correta)

B.5. Palavra com erro de grafia (*incorrect lexical choice* de Popovic e Burchardt (2011)): uma tradução correspondente ocorreu na TA e a escolha da palavra está correta, mas sua grafia não. Exemplo:

Fonte: *The idea was winning the game*

Referência: *A ideia era vencer o jogo*

TA: *A idéia era vencer o jogo*

C. N-grama errado

Escopo de marcação: necessariamente engloba várias palavras, senão o erro se enquadra em lexical (B).

Descrição: o erro engloba várias palavras que formam uma expressão, seja ela semântica ou não.

Forma de anotação: marcar as palavras da expressão na sentença Fonte, TA e Referência. Na dúvida, não marcar na Fonte ou Referência.

C.1. a C.4 - Subcategorias com mesmas definições de B.1 a B.4, só que agora englobam mais de uma palavra na marcação. Exemplo:

Fonte: *The dark horse won the game*

Referência: *O azarão ganhou o jogo*

TA: *O cavalo escuro ganhou o jogo*

D. **Ordem errada** (*reordering errors* de Popovic e Burchardt (2011))

Escopo de marcação: engloba uma ou mais palavras com erros que não se enquadram nas categorias A, B e C.

Descrição: a ordem das palavras está incorreta na sentença traduzida, ou seja, uma sequência de uma ou mais palavras que aparece na TA e na Referência só que em uma posição diferente e errada em relação à que é encontrada na Referência.

Forma de anotação: marcar da primeira até a última palavra com ordem incorreta na sentença Fonte, TA e Referência. Na dúvida, não marcar na Fonte ou Referência.

Exemplo:

Fonte: *It was the same for 1990 hosts*

Referência: *Foi a mesma coisa para as anfitriãs de 1990*

TA: *Foi a mesma coisa para as 1990 anfitriãs*

E. **Erros fonte-alvo**

Escopo de marcação: engloba uma sequência de palavras.

Descrição: a saída da TA está correta, só que não condiz com a sentença fonte.

Forma de anotação: marcar todas as palavras da TA que não condizem com a sentença Fonte.

Exemplo:

Fonte: *The girls went to school*

Referência: *As garotas foram à escola*

TA: *A garota foi à escola*

A TA está correta na língua alvo, portanto não se enquadra na categoria A. Também não é possível identificar qual palavra foi incorretamente traduzida (A, garota, foi). Assim, a melhor solução para esse caso é traduzir novamente e não pós-editar a tradução gerada.

A partir dessas categorias e subcategorias de erros, os dois anotadores humanos realizaram a anotação de cerca de 1300 sentenças seguindo as regras descritas na próxima seção.

3.2 Regras de anotação

Dado que a anotação de erros foi realizada por dois anotadores, algumas regras precisaram ser definidas e seguidas para assegurar as mesmas diretrizes para ambos. São elas:

1. Anotar erros nas categorias C (n-grama errado), D (ordem errada), A (erros lexicais), B (erros morfossintáticos) e E (erros fonte-alvo), nesta ordem¹⁵;
2. Anotar o mínimo necessário que precisaria ser alterado para tornar a saída da TA correta;
3. Não fazer a análise supondo a versão corrigida de um erro previamente anotado, por exemplo em:

Referência: *A casa do meu avô fica em São Paulo*

TA: *A meu tio casa fica em São Paulo*

Dois erros devem ser anotados: um de ordem incorreta (D) da sequência de palavras “meu tio casa” e outro de palavra incorretamente traduzida (B.4) “tio”. Neste caso não se deve marcar o erro de palavra ausente (B.2) para “do” porque se estaria supondo a correção do erro de ordem (D).

4. Podem ser anotadas várias categorias de erro na mesma sequência. Por exemplo em:

Referência: *As garotas foram à escola*

TA: *As garoto foram à escola*

Dois erros devem ser anotados: um de concordância de número (A.1) da palavra “garoto” e outro de concordância de gênero (A.2) da palavra “garoto”.

A partir dessas regras e das categorias e subcategorias de erros descritas na seção 3.1, os dois anotadores humanos realizaram a anotação de cerca de 1300 sentenças conforme detalhado na próxima seção.

3.3 Processo de anotação e resultados

Dois anotadores humanos realizaram a marcação dos erros de TA em uma parte do *corpus* FAPESP (AZIZ; SPECIA, 2011) não utilizada no treinamento do TAEIP: o *teste-a* contendo 1.314 pares de sentenças en-pt. Após a tradução automática das sentenças do *teste-a* usando o TAEIP, as sentenças fonte (src), traduzidas automaticamente (sys) e de referência (ref) foram fornecidas como entrada para a ferramenta Blast para permitir o processo de anotação manual de erros.

No início, os anotadores marcaram juntos 54 sentenças, seguindo as regras definidas

¹⁵ Essa ordem foi escolhida a fim de otimizar o processo de anotação dado que em experimentos anteriores notou-se que algumas categorias são influenciadas por outras.

na Seção 3.2, discutindo-as e adaptando-as caso julgassem necessário. Após essa fase inicial de familiarização com a tarefa de anotação, a ferramenta Blast, as categorias de erros e as regras que deveriam ser seguidas, os anotadores foram separados e marcaram um conjunto idêntico de 126 sentenças. Essa segunda etapa de marcação separada do mesmo conjunto de sentenças foi realizada com o intuito de medir a concordância entre os anotadores, a qual foi calculada em 63%. A concordância entre anotadores foi calculada considerando-se os erros que foram marcados exatamente da mesma forma por ambos os anotadores, ou seja: mesma categoria e mesmas palavras nas sentenças fonte (src), referência (ref) e alvo (sys). Além disso, no cálculo da concordância, uma sentença sem erros foi considerada como uma única instância, enquanto cada erro marcado por pelo menos um anotador foi considerado como uma instância.

Como o valor obtido para a concordância entre anotadores (63%) foi considerado satisfatório dada a complexidade/subjetividade da tarefa e o modo rigoroso como tal concordância foi calculado, a terceira e última etapa do processo de anotação foi realizada dividindo-se as sentenças restantes entre os dois anotadores. Assim, cada um deles anotou metade do total de sentenças restantes, totalizando as 1.314 sentenças do teste-a.

Os erros foram anotados marcando-se a palavra ou sequência de palavras incorreta, de acordo com as categorias A-E (Seção 3.1) e as regras de anotação (Seção 3.2) apresentadas anteriormente. É importante mencionar também que a Blast permite que uma palavra ou sequência de palavras seja categorizada com mais de um tipo de erro, funcionalidade esta que foi explorada no processo de anotação manual aqui descrito.

Das 1.314 sentenças do teste-a de Aziz e Specia (2011) anotadas, 33,10% estavam corretas (sem nenhum erro), ou seja, 66,90% das sentenças traduzidas pelo TAEIP apresentaram um ou mais erros. Das sentenças com erros, 24,27% apresentaram apenas 1 erro, 16,29% apresentaram 2 erros, 12,18% apresentaram 3 erros e assim por diante. O total de sentenças com um certo número de erros pode ser visto na Tabela 4 e os erros por categoria estão detalhados na Tabela 5.¹⁶

¹⁶ Das 1.314 sentenças anotadas, 126 tiveram os erros marcados paralelamente por dois anotadores. Para essas 126 sentenças as anotações de ambos anotadores foram computadas, ou seja, as anotações iguais foram contadas como uma única e as divergentes de cada anotador foram contadas separadamente.

Tabela 4 – Número de sentenças com um certo número de erros anotados por pelo menos um anotador.

Número de Erros	Sentenças
0	435
1	319
2	214
3	160
4	81
5	52
6	21
7	12
8	9
9	5
10	5

Tabela 5 – Total de Erros Anotados por Categoria e Subcategoria anotados por pelo menos um anotador.

	CATEGORIA	Total	%	Total	%
ERROS MORFOSSINTÁTICOS	concordância em número	305	13,94	840	38,39
	concordância em gênero	273	12,48		
	flexão verbal	213	9,73		
	PoS	49	2,24		
	outros	0	0,00		
ERROS LEXICAIS	extra	107	4,89	974	4,52
	ausente	326	14,90		
	não traduzida	217	9,92		
	incorretamente traduzida	310	14,17		
	grafia	14	0,64		
N-GRAMA	extra	0	0,00	173	7,91
	ausente	12	0,55		
	não traduzido	5	0,23		
ORDEM ERRADA	incorretamente traduzido	156	7,13	201	9,19
	ordem errada	201	9,19		
FONTE-ALVO	fonte-alvo	0	0,00	0	0,00

O processo total de anotação das 1.314 sentenças, englobando as etapas 1-3 descritas anteriormente, levou cerca de 4 meses para ser concluído. No total, 2.188 ocorrências de erros foram anotadas por pelo menos um anotador.

4 Conclusão

Este relatório descreveu o processo de anotação manual de erros de TA realizado com base na tipologia de erros e regras de anotação descritas no Capítulo 3. O *corpus* anotado resultante desse processo será utilizado no treinamento de algoritmos de AM com o intuito de realizar a identificação automática de erros, sendo este o primeiro passo na construção de um APE para inglês-português.

Pretende-se, com este trabalho, que novas abordagens envolvendo pós-processamento da TA sejam investigadas com o intuito de aumentar a qualidade dos textos traduzidos automaticamente e, desse modo, possibilite avançar nas pesquisas com pós-edição da TA, principalmente para português do Brasil.

Agradecimentos

Agradecemos à FAPESP e aos membros do NILC Lucas Vinícius Avanço e Profa. Dra. Maria das Graças Volpe Nunes pelo apoio ao desenvolvimento deste trabalho.

5 Referências Bibliográficas

- Armentano-Oller, C.; Carrasco, R. C.; Corbí-Bellot, A. M.; Forcada, L. M.; Ginestí-Rosell, M.; Ortiz-Rojas, S.; Pérez-Ortiz, J.; Ramírez-Sánchez, G.; Sánchez-Martínez, F.; Scalco, M. A. (2006). Open-source Portuguese-Spanish machine translation. In *Proceedings of the 7th International Workshop on Computational Processing of Written and Spoken Portuguese*, pp. 50–59, Itatiaia, RJ, Brazil.
- Aziz, W.; Specia, L. (2011). Fully Automatic Compilation of Portuguese-English and Portuguese-Spanish Parallel Corpora. In *Proceedings of the 8th Brazilian Symposium in Information and Human Language Technology*, pp. 234–238. Cuiabá, MT, Brazil.
- Aziz, W.; Sousa, S. C. M.; Specia, L. (2012). PET: a Tool for Post-editing and Assessing Machine Translation. In *Proceedings of the Eighth international conference on Language Resources and Evaluation (LREC 2012)*, pp. 3982–3987. Istanbul, Turkey.
- Banerjee, S.; Lavie, A. (2005). Meteor: An automatic metric for MT evaluation with improved correlation with human judgments. In *Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization*. Ann Arbor, Michigan.
- Béchara, H.; Ma, Y.; Genabith, J. Van. (2011) Statistical post-editing for a statistical MT system. In *Proceedings of the Thirteenth Machine Translation Summit (MT Summit XIII)*. pp. 308–315, Xiamen, China.
- Brown, P. F.; Cocke, J.; Pietra, S. A. D.; Pietra, V. J. D. (1990). A statistical approach to machine translation. *Computational Linguistics*, pp.79–85, 16(2).
- Brown, R. D. (1996). Example-based machine translation in the Pangloss system. In *Proceedings of the 16th International Conference on Computational Linguistics (COLING 1996)*, pp. 169–174, Copenhagen, Denmark.
- Callison-Burch, C.; Fordyce, C.; Koehn, P.; Monz, C.; Schroeder, J. (2007). (Meta-) Evaluation of Machine Translation. In *Proceedings of the Second Workshop on Statistical Machine Translation (ACL)*, pp. 136–158. Prague, Czech Republic. Disponível em: <<http://www.mt-archive.info/ACL-SMT-2007-Callison-Burch.pdf>>. Acesso em: 30 out. 2012.
- Caseli, H. M. (2007). *Indução de léxicos bilíngues e regras para a tradução automática*. Tese (Doutorado) – USP, São Carlos, São Paulo.
- Doddington, G. (2002) Automatic evaluation of machine translation quality using n-gram cooccurrence statistics. In *ARPA Workshop on Human Language Technology*.
- Doyon, J.; Doran, C.; Means, C. D.; Parr, D. (2008). Automated Machine Translation Improvement Through Post-Editing Techniques: Analyst and Translator Experiments. In *Proceedings of the 8th AMTA conference*. pp. 346–353, Hawaii.
- Elming, J. (2006). Transformation-based correction of rule-based MT. In *Proceedings of EAMT*.
- Farrús, M.; Costa-Jussá, M. R.; Mariño, J. B.; Fonollosa, J. A. R. (2010) Linguistic-based evaluation criteria to identify statistical machine translation errors. In *Proceedings of EAMT*, pp. 52–57. Saint Raphael, France. Disponível em: <<http://www.mtarchive.info/EAMT-2010-Farrus.pdf>>. Acesso em: 30 out. 2012.
- Felice, M.; Specia, L. (2012) Linguistic features for quality estimation. In *Proceedings of the 7th Workshop on Statistical Machine Translation*, pp. 96–103. Montreal, Canada.

- Fishel, M.; Sennrich, R.; Popovic, M.; Bojar, O. (2012). TerrorCat: a Translation Error Categorization-based MT Quality Metric. In *Proceedings of the 7th Workshop on Statistical Machine Translation*. Montreal, pp. 64–70. Canada.
- George, C.; Japkowicz, N. (2005). Automatic Correction of French to English Relative Pronoun Translations using Natural Language Processing and Machine Learning Techniques. In *Computational Linguistics In the North East*. Ottawa, Canada.
- Giménez, J.; Márquez, L. (2007). Linguistic features for automatic evaluation of heterogenous MT systems. In *Proceedings of ACL Workshop on Statistical Machine Translation*.
- Gomes, F. T.; Pardo, T. A. S. (2008). Trapezio - Translation Post Editor: um ambiente de pós-edição de traduções automáticas. In *Anais do Congresso da Academia Trinacional de Ciências (C3N)*, pp.1–10. Foz do Iguaçu, Paraná. Disponível em: <<http://www.icmc.usp.br/taspardo/C3N2008-TassarioPardo.pdf>>. Acesso em: 30 out. 2012.
- Hutchins, J. (2005). Towards a definition of example-based machine translation. In *Proceedings of MT Summit X*, pp. 63–70.
- Kawamorita, C.; Caseli, H. M. (2012). Memórias de Tradução: auxiliando o humano a traduzir. In *Anais do Encontro de Linguística de Corpus (ELC 2012)*.
- Kranias, L.; Samiotou, A. (2004). Automatic Translation Memory Fuzzy Match Post-Editing: A Step beyond Traditional TM/MT Integration. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation*. pp. 331– 334, Lisbon, Portugal. Disponível em: <<http://www.mt-archive.info/LREC-2004-Kranias.pdf>>. Acesso em: 30 out. 2012.
- Koehn, P.; Hoang, H.; Birch, A.; Callison-Burch, C.; Federico, M.; Bertoldi, N.; Cowan, B.; Shen, W.; Moran, C.; Zens, R.; Dyer, C.; Bojar, O.; Constantin, A.; Herbst, E. (2007). Moses: Open Source Toolkit for Statistical Machine Translation. In *Proceedings of the ACL 2007 Demo and Poster Sessions*, pp. 177–180. Prague.
- Krings, H. P. (2001). *Repairing Texts - Empirical Investigations of Machine Translation Post-Editing Processes*. The Kent State University Press.
- Lagarda, A. L.; Alabau, V.; Casacuberta, F.; Silva, R.; Díaz-De-Liaño, E. (2009). Statistical Post-Editing of a Rule-Based Machine Translation System. In *Proceedings of NAACL HLT.*, pp. 217–220. Boulder, Colorado.
- Llitjós, A. F. *Automatic Improvement of Machine Translation Systems*. Tese (Doutorado) – Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, July 2007. Disponível em: <<http://www.mendeley.com/research/automatic-improvement-of-machine-translation-systems/>>. Acesso em: 30 out. 2012.
- Martins, D. B. J. (2012). *Pós-edição automática de textos traduzidos automaticamente de inglês para português do Brasil*. Qualificação (Mestrado) – Programa de Pós-graduação em Ciência da Computação, UFSCar, São Carlos.
- Martins, D. B. J.; Avanço, L. V.; Nunes, M. G. V. (2013). Annotating translation errors in Brazilian Portuguese automatically translated sentences: first step to automatic post-edition. In *Corpus Linguistics Conference, 2013*, Lancaster, UK. *Proceedings of the Corpus Linguistics Conference (CL2013)*, 2013. p. 189-192.
- Niessen, S.; Och, F. J.; Leusch, G.; Ney, H. (2000). An evaluation tool for machine translation: Fast evaluation for machine translation research. In *Proceedings of the*

- Second International Conference on Language Resources and Evaluation (LREC)*., pp. 39–45. Athens, Greece.
- O'Brien, S. (2002). Teaching Post-editing: A Proposal for Course Content. (2002) In *Sixth EAMT Workshop "Teaching machine translation"*, pp. 99–106. Manchester, England.
- Och, F. J.; Ney, H. (2003). A Systematic Comparison of Various Statistical Alignment Models, *Computational Linguistics*, pp. 19–51, 29(1).
- Och, F. J.; Ney, H. (2004). The alignment template approach to statistical machine translation. *Computational Linguistics*, pp. 417–449, 30(4).
- Papineni, K.; Roukos, S.; Ward, T.; Zhu, W.-J. (2002). BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual meeting of the Association for Computational Linguistics*, pp. 311–318. Philadelphia. Disponível em: <<http://www.mt-archive.info/ACL-2002-Papineni.pdf>>. Acesso em: 30 out. 2012.
- Popovic, M. (2011). Hjerson: An Open Source Tool for Automatic Error Classification of Machine Translation Output. *The Prague Bulletin of Mathematical Linguistics*, pp. 59–67, 96.
- Popovic, M.; Burchardt, A. (2011). From Human to Automatic Error Classification for Machine Translation Output. In *Proceedings of the 15th Conference of the European Association for Machine Translation*, pp. 265-272. Leuven, Belgium.
- Potet, M.; Esperança-Rodier, E.; Blanchon, H.; Besacier, L. (2011). Preliminary Experiments on Using Users' Post-Editions to Enhance a SMT System. In Forcada, M. L.; Depraetere, H.; Vandeghinste, V. (Ed.). *Proceedings of the 15th conference of the European Association for Machine Translation (EAMT 2011)*. pp. 161–168, Leuven, Belgium.
- Silva, B. C. D. (2010). *Processamento Automático de Línguas Naturais*. Apostila. Araraquara, São Paulo, Brazil.
- Simard, M.; Goutte, C.; Isabelle, P. (2007). Statistical Phrase-based Post-editing. In *Proceedings of NAACL HLT*., pp. 508-515. Rochester, NY.
- Snover, M.; Dorr, B.; Schwartz, R.; Micciulla, L.; Makhoul, J. (2006). A study of translation edit rate with targeted human annotation. In *Proceedings of AMTA*., pp. 223-231. Cambridge, Massachusetts, USA. Disponível em: <<http://www.mtarchive.info/AMTA-2006-Snover.pdf>>. Acesso em: 30 out. 2012.
- Spearman, C. (1904). The proof and measurement of association between two things. *American J. Psychol.*, pp. 72-101, 15.
- Stymne, S. (2011a). BLAST: A Tool for Error Analysis of Machine Translation Output. In *Proceedings of the ACLHLT 2011 System Demonstrations*., pp. 56-61. Portland, Oregon.
- Stymne, S. (2011b). Pre- and Postprocessing for Statistical Machine Translation into Germanic Languages. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Student Session*. pp. 12–17, Portland, Oregon.
- Tillmann, C.; Vogel, S.; Ney, H.; Zubiaga, A.; Sawaf, H. (1997). Accelerated DP based search for statistical translation. In *European Conf. on Speech Communication and Technology*, pp. 2667-2670. Rhodes, Greece.
- Vilar, D.; Xu, J.; D'haro, L. F.; Ney, H. (2006). Error analysis of statistical machine translation output. In *Proceedings of the fifth International Conference on Language Resources and Evaluation*. pp. 22-28. Italy.

- Zeman, D.; Fishel, M.; Berka, J.; Bojar, O. (2011). Addicter: What Is Wrong with My Translations? *The Prague Bulletin of Mathematical Linguistics*, pp. 79-88, 96.
- Zhou, L.; Lin, C.-Y.; Hovy, E. (2006). Reevaluating machine translation results with paraphrase support. In *Proceedings of EMNLP*.