

3

Avaliação de tradução automática: teoria e iniciativas

Este capítulo está organizado da seguinte forma: primeiramente apresentamos os principais conceitos referentes à avaliação de tradução automática segundo Sarmiento *et al* (no prelo) e algumas considerações de Hovy *et al* (2002). Em seguida, comentamos a taxonomia proposta pelo FEMTI para a avaliação de TA, que será aplicada em nosso estudo de caso. Finalmente, apresentamos três estudos sobre avaliação de TA para o português, descritos em Dayrell (1999), Oliveira *et al* (2000) e Sarmiento (no prelo) e Maia&Barreiro (no prelo).

3.1.

Avaliação de tradução automática: teoria

Com a retomada da tradução automática no final dos anos 80, a avaliação nessa área vem recebendo um grande destaque. Além do clássico ALPAC, de 1966, comentado no capítulo 2, e das avaliações financiadas pela DARPA (*United States Defense Advanced Research Projects Agency*), realizadas entre 1992 e 1994, algumas outras iniciativas de avaliação mencionadas por Hovy *et al* (2002) são o projeto europeu EAGLES (*Expert Advisory for Language Engineering Standards*), iniciado em 1993 e dedicado à avaliação de processamento de linguagem natural em geral, e a campanha norte-americana do NIST (*United States National Institute of Standards and Technology*), realizada entre 2001 e 2002, voltada para a avaliação humana e automática de TA. A taxonomia do FEMTI utilizada na nossa pesquisa foi elaborada pelo projeto ISLE, que é a continuação, em nível internacional, do projeto EAGLES.

É a partir das avaliações que os usuários podem determinar o sistema mais adequado para suas necessidades, e usuários e fabricantes podem identificar os problemas dos programas atuais e buscar soluções para eles, melhorando, assim, a qualidade final das traduções geradas de acordo com as necessidades.

Como observam Sarmento *et al.* (no prelo), os critérios para avaliação de tradução automática ou humana podem variar muito, devido à complexidade do processo tradutório e aos objetivos da tradução e da avaliação, abrangendo tópicos tão diversos como a correção lingüística formal e a função do material traduzido na cultura de chegada. Conscientes dessa diversidade, os pesquisadores responsáveis pelo FEMTI (*Framework for the Evaluation of Machine Translation in ISLE*) procuraram reunir os vários métodos e critérios de avaliação existentes para TA e propor uma taxonomia, aplicando também as normas ISO/IEC 9126 e 14598 de avaliação de software aos sistemas de TA. Essa taxonomia será comentada na próxima seção e será aplicada no estudo de caso do capítulo 4.

Historicamente, as avaliações não têm considerado o contexto de uso, ficando restritas aos aspectos lingüísticos com base em dois conceitos principais: fluência e fidelidade. Esses conceitos, discutidos e questionados no âmbito dos estudos da tradução, são apresentados de forma bastante tradicional na área de avaliação de TA: a fluência costuma ser associada à capacidade do sistema de gerar frases bem-formadas lexical e sintaticamente; já a fidelidade é determinada em termos semânticos, em relação à preservação do significado original.

Quando a avaliação inclui critérios extralingüísticos, os mais frequentes são preço do sistema, possibilidade de expansão (facilidade de inclusão de novas palavras e regras gramaticais pelo usuário) e abrangência (capacidade de especialização do sistema aos domínios de interesse) (Hovy *et al.*, 2002).

Em relação aos objetivos da avaliação, a literatura identifica três categorias principais (Oliveira *et al.*, 2000; Maia&Barreiro, no prelo): avaliação de diagnóstico, de desempenho e de adequação. A avaliação de diagnóstico procura identificar as limitações, os erros e as deficiências dos sistemas; a avaliação de desempenho analisa seus estágios de desenvolvimento; e a avaliação de adequação procura determinar se o sistema atende a determinados contextos operacionais. As avaliações de diagnóstico e de desempenho costumam ser realizadas por pesquisadores e fabricantes, enquanto a avaliação de adequação geralmente é feita por usuários.

Em termos de metodologia, a avaliação pode ser interna (*blackbox*, ou caixa preta), com acesso ao funcionamento interno do programa, ou externa (*glassbox*, ou caixa de vidro), com acesso somente ao texto original e ao texto traduzido. A

avaliação externa é a mais comum, dado que o acesso aos componentes dos sistemas é restrito aos fabricantes. Em ambos os casos, a avaliação pode ser automática, baseada em algoritmos, ou humana, realizada por usuários leigos (que não sejam da área de linguagem nem tradutores) ou por lingüistas, tradutores ou fabricantes de software.

A avaliação externa, como não tem acesso aos componentes do programa, normalmente se concentra na qualidade da tradução. Segundo Sarmiento *et al* (no prelo) essa qualidade é geralmente determinada de acordo com os seguintes critérios: em função do objetivo, em comparação com traduções humanas, por correção formal e em função do esforço de pós-edição. Em relação à determinação da qualidade em função do objetivo, o exemplo apresentado é o de um usuário que recorre à TA para encontrar a definição de um termo num *website*, num idioma que ele não domina. Caso o usuário encontre a definição e a compreenda, a TA poderá ser considerada de qualidade. Os próprios autores observam que essa noção de qualidade é bastante limitada, não apenas porque envolve a capacidade do usuário de lidar com o sistema, mas também porque mesmo um sistema que utilizasse o método de tradução palavra por palavra poderia atender a esse objetivo específico, sem, no entanto, poder ser considerado um programa de qualidade em termos mais abrangentes. Além disso, esse critério dificultaria a comparação dos resultados de vários programas simultaneamente.

Já sobre o critério de comparação da tradução gerada automaticamente com traduções humanas, os autores comentam duas principais dificuldades para sua implementação prática: a primeira seria como determinar a distância aceitável entre os textos traduzidos, especialmente considerando-se as várias possibilidades de tradução para uma mesma frase ou segmento de texto. A segunda dificuldade é de ordem financeira, já que seria necessário construir um *corpus* de referência com várias traduções humanas dos textos a serem utilizados nos testes para comparação, o que implicaria custos muito elevados para que o *corpus* fosse realmente abrangente. Alguns métodos automáticos de avaliação utilizam métricas para a medição da distância entre as frases, como o padrão BLEU, descrito em Papineni *et al* (2001, *apud* Sarmiento *et al*, no prelo).

A qualidade em relação à correção formal é avaliada em termos de morfologia, léxico e sintaxe. Uma das principais falhas está no fato de o aspecto

semântico geralmente ser desconsiderado, apesar de ser um dos principais problemas para a TA. Esse tipo de avaliação normalmente toma como base uma categorização de erros pré-definida. Por um lado, esse método representa vantagens em termos de sistematização, possibilidade de comparação entre vários sistemas e análises específicas por tipo de erro, mas, por outro lado, depende da complexa tarefa de elaboração de uma categorização rigorosa, voltada para erros exclusivamente de TA. Essa categorização, por sua vez, depende da coleta e da análise de grandes quantidades de dados e é dificultada pela natureza não estanque dos erros, que podem ter causas diferentes e podem ser observados de vários pontos de vista, como veremos nas tipologias propostas para o português na seção 3.3. Outras dificuldades são o fato de a categorização precisar lidar com duas línguas diferentes (alvo e fonte) e a determinação, de caráter subjetivo, da gravidade de cada erro. Por fim, os autores observam que, para a viabilidade da avaliação em termos de correção formal, a categorização deveria ser elaborada por pesquisadores com formação em lingüística, tradução humana e tradução automática, não havendo ainda muitos profissionais com esse perfil mais diversificado.

O último critério de determinação de qualidade apresentado em Sarmiento *et al* (no prelo) é em função da pós-edição. De acordo com a definição de Allen (2003), pós-edição é o “termo usado para a correção do produto da tradução automática por revisores ou lingüistas humanos” (nossa tradução)². Dessa forma, quanto menos correções forem necessárias na pós-edição, melhor será a qualidade do sistema. Apesar de ter implicações teóricas para a avaliação de mais de um sistema, como problemas de julgamentos subjetivos de cada revisor em relação ao que deve ser corrigido em cada texto, no mercado de localização esse pode ser um critério muito eficiente. Os padrões de qualidade desejados para o texto final já serão conhecidos, pois provavelmente serão os mesmos já empregados para a tradução humana. A qualidade da TA, então, poderá ser medida em termos de tempo e custo. Se, para atingir a mesma qualidade final, o tempo dedicado ao processo de TA e à pós-edição for inferior ao tempo gasto com o processo de tradução e revisão humanas, então o sistema de TA poderá ser uma boa opção.

² “[Post-editing is the] term used for the correction of machine translation output by human linguists/editors.”

Ainda que o tempo seja igual ou um pouco superior, o uso do tradutor automático pode se justificar em termos de redução de custo ou retorno do investimento no sistema, já que não haverá o custo da etapa de tradução.

Voltando aos métodos de avaliação, Hovy *et al* (2002) descrevem a tradução reversa (*back translation*) como um dos mais antigos. Nesse método, uma frase é traduzida para outra língua e depois traduzida novamente para a língua-fonte, para uma comparação com a frase original. Segundo os autores, esse é um método muito pouco sistemático e que dificulta a comparação entre sistemas. Uma implementação desse método para o português é o *Boomerang*, da Linguateca, descrito em Sarmento (no prelo). Sarmento argumenta que o método pode ser útil para identificar os tipos de estruturas que representam mais dificuldade para o sistema e aqueles com as quais o sistema já lida melhor, por exemplo. Apesar de o projeto ter sido interrompido, a ferramenta continua disponível no endereço eletrônico <http://poloclup.linguateca.pt/ferramentas/boomerang/> (último acesso em 20 de março de 2006).

Mais adiante neste capítulo, discutiremos três estudos elaborados especificamente para o português. Antes, porém, vale notar que eles se concentram nas questões lingüísticas e, como argumentam os pesquisadores do FEMTI, a consideração do contexto de uso é fundamental para a avaliação. Portanto, para fazer uma avaliação mais abrangente da utilização da TA no contexto do mercado de localização, adotamos a classificação proposta no FEMTI.

3.2. A proposta de avaliação do FEMTI

O objetivo do projeto FEMTI é sistematizar os vários métodos de avaliação de TA existentes considerando também as normas ISO/IEC para avaliação de qualidade de software. Como vimos, seu principal diferencial é considerar o contexto de uso do sistema, o que é especialmente relevante no nosso estudo de caso.

Com base nos vários métodos e nas normas ISO/IEC, foi elaborada uma taxonomia bastante extensa, que refina muitos dos aspectos gerais de avaliação

apresentados na seção anterior. Em relação aos objetivos da avaliação, por exemplo, a taxonomia apresenta sete diferentes possibilidades: Avaliação de viabilidade, Explicitação das necessidades, Avaliação interna, Avaliação diagnóstica, Avaliação declarativa, Avaliação operacional e Avaliação de utilidade.

A taxonomia está dividida em duas partes. A primeira, Requisitos da avaliação, tem como objetivo auxiliar avaliadores, pesquisadores e fabricantes a definir o contexto de uso. De acordo com essa definição, são indicadas as categorias a serem avaliadas na segunda parte, intitulada Características do sistema a serem avaliadas. Por exemplo, se na parte 1 ficar definido que o objetivo da avaliação é verificar a viabilidade do sistema, as duas categorias a serem avaliadas na parte 2 serão Consideração dos aspectos específicos ao *corpus* e Precisão. Dentro de cada categoria são apresentadas várias métricas usadas na comunidade científica para avaliação daquele determinado aspecto. O projeto é descrito de forma mais detalhada em Hovy *et al* (2002).

O FEMTI representa um grande avanço em termos de sistematização, com um alto nível de detalhamento. Outra grande contribuição do projeto é mostrar que o conceito de qualidade depende diretamente do contexto de utilização. Como observam os autores, assim como não faz sentido perguntar *Qual é a melhor casa?*, não faz sentido perguntar *Qual é o melhor tradutor automático?*. “Até mesmo um sistema de baixa qualidade pode ser útil, ou mesmo ideal, nas devidas circunstâncias” (Hovy *et al*, 2002:43, nossa tradução³). A relação entre contexto de uso e qualidade fica clara na definição das normas ISO/IEC: qualidade é “a totalidade dos recursos e características de um produto ou serviço que estão relacionados à sua capacidade de satisfazer as necessidades explícitas e implícitas” (ISO/IEC, 1991:2, *apud* Hovy *et al*, 2002, nossa tradução⁴).

Para avaliar o contexto, a proposta inclui categorias voltadas para as características da tarefa tradutória, do público-alvo (consumidor da tradução), do usuário da TA e do tipo de texto e de autor. Em relação ao software, são

³ “...even poor MT can be useful, or even ideal, in the right circumstances.”

⁴ “the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs.”

consideradas as características específicas e externas do sistema, como os recursos disponíveis, a funcionalidade e os custos envolvidos.

Por questões de espaço e praticidade, apresentaremos apenas a tradução das categorias, sem as definições. As definições detalhadas, bem como as métricas sugeridas para avaliação, os agentes envolvidos em cada etapa e as referências bibliográficas para cada categoria podem ser encontradas on-line em inglês no endereço <http://www.issco.unige.ch/projects/isle/femti/>. Como comentado no *website*, essa é uma taxonomia “em construção”, ainda incompleta e em constante atualização com base nos comentários dos usuários.

CLASSIFICAÇÃO DO FEMTI

1 REQUISITOS DA AVALIAÇÃO

1.1 Propósito da avaliação

1.1.1 Avaliação de viabilidade

1.1.2 Explicitação das necessidades

1.1.3 Avaliação interna

1.1.4 Avaliação diagnóstica

1.1.5 Avaliação declarativa

1.1.6 Avaliação operacional

1.1.7 Avaliação de utilidade

1.2 Objeto da avaliação

1.2.1 Um componente do sistema de TA

1.2.2 O sistema de TA em si

1.2.3 O sistema de TA como parte de um sistema mais amplo

1.3 Características da tarefa tradutória

1.3.1 Assimilação

1.3.1.1 Repasse/distribuição de documentos

1.3.1.2 Sumarização/extração de informações

1.3.1.3 Pesquisa

1.3.2 Divulgação

1.3.2.1 Publicação interna

1.3.2.1.1 Rotineira**1.3.2.1.2 Experimental/pesquisa****1.3.2.2 Publicação externa****1.3.2.2.1 Um tipo de cliente****1.3.2.2.2 Vários tipos de clientes****1.3.3 Comunicação****1.3.3.1 Síncrona****1.3.3.2 Assíncrona****1.4 Características dos usuários****1.4.1 Usuário da TA****1.4.1.1 Formação****1.4.1.2 Proficiência na língua-fonte****1.4.1.3 Proficiência na língua-alvo****1.4.1.4 Familiaridade com o computador****1.4.2 Consumidor da tradução****1.4.2.1 Proficiência na língua-fonte****1.4.2.2 Proficiência na língua-alvo****1.4.3 Usuário corporativo****1.4.3.1 Volume de tradução****1.4.3.2 Tamanho da equipe****1.4.3.3 Tempo disponível para a tradução****1.5 Características da entrada (autor e texto)****1.5.1 Tipo de documento****1.5.1.1 Gênero****1.5.1.2 Domínio/campo de aplicação****1.5.2 Características do autor****1.5.2.1 Proficiência na língua-fonte****1.5.2.2 Treinamento profissional****1.5.3 Características relacionadas às fontes de erros****1.5.3.1 Fontes de erros intencionais****1.5.3.2 Fontes de erros relacionadas à mídia****1.5.3.3 Erros relacionados ao desempenho****2. CARACTERÍSTICAS DO SISTEMA A SEREM AVALIADAS**

2.1 Características específicas do sistema de TA

2.1.1 Modelos de processo de tradução

2.1.1.1 Metodologia

2.1.1.1.1 Modelos baseados em regras

2.1.1.1.2 Modelos baseados em estatística

2.1.1.1.3 Modelos baseados em exemplos

2.1.1.1.4 Modelos com memória de tradução

2.1.1.2 Modelos

2.1.1.2.1 Direto

2.1.1.2.2 Transferência

2.1.1.2.3 Interlíngua

2.1.2 Recursos lingüísticos

2.1.2.1 Idiomas

2.1.2.2 Dicionários

2.1.2.3 Listas de palavras, glossários

2.1.2.4 *Corpora* comparáveis e paralelos

2.1.2.5 Gramáticas

2.1.3 Características do processo

2.1.3.1 Atividades de preparação da tradução

2.1.3.2 Atividades posteriores à tradução

2.1.3.3 Atividades interativas

2.1.3.4 Atualização do dicionário

2.1.3.5 Gerenciamento do processo

2.2 Características externas do sistema

2.2.1 Funcionalidade

2.2.1.1 Adequação

2.2.1.1.1 Somente língua-alvo

2.2.1.1.1.1 Legibilidade (ou fluência, compreensibilidade, clareza)

2.2.1.1.1.2 Inteligibilidade

2.2.1.1.1.3 Coerência

2.2.1.1.1.4 Coesão

2.2.1.1.2 Contrastivo/interlingual

**2.2.1.1.2.1 Consideração dos fenômenos
específicos ao *corpus***

2.2.1.1.2.2 Estilo

2.2.1.2 Precisão

2.2.1.2.1 Fidelidade

2.2.1.2.2 Uniformidade

2.2.1.2.3 Terminologia

2.2.1.3 Boa formação

2.2.1.3.1 Pontuação

2.2.1.3.2 Léxico/escolha lexical

2.2.1.3.3 Gramática/sintaxe

2.2.1.3.4 Morfologia

2.2.1.4 Interoperabilidade

2.2.1.5 Conformidade

2.2.1.6 Segurança

2.2.2 Confiabilidade

2.2.2.1 Maturidade

2.2.2.2 Tolerância a falhas

2.2.2.3 Frequência de travamento

2.2.2.4 Facilidade de recuperação

2.2.2.5 Conformidade em relação à confiabilidade

2.2.3 Facilidade de uso

2.2.3.1 Facilidade de compreensão

2.2.3.2 Aprendizagem

2.2.3.3 Operabilidade

2.2.3.4 Documentação

2.2.3.5 Atratividade

2.2.3.6 Conformidade em relação à facilidade de uso

2.2.4 Eficiência

2.2.4.1 Rapidez

2.2.4.1.1 Tempo de pré-processamento

2.2.4.1.1.1 Tempo de pré-edição

2.2.4.1.1.2 Conversão de códigos

2.2.4.1.1.3 Tempo de preparação**2.2.4.1.2 Velocidade de tradução (entrada e saída)****2.2.4.1.3 Tempo de pós-processamento****2.2.4.1.3.1 Tempo de pós-edição****2.2.4.1.3.2 Conversão de códigos****2.2.4.1.3.3 Tempo de atualização****2.2.4.2 Utilização de recursos****2.2.4.2.1 Memória****2.2.4.2.2 Léxico****2.2.4.2.3 Limpeza dos arquivos****2.2.4.2.4 Tamanho do programa****2.2.5 Facilidade de manutenção****2.2.5.1 Analisibilidade****2.2.5.2 Flexibilidade****2.2.5.2.1 Facilidade de atualização de aspectos multilíngües do sistema****2.2.5.2.2 Capacidade de aprimoramento****2.2.5.2.3 Facilidade de atualização do dicionário****2.2.5.2.4 Facilidade de alteração das regras gramaticais****2.2.5.3 Estabilidade****2.2.5.4 Testabilidade****2.2.5.5 Conformidade em relação à facilidade de manutenção****2.2.6 Portabilidade****2.2.6.1 Adaptabilidade****2.2.6.2 Facilidade de instalação****2.2.6.3 Conformidade em relação à portabilidade****2.2.6.4 Facilidade de substituição****2.2.6.5 Coexistência****2.2.7 Custo****2.2.7.1 Custo inicial**

2.2.7.2 Custo de manutenção

2.2.7.3 Outros custos

3.3.

Três estudos sobre avaliação para o português

Apesar de incluir categorias para a avaliação lingüística, foge ao escopo da taxonomia do FEMTI apresentar tipologias ou critérios de avaliação para uma língua em particular. Para complementar a taxonomia com dados específicos do português, recorreremos a três fontes: a dissertação de mestrado de Carmen Dayrell, intitulada *Sistemas de Tradução Automática: Avaliação e Propostas de Melhoria* (Dayrell, 1999), um estudo do NILC (Núcleo Interinstitucional de Lingüística Computacional) relatado no artigo “A critical analysis of the performance of English-Portuguese-English MT systems” (Oliveira *et al*, 2000) e o projeto TrAva, desenvolvido pela Linguateca, em Portugal, e descrito nos artigos “Ferramentas para experimentação, recolha e avaliação de exemplos de tradução automática” (Sarmiento, no prelo) e em “Uma experiência de recolha de exemplos classificados de tradução automática de inglês para português” (Maia&Barreiro, no prelo). Daqui em diante, os trabalhos serão referidos como “Dayrell”, “Oliveira *et al*”, ” e “TrAva”.

Como os problemas de qualidade da tradução automática ainda são muito variados e complexos, consideramos especialmente relevante apresentar as discussões dessas avaliações para o português. Como elas indicam problemas lingüísticos que também ocorrem com o tradutor automático que avaliaremos no capítulo 4, seus resultados podem ser aplicados para a melhoria de vários aspectos do programa, sendo complementares aos nossos resultados mais especificamente voltados para o mercado de localização.

Oliveira et al e o *TrAva* fazem uma avaliação externa (ou seja, sem acesso ao funcionamento do programa, levando em consideração somente o texto original e a tradução gerada pelo programa). Já *Dayrell* faz uma avaliação híbrida, com acesso a parâmetros do programa para implementar e testar soluções para alguns problemas. Os três estudos recorrem a vários tradutores automáticos.

Em todos os casos, o enfoque é estritamente lingüístico (principalmente morfológico, lexical e sintático e, no caso de *Oliveira et al*, também semântico-

pragmático), não envolvendo questões contextuais como aquelas discutidas sobre o FEMTI na seção anterior.

Devido à variedade e à quantidade de erros, todas as análises se concentram no nível da frase ou da oração, sem considerar o nível discursivo ou fatores extralingüísticos. A esse respeito, Sarmiento (no prelo) argumenta ainda que “o limite de cinco palavras pode ser considerado aceitável, tendo em conta que a tecnologia actual de TA que consultámos parece operar apenas sobre estruturas de duas ou três palavras, raramente analisando a frase como um todo”. Vejamos, a seguir, os principais aspectos de cada um dos estudos.

3.3.1.

Aspectos gerais dos estudos

Programas avaliados

Os tradutores automáticos avaliados foram:

- *Dayrell - Globalink, TraduzTudo, Transcend, A.R.T. e Systran* (acesso disponível na Web. No caso dos sistemas comerciais, foram utilizadas as versões de demonstração gratuitas oferecidas na Web.);
- *Oliveira et al – TranslatorPro e Tradunet* (vendidos comercialmente na época da análise) e *Alta Vista, Intertran, GO Translator e Enterprise Translator Server* (disponíveis na Web);
- *TrAva - FreeTranslation, Systran, E-T Server e Amikai* (acesso disponível na Web. No caso dos sistemas comerciais, foram utilizadas as demonstrações gratuitas oferecidas na Web.).

Apesar das diferenças de nomenclatura, observamos que alguns dos sistemas parecem se repetir, como o *E-T Server*, que supomos ser o mesmo que *Enterprise Translator Server*, e o *TranslatorPro*, possivelmente o mesmo que *Globalink TranslatorPro*. O *Alta Vista*, disponível na Web, é baseado no *Systran*. Aparentemente, todos os sistemas são baseados no método de tradução indireta por transferência, à exceção do *Intertran*, que parece adotar o método de tradução direta, palavra por palavra.

Direção

Vários desses tradutores automáticos oferecem a possibilidade de tradução entre vários pares de idiomas, mas os três estudos utilizam somente o inglês e o português.

Dayrell e *TrAva* analisam somente traduções na direção inglês-português, enquanto *Oliveira et al* avaliam também a direção português-inglês.

Objetivos

Os três estudos partem de um objetivo comum, o levantamento de problemas de TA envolvendo português, mas assumem enfoques específicos bastante variados.

Enquanto *Dayrell* tem o objetivo pragmático de promover a melhoria efetiva dos programas, inclusive propondo soluções para os erros encontrados, *Oliveira et al* e o *TrAva* demonstram ter uma preocupação mais teórico-conceitual, com a finalidade de tomar os dados como base para uma maior compreensão do funcionamento e dos desafios da TA.

De acordo com os autores, os objetivos são:

1. *Dayrell*

“O objetivo deste trabalho é avaliar o desempenho de cinco sistemas de tradução automática, na tradução do inglês para o português, e propor algumas soluções possíveis de ser implementadas, visando à melhoria dos sistemas”. (Dayrell, 1999:1)

2. *Oliveira et al*

“[...] primeiramente analisamos o desempenho dos sistemas na tradução de trechos de jornais de inglês para português e vice-versa. Além de identificar as principais limitações, discutimos por que tais limitações existem e as classificamos segundo o processamento automático de informações lexicais, sintáticas e semântico-pragmáticas” (Oliveira *et al* 2000, nossa tradução⁵).

⁵ “...we first analyze several systems’ performances in translating newspaper excerpts from English into Portuguese and vice-versa. Besides identifying their main limitations, we discuss why such limitations exist and classify them according to automatic processing of lexical, syntactic, and semantic-pragmatic information”.

3. *TrAva*

“Durante este período de experimentação com o METRA, acabou por parecer oportuno desenvolver algumas funcionalidades que permitissem não só observar os fenómenos associados à TA, mas também *compilá-los, organizá-los e categorizá-los sistematicamente* [...] [a fim de] *proporcionar uma boa base para estudos mais aprofundados sobre TA e sua avaliação*. [...] ajudaria também a estimar, ainda que empiricamente, *o tipo de erros mais frequentes e quais os casos que os originam*” (Sarmiento, no prelo, nosso grifo).

Abrangência

Enquanto *Dayrell* e *Oliveira et al* fazem uma análise pontual, partindo de traduções concluídas de textos relativamente pequenos para identificar algumas limitações dos sistemas, o projeto *TrAva* é uma iniciativa contínua disponível na Web, com enfoque principalmente na coleta de grandes volumes de dados sobre erros de TA para posterior análise.

Assim, *Dayrell* e *Oliveira et al* têm menor abrangência porque trabalham com dados mais limitados, avaliados por uma equipe reduzida, mas, por outro lado, já apresentam os resultados de suas análises. O *TrAva*, por sua vez, pretende recolher uma grande quantidade de dados, introduzidos e classificados por todos os avaliadores que desejarem contribuir com o projeto. As frases já classificadas estão reunidas no *corpus* CorTA, disponível na Web. Uma análise de dados do *TrAva* com fins pedagógicos é apresentada em Maia&Barreiro (no prelo).

Tipos de textos e tamanho dos *corpora*

Em termos de extensão do material, O *TrAva* reúne apenas frases com no máximo cinco palavras, para simplificar a classificação dos erros pelos usuários. Já *Oliveira et al* compilaram um *corpus* com 20 trechos de jornal. No caso de *Dayrell*, foram utilizados dois textos completos, de aproximadamente uma página cada um.

Vale observar que a extensão e a unidade textuais não são consideradas pelos programas analisados. Como afirma Sarmiento (no prelo), “...foi possível verificar que, mesmo ao nível da frase, os serviços de TA desprezavam a informação de contexto, executando traduções aparentemente ao nível de conjuntos de uma ou duas palavras apenas”. Pelos resultados, podemos considerar

que essa seja uma característica dos sistemas utilizados nos outros estudos também.

Em termos de linguagem, os três estudos têm em comum o fato de empregarem textos considerados de linguagem geral, à exceção de um dos textos de *Dayrell*, como veremos adiante. *Oliveira et al* e o *TrAva* afirmam que os sistemas estão voltados para a linguagem mais geral, sobretudo aqueles disponíveis on-line, que podem ser utilizados para fins muito variados e não oferecem recursos para linguagem especializada, como dicionários técnicos, por exemplo. Como afirmam Maia&Barreiro (no prelo): “...pensamos que começar com a linguagem geral é a escolha mais apropriada, [...] porque os programas de TA *online* estão preparados para fins relativamente genéricos”.

Já no caso de *Dayrell*, um dos objetivos secundários era confirmar a hipótese de que os sistemas apresentam um desempenho melhor com linguagem especializada, num domínio mais restrito. Por isso, a autora optou por usar dois textos, com assuntos, registros e estilos diferentes, sendo um de linguagem mais geral e o outro de linguagem mais especializada. O primeiro texto de *Dayrell* foi retirado de uma revista feminina, tendo como tema o relacionamento humano. Já o segundo texto apresenta linguagem um pouco mais técnica, da área de lingüística, sobre tradução automática.

Segunda a autora, o primeiro texto é mais informal e emprega uma linguagem mais geral, enquanto o segundo utiliza uma linguagem mais restrita, com terminologia específica de lingüística e com menos variedade de estruturas sintáticas e itens lexicais.

Dos 20 trechos de jornal selecionados por *Oliveira et al*, dez estavam em português, retirados do jornal *Folha de São Paulo* e dez estavam em inglês, extraídos do *The New York Times*. As seções escolhidas foram editoriais, noticiário nacional e internacional e cartas de leitores.

No caso do *TrAva*, não há controle das frases introduzidas pelos usuários. A única recomendação é que sejam usadas frases originalmente escritas em inglês, sendo sugerida a utilização do BNC (*British National Corpus*).

Os parâmetros de avaliação

Os três estudos realizam avaliações humanas. No caso do *TrAva*, os avaliadores são os vários usuários do sistema. Em *Oliveira et al*, aparentemente os próprios pesquisadores fizeram as avaliações nos níveis lexical e sintático. No nível semântico-pragmático, como o enfoque era a inteligibilidade dos trechos traduzidos automaticamente, sem pós-edição nem consulta aos textos originais, os pesquisadores recorreram a outros avaliadores humanos não envolvidos na pesquisa. O artigo não apresenta mais informações sobre os avaliadores. Já a avaliação de *Dayrell* aparentemente foi feita pela própria equipe envolvida na pesquisa.

Crítérios de classificação

Assim como os desafios em TA, os critérios para abordá-los são muitos e muito variados entre si. Provavelmente por isso, as pesquisas assumem um caráter bastante empírico, em que os dados praticamente determinam as questões lingüísticas a serem pesquisadas e muitas vezes o próprio embasamento teórico mais adequado para analisá-los.

Em *Dayrell*, no *TrAva* e na nossa própria análise, que será apresentada no capítulo 4, as categorias foram determinadas a partir dos erros observados, procurando agrupá-los. Já em *Oliveira et al*, parece ter havido uma pré-seleção de alguns aspectos gerais a serem observados, mas as categorias de erros sintáticos foram criadas com base nos erros identificados na própria avaliação. Parece haver um consenso quanto às macrocategorias, estando todas as análises divididas em questões lexicais e morfológicas e questões sintáticas. *Oliveira et al* acrescentam ainda o aspecto semântico-pragmático.

Já em relação às subcategorias, a divisão não é tão consensual. As variações entre os erros analisados e a dificuldade de determinar precisamente uma única categoria para cada um tornam a tarefa bastante complicada. Há também a dificuldade de determinar se as categorias serão baseadas nos erros da frase traduzida, nas características da língua-fonte ou ainda no próprio processo tradutório.

Dayrell parte dos dados para identificar 15 categorias de erros, que serão descritas na próxima seção. Como na época a autora não teve acesso a categorias de erros especificamente voltados para a tradução automática, ela tomou a descrição de erros de tradução humana de Mona Baker (1992, *apud* Dayrell, 1999) como base teórica para analisar os erros.

Já para o *TrAva*, o fato de os erros de TA muitas vezes não estarem relacionados a erros humanos fez com que as categorias fossem definidas a partir da observação apenas de erros de TA. A principal diferença em relação aos outros projetos é que a quantidade de dados observados para a definição das categorias foi muito mais expressiva, o que possibilitou a criação de uma categorização muito mais abrangente e detalhada. Em linhas gerais, ela adota as classes estabelecidas pela gramática tradicional, acrescentando algumas mais específicas do processo tradutório, como “resolução incorrecta do POS-homógrafo” para dar conta do problema da homografia, por exemplo.

Diferentemente dos outros estudos, as análises dos dados do *TrAva* apresentadas em Maia&Barreiro (no prelo) não se iniciaram a partir dos textos, devido à própria natureza do sistema. Nesse caso, alunos do curso de mestrado em tradução escolheram arbitrariamente uma das categorias preestabelecidas no sistema para então analisar as frases que haviam sido classificadas sob aquela categoria.

No caso de *Oliveira et al*, aparentemente houve uma seleção prévia de algumas questões já sabidamente problemáticas para TA. Para a análise lexical, as questões foram: dicionarização, homonímia, conotação e expressões. Para a análise sintática, o enfoque foi a capacidade do programa de recuperar as relações de dependência da frase original e preservá-las para possibilitar a interpretação da tradução gerada. Com esse enfoque, foram identificadas 11 categorias de problemas sintáticos, que serão apresentadas na próxima seção.

Como comentamos, além das macrocategorias “léxico e morfologia” e “sintaxe”, o estudo de *Oliveira et al* inclui um terceiro nível de análise, com um enfoque semântico-pragmático. Nessa categoria, são consideradas duas perspectivas: co-referencialidade e interpretabilidade das frases traduzidas.

3.3.2. Categorias de erros

Tomando como base os critérios apresentados na seção 3.3.1, os estudos estabeleceram as categorias de erros apresentadas a seguir.

3.3.2.1. As categorias de *Dayrell*

Dayrell identificou 15 tipos de erros e agrupou-os nas categorias “Erros primários” e “Erros devidos a problemas tradutológicos”.

Foram considerados erros primários aqueles não relacionados diretamente às dificuldades do processo tradutório, estando ligados à lexicografia ou a erros de programação do tradutor automático “causados pela implementação incorreta, ou não implementação, de regras gramaticais, inclusão incorreta do dicionário, ou pela má elaboração do programa” (Dayrell, 1999:79). Nessa categoria foram incluídas as questões que se seguem. Para todas as categorias, selecionamos aleatoriamente exemplos dentre as traduções dos vários programas utilizados. Incluímos as traduções sugeridas pela autora apenas quando consideramos que o erro não era óbvio e dependia de contexto:

1. dicionário incorreto – Exemplo: “would” correspondendo somente a “poder” e “rather” correspondendo somente a “melhor”
2. item mantido na língua-fonte (por não constar no dicionário) – Exemplo: “jealous”
3. formação lexical e ortografia – Exemplo: “confiançamente” e “exijiu”
4. formatação e acentuação – Exemplo: “nao”

Já a categoria “Erros devidos a problemas tradutológicos” inclui dificuldades do processo tradutório. As questões identificadas foram analisadas de acordo com a classificação de erros de tradução humana de Mona Baker (1992, *apud* Dayrell, 1999).

A primeira categoria de Mona Baker selecionada foi a “não equivalência lexical”. Nela foram incluídos os seguintes tipos de erros:

5. ambigüidade lexical – Exemplo: “split”, traduzido como “divisão” enquanto a tradução sugerida foi “separação”
6. expressões – Exemplo: “how come” traduzido como “como vem”
7. inadequação estilística – Exemplo: “jealous” traduzido como “cioso”
8. inclusão de itens (itens que ficaram faltando na tradução) – Exemplo:
Original: “let’s call him Dave”
TA: “chamemos o Dave”
9. exclusão de itens (itens que deveriam ter sido excluídos da tradução) - Exemplo:
Original: “It was your decision to end it”
TA: “Era sua decisão [para] terminar”

Na segunda categoria de Baker, “erros lexicais causados por erros gramaticais”, foi incluído o seguinte tipo de erro:

10. classe gramatical
Exemplo:
Original: “why not try”
TA: “Porque não tentativa”

A terceira categoria de Baker era “não equivalência gramatical”. Nesse caso, *Dayrell* organizou os cinco tipos de erros restantes dentro das subcategorias de Baker, a saber: gênero e número, pessoa, sistema verbal, voz, ordem das palavras e estruturas incompatíveis. A autora procurou fazer uma correspondência entre os erros identificados e essas subcategorias de Baker, da seguinte forma:

11. concordância (que exemplifica a subcategoria “gênero e número”)
Exemplo:
Original: “she was so angry”
TA: “foi ela assim zangado”
12. pronome incorreto (subcategoria “pessoa”)

Exemplo:

Original: “...made any sense to me”

TA: “qualquer sentido para **me**”

13. sistema verbal

Exemplo (modal *would* seguido de verbo sempre traduzido como futuro do pretérito simples):

Original: “Some would claim...”

TA: “Alguns reivindicariam...”

Tradução sugerida: “alegavam”

14. ordem das palavras

Exemplo (distribuição de adjetivo):

Original: “the new man in”

TA: “O homem novo em”

15. estruturas incompatíveis

Exemplo:

Original: “the early 1980s”

TA: “os cedo 1980s”

Na discussão da subcategoria “voz”, *Dayrell* inclui exemplos discutidos na categoria “sistema verbal”. Todas as categorias são detalhadamente discutidas pela autora, sendo apresentados e comentados todos os exemplos diferentes de cada categoria. O estudo propõe ainda soluções para cada problema especificamente.

3.3.2.2.

As categorias de *Oliveira et al*

As categorias de *Oliveira et al* foram organizadas nos níveis lexical, sintático e semântico-pragmático.

No nível lexical, as categorias selecionadas para análise foram dicionarização, homonímia, conotação e expressões.

1. Dicionarização - os principais problemas encontrados foram:

a) reconhecimento de nomes próprios e nomes deles derivados -

Exemplo de termo mantido em inglês: “Hungarian”. Exemplo

de termo traduzido indevidamente: “World Trade Center”, traduzido como “Centro de Comércio Mundial”

- b) relações do dicionário – Exemplo: “Páscoa” traduzido como “Passover” em vez de “Easter”
 - c) flexão – inclusão de formas flexionadas como uma entrada separada do dicionário. Exemplo: “tradicionais”.
2. No caso da homonímia, são discutidas as alterações de classe gramatical e as variações dentro de uma mesma classe.

- a. alteração de classe gramatical

Exemplo:

Original: “São Paulo”

TA: “are Paulo”

- b. Variação de significado na mesma classe gramatical (palavra *ever*)

Exemplo:

Original: “Hungary has ceded more sovereignty than many other nations – including the United States – would *ever* consider”

TA: “sempre”

3. Conotação - relacionada ao uso específico de palavras ou expressões em um contexto cultural.

Exemplo (expressão “pegar carona”):

Original: “Os comerciantes *pegaram carona* na celebração de Páscoa e rechearam suas vitrines com topo tipo de opções de presentes”

TA: “hitchhike”

4. Expressões - significado do todo não corresponde à soma dos significados das palavras que as compõem.

O artigo não apresenta exemplos das traduções automáticas das expressões, mas cita os seguintes exemplos do original: “as far as”, “provided that”, “da

mesma forma” e “abrir mão de”. Estão incluídos nesta categoria também os *phrasal verbs*.

Exemplo de *phrasal verb*:

Original: “pulled out”

TA: “puxado sem”

Já no nível sintático, o foco era a capacidade dos sistemas de recuperar as relações de dependência do original para possibilitar a interpretação da oração traduzida. A esse respeito, são identificados dois problemas: a recuperação das relações sintáticas entre os itens lexicais e a seleção da relação mais apropriada dentre aquelas possíveis. Com esse enfoque, foram observados os seguintes tipos de problemas, seguidos de exemplos (que eram pertinentes no contexto):

1. concordância entre artigo e substantivo: “um retirada”
2. concordância entre substantivo e verbo: “outras nações ia considerar”
3. emprego de tempo verbal: “had been condemned” em vez de “were condemned”
4. emprego de preposição: “ao fim de semana”
5. emprego de artigo: “the fraternity”
6. emprego de pronome: “its” em vez de “his”
7. grau comparativo: “more early”
8. falta de preposição: “direito levar armas”
9. falta de artigo: “candidate” em vez de “the candidate”
10. falta de pronome reflexivo: “he controlled” em vez de “he controlled himself”
11. falta de conjunção: “dizem Senhor Mogilevich contratou” em vez de ““dizem *que* Senhor Mogilevich contratou”

Finalmente, na análise semântico-pragmática, o foco de interesse era a interpretabilidade das frases traduzidas automaticamente, ou seja, se os leitores conseguiam compreender as traduções geradas pelos sistemas. Os leitores consideraram as passagens inteligíveis em apenas 30% dos casos para os sistemas

Alta Vista e *Enterprise Translator Server*, 20% para o *TranslatorPro* e o *Tradunet* e 10% para o *Intertran*.

3.3.2.3.

As categorias do *TrAva*

As categorias elaboradas para o sistema *TrAva* estão reunidas nas tabelas a seguir, retiradas de Sarmento (no prelo). A tabela 1 apresenta os aspectos morfológicos e lexicais. Na tabela 2 são listados os aspectos sintáticos decorrentes do processo tradutório. Não são apresentados exemplos nas tabelas.

No estudo sobre a aplicação pedagógica do *TrAva*, Maia&Barreiro (no prelo) analisam alguns exemplos de erros categorizados com uso da ferramenta, selecionados pelos alunos do mestrado em tradução. São eles:

1. Homógrafos e polissemia
 - a. Lexemas polissêmicos, como “run”, “miss” e “play”
 - b. Homógrafos terminados em *-ing* (substantivo, verbo, adjetivo etc)
 - c. Sintagmas preposicionais/adverbiais
2. Sintagmas nominais
 - a. Substantivos pluralizados em posição inicial de frase
 - b. Sintagmas nominais com mais de um adjetivo
3. Sintagmas verbais
 - a. verbos modais e semimodais (“had better”, would rather”, “wonder if”, “would you mind”, “supposed to”, “need/s/ed/ing” e “have to”, “going to”, “can”)
4. “Lexical bundles”, definido pelas autoras como “combinatórias lexicais que atravessam as fronteiras sintáticas clássicas” (“I think that”, “is meant to”)
5. Outros
 - a. Estruturas interessantes como “to sing the baby to sleep”
 - b. Vários usos de “any”
 - c. “Questions tags”
 - d. There is/are
 - e. Verbos de emoção como “to be annoyed”

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

Morfologia e Léxico	
1. Substantivos a. Escolha Lexical b. Plural dos Substantivos c. Plural dos Substantivos Compostos d. Grau dos Substantivos 2. Adjectivos a. Escolha Lexical b. Plural dos Adjectivos c. Adjectivos Uniformes d. Grau dos Adjectivos - Comparativo e. Grau dos Adjectivos - Superlativo 3. Advérbios a. Escolha Lexical b. Grau dos Advérbios - Comparativo c. Grau dos Advérbios - Superlativo d. Locuções adverbiais 4. Determinantes a. Escolha Lexical b. Artigos c. Outros Determinantes 5. Pronomes a. Escolha Lexical b. P. Pessoais c. P. Pessoais (Função Sujeito) d. P. Pessoais (Função Predicativa) e. P. Pessoais (Função OD) f. P. Pessoais (Função OI - forma átona) g. P. Pessoais (Função OI - forma tónica) h. P. Pessoais (Função Objecto de circunstância)	(continuação) i. P. Pessoais (Forma combinada OI + OD) j. P. Possessivos k. P. Demonstrativos l. P. Relativos m. P. Interrogativos n. P. Indefinidos 6. Numerais a. Escolha Lexical b. Cardinais c. Ordinais 7. Preposições a. Escolha Lexical b. Contracção com Artigos c. Contracção com Pronomes d. Locuções Prepositivas 8. Verbos a. Escolha Lexical b. Tempos Simples c. Tempos Compostos d. Modo - poder, dever, ter de/que e. Voz Passiva f. Verbos Reflexivos g. Locuções Verbais 9. Conjunções a. Escolha Lexical b. Conjunções Coordenativas c. Conjunções Subordinadas d. Locuções Conjuncionais

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

Sintaxe	
1. Concordância - Gênero no interior do SN - Gênero entre o N e o Part. Pass. - Número no interior do SN - Número entre o SN e o Verbo - Número entre o N e o Part. Pass. 2. Ordem das Palavras - Interior do SN - Interior do SV - Interior do SP - Interior da Oração - Interior da Frase 3. Elisão - Artigo em início de frase - Artigo com nomes próprios - Artigo antes de um pronome possessivo - Sujeito - Pronome - Preposição 4. Coordenação - SNs SVs - Orações - Outra	5. Resolução incorrecta do POS-homógrafo -ing N/V (ex: building, dancing, writing) -ing N/Adj (ex: interesting, lasting) -ing V/Adj (ex: running, loving, working) -ing Adj/Ger_Part. Pres. (ex: promising, following) -ed/-en Adj/V_Part. Pass. (ex: alleged, employed) -ed/-en Adj/V_Pretérito (ex: experienced, broken) -ed V_Pret./V_Part. Pass. (ex: employed, transmitted) - Adv/Prep/Subordinador (ex: before) - Adv/Subordinador (ex: as) - Adj/Adv (ex: early) - N/V (ex: fight) - V/Adj (ex: narrow) - N/Adj (ex: red) - N/V/Adj/Adv/Prep (ex: round) - N/Adj/Adv (ex: weekly) - Prep/V (ex: following) - Prep/Adj (ex: opposite) - Prep/Adv/Adj (ex: outside) - Prep/Subordinador (ex: than) - Outra

3.3.3. Resultados dos estudos

Cada uma dos estudos apresenta muitas conclusões, variando de acordo com os níveis de análise, os problemas discutidos e os objetivos propostos, não sendo possível comentar todos eles aqui detalhadamente.

Em relação ao desempenho dos sistemas, *Dayrell* conclui que o *Globalink* é o mais eficiente e o *TraduzTudo*, o menos eficiente, em termos de quantidade de erros. Já no caso de *Oliveira et al*, à exceção do *Intertran*, o desempenho dos outros sistemas é considerado similar. A partir da análise das tabelas de frequências de erros lexicais, sintáticos e semântico-pragmáticos, podemos concluir que o *TranslatorPro* (que supomos ser o mesmo sistema que *Dayrell*

chama de *Globalink*) também foi o mais eficiente, apresentando o menor número de erros lexicais e sintáticos e o melhor desempenho em termos de interpretabilidade, no critério semântico-pragmático. O pior desempenho foi do *Intertran*, que aparentemente adota o método de tradução palavra por palavra, sem concordância entre substantivos e adjetivos ou substantivos e verbos, por exemplo. Como o objetivo do *TrAva* não é a comparação dos sistemas, não há comentários a esse respeito nos artigos.

A respeito da linguagem utilizada, *Dayrell* conclui que os sistemas apresentam melhor desempenho no texto de linguagem especializada do que na linguagem geral. A autora afirma que “no momento, a maneira mais provável de alcançar uma tradução de alta qualidade é através da seleção de um domínio, controle da linguagem de entrada e personalização do sistema” (Dayrell, 1999:152), o que corrobora uma das estratégias mais adotadas nas pesquisas de TA recentes (Martins *et al*, 2004).

O estudo de *Oliveira et al* aponta três causas principais para os problemas identificados, as quais consideramos que resumem também algumas conclusões das outras pesquisas: a primeira delas seria a inexistência de recursos lingüísticos nos sistemas ou a baixa qualidade daqueles disponíveis, exemplificados pelos problemas de dicionarização e pelas estratégias de acesso aos dicionários. A segunda causa seriam hipóteses equivocadas sobre as semelhanças e as diferenças entre o inglês e o português e sobre as supostas correspondências entre as línguas, ignorando-se que as estruturas semânticas são dependentes do contexto e da cultura. A terceira e última causa, que todos comentam ser a mais desafiadora, seria a própria complexidade do processo tradutório.

Sobre as categorias do *TrAva*, Maia&Barreiro concluem que a categorização de acordo com a terminologia sintática não dá conta de vários problemas de tradução. Além disso, ela exige que o avaliador tenha um grande conhecimento de lingüística para poder relacionar corretamente os problemas às categorias.

Dentre os problemas mais detalhadamente discutidos nas três pesquisas estão a homografia e a polissemia, a recuperação das relações de dependência e a falta de correspondência entre as estruturas de línguas diferentes.

Sobre a homonímia, a pesquisa de *Oliveira et al* conclui que somente a dicionarização não é suficiente, sendo necessário desenvolver estratégias de

seleção entre as várias opções dicionarizadas (desambiguação). Os autores observam também que essa questão tem implicações em vários níveis, podendo não afetar tanto o significado da oração se o termo tiver uma posição mais secundária, mas chegando a comprometer a coerência e a interpretabilidade da oração caso o item lexical ocupe uma posição mais central, como a de sujeito ou predicado.

Ainda a respeito da homonímia e da polissemia, tanto as pesquisas do *TrAva* (em Maia&Barreiro, no prelo) quanto a de *Dayrell* apontam como solução a melhoria do sistema de análise sintática (*parsing*). Apesar de ser satisfatória para os casos de homografia entre classes gramaticais diferentes, essa solução não esgota os problemas de homografia na mesma classe, o que é demonstrado pelo número crescente de estudos sobre desambiguação lexical (Ide&Véronis, 1998). Algumas referências para o português sobre esse tema encontram-se nas pesquisas de Lucia Specia (Specia&Nunes, 2004, e outras publicações disponíveis no *website* da pesquisadora: www.kmi.open.ac.uk/people/lucia).

Maia&Barreiro (no prelo) observam ainda que a resolução da polissemia é ainda mais complexa, exigindo que o sistema reconheça contextos mais amplos para determinar o significado apropriado.

A respeito das relações de dependência, em geral os autores consideram que esse ainda seja um problema de difícil resolução e sugerem que alguma estratégia que leve em consideração o contexto possa ser útil nesse sentido, além do investimento em maior formalização da análise sintática (Maia&Barreiro, no prelo) e dos processos de linearização (Oliveira *et al*, 2000).

Finalmente, a respeito da falta de correspondência entre as línguas, é interessante observar que praticamente todos os sistemas avaliados são baseados na tradução indireta por transferência, uma estratégia fundamentada nessa suposta correspondência. São sugeridas algumas soluções pontuais baseadas nos dicionários e nas regras gramaticais, mas a conclusão de Oliveira *et al* parece ser a mais abrangente: os sistemas baseiam-se em hipóteses errôneas sobre as semelhanças e as diferenças entre as línguas, sendo necessária uma mudança de perspectiva no tratamento das particularidades de cada língua.

De modo geral, as três pesquisas confirmam o senso comum sobre o desempenho dos tradutores automáticos para o português brasileiro: “os sistemas

disponíveis no mercado não funcionam corretamente e suas traduções raramente são úteis sem pós-edição ou conhecimento avançado da língua-fonte” (Oliveira *et al*, 2000, nossa tradução⁶). As traduções apresentadas confirmam o quadro descrito por Martins *et al* (2004, nosso grifo):

“...o resultado é visivelmente desalentador. E os problemas se espalham nas mais variadas direções: há falta de correspondências nos dicionários bilíngües que servem às ferramentas; há o clássico problema do tratamento das formas homônimas e das expressões idiomáticas; há a generalização, para a língua-alvo, de construções figuradas de validade restrita à língua-fonte; há o problema da ordem dos itens lexicais na sentença (principalmente no interior dos sintagmas nominais); há os problemas sintáticos relativos à concordância e à regência; a recuperação de relações anafóricas e o preenchimento de elipses; etc. O grande desafio da tradução automática para a língua portuguesa continua sendo, portanto, operar, com razoável eficácia, não apenas quando o português replica a estrutura do inglês, mas também nas situações - muito mais comuns, é preciso dizer - em que as duas línguas se distanciam” (p. 43).

O motivador, contudo, é que, mesmo diante dessas dificuldades, as pesquisas concluem que a TA é viável e a melhoria do seu desempenho, plenamente possível:

“...é absolutamente viável aprimorar a qualidade das traduções produzidas automaticamente, pois muitas deficiências encontradas nos sistemas podem ser resolvidas imediatamente”. (Dayrell, 1999:153)

“...observamos que os obstáculos não são de forma alguma intransponíveis, podendo ser facilmente superados se houvesse uma mudança de perspectiva ao considerar-se as idiosincrasias de cada língua” (Oliveira *et al*, 2000, nossa tradução⁷).

Para que tal melhoria seja alcançada, os estudos reconhecem a necessidade de formalização das línguas e de integração entre as mais diversas áreas:

“No entanto, para que a TA atinja uma qualidade satisfatória, são necessárias pesquisas não somente na área de tradução automática, como também nas diversas outras áreas que servem de apoio a elas, tais como a lexicografia, a sintaxe, a semântica, a pragmática e as ciências cognitivas”. (Dayrell, 1999:153)

⁶ “[...] the commercially available systems do not work properly, and their outputs are seldom useful without post-editing or expertise in the source language.”

⁷ “[...] we observed that the obstacles are not insurmountable at all, they could be easily overcome if there was a change in perspective when considering the idiosyncrasies of each language.”

“...o seu desenvolvimento [da pesquisa inicial com o TrAva] provou que este futuro depende da angariação e organização de uma equipe de linguistas, tradutores e engenheiros informáticos” (Maia&Barreiro, no prelo).

3.4. Comentários finais

Neste capítulo, apresentamos alguns dos principais aspectos da avaliação de tradução automática. A avaliação proposta pelo FEMTI aborda em detalhes muitos aspectos contextuais do uso de TA. Para aprofundar os aspectos da avaliação lingüística e reunir questões já discutidas sobre o português, apresentamos as pesquisas de Dayrell (1999); Oliveira *et al* (2000) e Sarmento (no prelo) e Maia&Barreiro (no prelo).

Os problemas de linguagem detalhadamente discutidos e as soluções sugeridas pelas pesquisas sobre português são fontes de informações úteis não só para a melhoria efetiva do desempenho dos sistemas, mas também para pesquisas nas diversas áreas afins, como tradução, processamento de linguagem natural, semântica, lingüística, lexicografia, ciências cognitivas, inteligência artificial, entre outras. Uma classificação quantitativa, como as de *Dayrell* e de *Oliveira et al*, pode ser útil para a priorização dos problemas a serem resolvidos, por exemplo, e o esforço de coleta de grandes quantidades de dados, como pretendido pelo *TrAva*, será relevante para a identificação das questões mais recorrentes.

O fato de as três pesquisas se concentrarem em sistemas baseados em tradução indireta por transferência indica que esse é o método mais utilizado nos sistemas disponíveis comercialmente e na Web, envolvendo português. Mesmo estando limitadas a esse tipo de sistema, as pesquisas mostram a diversidade e a complexidade dos problemas de TA e alguns enfoques possíveis para abordá-los, além de indicar as possibilidades e limitações desses sistemas.

Concordamos com Martins&Nunes (2005) quando afirmam que:

a ambigüidade não é um fenômeno periférico e marginal [...] nos enunciados em língua natural. Ela é constitutiva da própria linguagem, na medida em que todos os enunciados sofrem de vagueza e de indeterminação, se isolados os índices contextuais (relativos ao contexto extratextual) e co-textuais (relativos ao contexto intratextual) que provocam, com frequência, a ilusão de que os enunciados seriam exatos e precisos (p. 9).

Por esse motivo, segundo os autores, a delimitação do domínio e a elaboração de dicionários e gramáticas não são suficientes para solucionar todas as questões de ambigüidade na interpretação semântica da língua-fonte e na geração da língua-alvo. No entanto, como indica o estudo de *Dayrell*, o desempenho pode ser melhor com texto de linguagem mais especializada. Assim, consideramos que pode haver um ganho significativo de qualidade se os problemas a serem resolvidos na pós-edição, por exemplo, se limitarem àqueles que sejam mais complexos ou impossíveis de se resolver pela delimitação do domínio e pelos recursos lingüísticos. Entendemos que a finalidade da delimitação do domínio seja exatamente reduzir o universo de possibilidades semânticas dependentes do contexto extratextual.

Mesmo os sistemas voltados para a “linguagem geral”, que pretendam atender a um público mais amplo, poderiam recorrer a delimitações de domínio para melhorar o desempenho. Em termos de processamento computacional, a linguagem geral pode ser entendida como uma tentativa de dar conta de “qualquer linguagem”, ou de todas elas. No entanto, textos considerados de linguagem geral, mesmo que não incluam terminologia específica ou um jargão, tratam de um determinado assunto, em um determinado contexto e têm uma determinada estrutura. Os sentidos, então, não existiriam por si só, mas seriam construídos nesse contexto (Kilgarriff, 1997). O contexto, por sua vez, determinaria as informações, inclusive extralingüísticas (culturalmente compartilhadas pelos leitores), que precisam ser recuperadas para que o texto seja interpretado de forma semelhante por todos. Como os tradutores automáticos não são capazes de “construir” os sentidos e os sistemas baseados em conhecimento lingüístico não têm recursos extralingüísticos, um caminho viável pode ser a delimitação prévia do domínio. Se não resolve totalmente todos os problemas encontrados na tradução automática, a especialização lingüística, de implantação razoavelmente fácil no mercado de localização, pode trazer um avanço considerável no resultado final.