

4

Estudo de caso de tradução automática para o mercado de localização

Neste capítulo, fazemos a avaliação do uso de um tradutor automático no mercado de localização. Por isso, consideramos relevante aplicar uma proposta de avaliação que não ficasse restrita às questões lingüísticas e considerasse também os aspectos contextuais desse mercado. Para tal, recorreremos à taxonomia do FEMTI, apresentada no capítulo 3.

Além de incluir o aspecto contextual, outras diferenças entre a nossa avaliação e aquelas apresentadas no capítulo 3 são a utilização de apenas um sistema e a escolha de um texto de linguagem técnica, específico da área de localização, com terminologia de informática e telefonia celular. Com essa escolha, buscamos verificar também se a delimitação do domínio de fato facilita o processo de TA, além de observar algumas das características específicas das opções de interface e dos manuais.

Procuramos classificar os erros encontrados na tradução automática da lista de menus do aparelho e relacioná-los às categorias da taxonomia do FEMTI. Sempre que possível, são sugeridas soluções para as questões identificadas. Vale observar que as sugestões são pontuais, voltadas para os problemas encontrados nesse contexto limitado. Elas precisam ser testadas em *corpora* representativos da área para confirmar se efetivamente resolvem as questões sem provocar novos problemas em outras situações. Além disso, como não tivemos acesso ao programa, não é possível garantir se as soluções estão de acordo com a sua estrutura nem se poderão ser implementadas. A lista completa dos menus com as respectivas traduções é apresentada no anexo I.

O fato de nos concentrarmos nos aspectos pertinentes à utilização deste tradutor automático no mercado de localização de forma alguma significa que o programa tenha superado outras dificuldades relacionadas à linguagem dita geral. Acreditamos, contudo, que as pesquisas sobre avaliação que apresentamos no capítulo 3 fazem contribuições bastante relevantes para a melhoria do programa, o

que possibilita que nossa análise se concentre nos aspectos relacionados à localização.

Nas categorias lingüísticas, os *problemas* ou *erros* foram determinados com base nos padrões atualmente aceitos no mercado de localização, conforme apresentado resumidamente na seção sobre controle de qualidade, no capítulo 2, e mais detalhadamente discutido em Esselink (2000).

4.1. Avaliação com base na taxonomia do FEMTI

Nesta seção, aplicamos parte da taxonomia do FEMTI para avaliação do sistema de TA utilizado em nossa pesquisa. Como essa é uma avaliação externa, sem acesso ao programa, determinadas categorias não puderam ser avaliadas. Mesmo assim, procuramos comentar questões relevantes que deverão ser levadas em consideração numa avaliação completa. As categorias que dizem respeito ao funcionamento do software (Interoperabilidade, 2.2.1.4, até Coexistência, 2.2.6.5) foram excluídas deste capítulo. Foram mantidas apenas as categorias que consideramos relevantes comentar ou para quais tínhamos informações suficientes. Além disso, em alguns casos, a escolha de uma categoria eliminava outras. Por exemplo, no nosso caso a comunicação é assíncrona, então a categoria “Síncrona” foi excluída. Por isso, há saltos na numeração (p. ex.: de 1.1 para 1.1.5). A lista completa, com todas as categorias, foi apresentada no capítulo 3.

Vale observar que a taxonomia é complexa e que são sugeridas muitas métricas de avaliação para cada categoria, portanto uma avaliação completa seria muito longa e dependeria não só do acesso a informações confidenciais da empresa, mas também da participação de profissionais de áreas diferentes. Nesta avaliação inicial, comentaremos as categorias de forma mais livre, sem uma preocupação estrita com as métricas sugeridas, mas procurando considerar aspectos relevantes em relação a cada categoria.

Faremos uma avaliação mais detalhada das categorias lingüísticas, com exemplos do nosso *corpus* e sugestões de solução sempre que possível. Como as categorias lingüísticas do FEMTI são abrangentes e não dão conta de fenômenos específicos do português, do mercado de localização e dos domínios, propusemos algumas subcategorias, que estão indicadas com marcação alfabética. Por

exemplo, na categoria 1.5 Características da entrada (texto e autor) incluímos as subcategorias a) Formatação, b) Correção ortográfica e c) Variante lingüística.

CLASSIFICAÇÃO DO FEMTI

1 REQUISITOS DA AVALIAÇÃO

1.1 Propósito da avaliação

1.1.5 Avaliação declarativa

Esse é o propósito de avaliação que mais se encaixa em nosso estudo, por considerar o desempenho atual do programa em função de fatores como a abrangência lingüística e o tratamento de textos reais. As características do sistema que deverão ser avaliadas nesse caso são Modelos de processo de tradução, Recursos lingüísticos, Adequação, Precisão e Boa formação. Discutiremos todas elas, à exceção dos Recursos lingüísticos porque não tivemos acesso ao programa.

1.1.6 Avaliação operacional

Apesar de não termos acesso às informações necessárias para fazer esse tipo de avaliação mais técnica, consideramos que esse seja um aspecto crucial para a implantação efetiva da TA no mercado de localização, por isso decidimos comentá-lo. Segundo a definição, o foco da avaliação operacional é analisar se o sistema de TA irá de fato servir ao propósito pretendido no contexto operacional de uso. Uma das questões fundamentais é o custo-benefício de integrar o sistema de TA ao processo geral. As características a serem avaliadas são Recursos lingüísticos, Interoperabilidade, Confiabilidade, Facilidade de manutenção, Portabilidade e Custos, às quais não tivemos acesso.

1.2 Objeto da avaliação

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

Na nossa pesquisa, consideraremos o sistema de TA como parte do processo de localização. Na definição dessa categoria, os autores comentam que a avaliação irá variar muito de acordo com o processo ao qual o sistema será integrado e por isso essa questão ainda não é abordada na versão atual da taxonomia. Isso mostra

que essa integração é um campo de pesquisa a ser explorado e que a iniciativa da empresa de aplicar TA no processo de localização é realmente pioneira.

Apesar de ainda não haver métricas propostas para essa categoria, comentaremos alguns aspectos da integração em outras categorias, como a relação com a memória de tradução e com arquivos de formatos variados.

1.3 Características da tarefa tradutória

1.3.2 Divulgação

1.3.2.2 Publicação externa

1.3.2.2.2 Vários tipos de clientes

A finalidade da tradução no processo de localização é a **divulgação**, com **publicação externa** do material. O público-alvo de softwares é muito abrangente (vários tipos de clientes com necessidades diferentes). Outros aspectos sobre o público-alvo serão discutidos na categoria 1.4.2 “Consumidor da tradução”.

1.3.3 Comunicação

1.3.3.2 Assíncrona

No processo tradicional de localização, a tradução não é usada para interação simultânea entre usuários. Os produtos são primeiramente localizados para, então, serem comercializados.

1.4 Características dos usuários

1.4.1 Usuário da TA

Caso a TA seja integrada ao processo de localização, podemos esperar que haja vários usuários: engenheiros, que prepararão os arquivos, linguistas (provavelmente tradutores), que atualizarão as regras e os dicionários, e tradutores e revisores que serão responsáveis pela pré-edição e pela pós-edição do texto. A aceitação desses usuários poderá variar de acordo com as implicações que o uso do sistema tiver para cada área. A capacidade de integração com as ferramentas de memória de tradução, por exemplo, afetará a aceitação dos engenheiros. Os aspectos mais relacionados ao funcionamento do software, como utilização de recursos, frequência de travamento, facilidade de utilização, afetarão a aceitação de todos os usuários em relação à produtividade. Já a adequação do texto de entrada afetará os tradutores responsáveis pela pré-edição do material, enquanto a qualidade do texto gerado pelo sistema afetará os responsáveis pela pós-edição, por exemplo.

1.4.1.2 Proficiência na língua-fonte e

1.4.1.3 Proficiência na língua-alvo

Em geral, nas empresas de localização no Brasil, todos os envolvidos no processo têm algum conhecimento da língua-fonte (mais comumente o inglês). Tradutores, revisores e outros lingüistas envolvidos no processo são proficientes em ambas as línguas.

1.4.1.4 Familiaridade com o computador

Devido à própria natureza do processo de localização e aos diferentes softwares e formatos de arquivos utilizados nessa área, todos os envolvidos no processo têm total familiaridade com o computador.

1.4.2 Consumidor da tradução

1.4.2.1 Proficiência na língua-fonte e

1.4.2.2 Proficiência na língua-alvo

O perfil do consumidor de produtos localizados é muito amplo, variando de acordo com o tipo e a finalidade dos softwares e dos equipamentos de tecnologia disponíveis. Nesse caso, a proficiência na língua-fonte não será relevante porque os padrões de qualidade deverão ser os mesmos empregados no mercado profissional atualmente, então o uso de TA não deverá ser percebido pelo consumidor da tradução. Para tal, ainda que haja melhorias na qualidade do sistema de TA, faz-se necessário manter as etapas de controle de qualidade do material, que já fazem parte do processo de localização, além de incluir novas etapas, como a pré- e a pós-edição.

1.4.3 Usuário corporativo

A nossa avaliação é voltada para um usuário corporativo que é a empresa de localização que pretende integrar o sistema de TA ao processo de localização.

1.5 Características da entrada (autor e texto)

1.5.1 Tipo de documento

Nesse tópico, a taxonomia apresenta as subcategorias 1.5.1.1 Gênero e 1.5.1.2 Domínio/campo de aplicação. No entanto, Bruckner&Plitt (2001), num estudo em que aplicam a taxonomia do FEMTI para avaliar a integração da TA à memória de tradução no processo de localização, propõem a inclusão das seguintes subcategorias especificamente para esse mercado: *tipo* (ex.: tutoriais, manuais do usuário, referências para programadores), *domínio* (ferramentas para

escritório, sistemas CAD, software da área de negócios), *formato de arquivo* (RTF, Framemaker, XML, HTML etc) e *produto*. Seguindo essa subcategorização proposta pelos autores, o tipo de texto utilizado na nossa avaliação foi um manual de usuário de um telefone celular, disponível na internet. O arquivo original estava no formato pdf e teve de ser convertido para txt para ser processado pelo programa. O produto é um telefone celular, de uma das marcas mais populares no Brasil.

O formato do arquivo do texto de entrada é uma questão crucial em relação ao sistema de TA avaliado nesse estudo. Para o processamento do texto na memória de tradução em geral o arquivo é convertido do formato original para um formato aceito pela ferramenta de memória, como rtf. Nessa conversão, são incluídos códigos de formatação, que indicam as características do formato original para que elas sejam preservadas quando o arquivo for convertido de volta para o formato original após a tradução. Como o sistema de TA só aceita texto sem formatação, essas indicações das características originais seriam perdidas. Portanto, é necessário analisar se essa limitação do sistema pode ser alterada ou se é viável criar uma etapa de recuperação da formatação original ao final do processo.

A seguir, apresentaremos informações mais detalhadas sobre o texto selecionado, que serão relevantes para a avaliação lingüística.

Decidimos utilizar esse manual disponível na internet pela facilidade de obtê-lo e para evitar complicações referentes a direitos autorais e também para possibilitar o processamento eletrônico do arquivo, evitando, assim, dificuldades com digitalização de textos e correção dos problemas decorrentes da digitalização.

A princípio, um telefone celular pode parecer fugir do escopo da área de localização de software. No entanto, como discutido no capítulo 2, o termo *localização* atualmente se aplica ao processo de tradução de qualquer produto de tecnologia. Um telefone celular é um aparelho (um hardware) que executa um sistema operacional específico e acessa uma série de outros hardwares e softwares em rede para funcionar e disponibilizar serviços aos assinantes. Decidimos, então, optar por um manual de telefone celular porque em termos do processo de localização, o telefone celular apresenta uma estrutura similar à de um software

convencional, porém relativamente mais simples, o que facilita o nosso estudo. As informações exibidas no visor do aparelho são os itens de interface com o usuário e o manual é a documentação de apoio para utilização do aparelho. Um telefone celular tem muito menos itens de interface do que um software de uso geral, como o *Microsoft Word* ou *Outlook*, por exemplo, e a sua documentação também é mais simples, devido à menor complexidade de um aparelho básico e à maior familiaridade do público em geral com o uso do celular e o conteúdo do manual.

Outro fator decisivo para a escolha do *corpus* foi a minha experiência profissional com a tradução e a revisão de material da área de telefonia celular, o que facilita o reconhecimento de questões relevantes para a discussão e a sugestão de soluções.

Por fim, com a integração dos telefones celulares à internet, a relação entre um telefone celular e a área de informática fica ainda mais estreita, com a possibilidade de uso de *e-mail* e *download* de aplicativos, jogos e músicas, por exemplo, fazendo com que, em termos de localização, o processo e a linguagem utilizados fiquem cada vez mais próximos.

Utilizamos um manual cuja versão traduzida para português também estava disponível na internet. Assim, temos a tradução publicada, autorizada pelo fabricante, e, que, por isso, provavelmente emprega uma linguagem considerada adequada dentro dos padrões das traduções realizadas por tradutores humanos. Dessa forma, elas podem servir de referência para os resultados do tradutor automático.

O manual original em inglês tem 97 páginas, com um total de 17.889 palavras, de acordo com a análise da ferramenta de memória de tradução *Trados Translator's Workbench*.

O material a ser traduzido foi selecionado por nós e enviado à agência de localização de software proprietária do sistema de TA. Inicialmente, selecionamos dez manuais de fabricantes e modelos diferentes. A agência de localização nos forneceu a tradução de todos eles. No entanto, como a estrutura textual dos manuais era muito parecida e a delimitação do aspecto lingüístico a ser analisado revelou-se um grande desafio, decidimos utilizar apenas um dos manuais, considerado bastante representativo dos problemas encontrados no *corpus* geral. Mais especificamente, para os aspectos lingüísticos desta avaliação, selecionamos

apenas a lista de menus do aparelho apresentada no manual, visando limitar o escopo da avaliação. Como vimos no capítulo 2, a interface com o usuário é composta pelos ícones, menus, mensagens e caixas de diálogo exibidos no próprio aparelho. Por conter o texto de interação do usuário com o aparelho, a interface é um componente de fundamental importância no processo de localização. Quanto mais clara e simples for a interface, mais fácil será a utilização do produto, no nosso caso, o telefone. Vale observar que essa característica representa um grande desafio para a tradução automática, porque a concisão e as abreviações nos obrigam a fazer muito mais associações e inferências e a máquina não é capaz de fazê-las.

Além desse aspecto específico para o mercado de localização, a lista de menus mostrou ser também uma fonte relevante para avaliação de erros e desvios de língua geral recorrentes ao longo do manual e também em outros sistemas, conforme apresentado nas pesquisas do capítulo 3.

Para facilitar a análise do material e a identificação das respectivas traduções atribuídas pelo tradutor automático, foi solicitado a um engenheiro de localização que alinhasse o texto original e o texto traduzido automaticamente. Esse é o método usado no mercado de localização para gerar uma memória de tradução a partir de um texto original e sua respectiva tradução. Assim, o texto original foi dividido em segmentos (que em geral correspondem a uma oração ou título ou item de uma lista, dependendo da configuração da ferramenta) e alinhado ao texto traduzido. Foi gerado um arquivo de texto txt com esse conteúdo, que foi automaticamente exportado para o formato do *Excel*, para facilitar a organização dos segmentos originais ao lado das respectivas traduções.

No nosso *corpus*, identificamos três aspectos que não havíamos previsto do texto original que afetam diretamente a qualidade da tradução automática: a formatação, a correção ortográfica e a variante lingüística. Reiteramos que essas subcategorias não fazem parte da taxonomia do FEMTI e foram incluídas por nós.

a) Formatação

Exemplo:

Original – 1) 10.Enhancement settings1

Tradução automática - 10.Enhancement settings1

Tradução do fabricante – Configurações de acessório

Comentários – Nesse exemplo, duas questões de formatação do original afetaram a tradução, fazendo com que a opção de interface não fosse traduzida. A primeira, o número “10” na frente do item, indica que havia uma lista numerada. Como faltou um espaço entre a palavra e o número, o programa não reconheceu a palavra “Enhancement”. A segunda questão está relacionada à formatação de uma nota de rodapé, indicada pelo algarismo “1” junto à palavra “settings”. O algarismo foi mantido junto à palavra e ela não foi traduzida, porque não foi reconhecida. As palavras “enhancement” e “settings” foram traduzidas em outras ocorrências e, portanto, constam no dicionário do sistema.

Solução - Depende da capacidade de alteração do sistema para aceitar códigos de formatação que indicam negrito, itálico, fontes, aspas, listas numeradas ou com marcadores, notas de rodapé, entre outras, no original.

Outra solução seria fazer as alterações de formatação possíveis durante a pré-edição do original, acrescentando espaços entre as palavras e as numerações de lista e nota de rodapé, por exemplo.

b) Correção ortográfica

Exemplos:

Original – 1) Anykey 2) Keyguard 3) Multimedia msgs.

Tradução automática – Nos itens 1 e 2, as palavras foram mantidas em inglês. 3) Msgs de multimídia

Tradução do fabricante – 1) qualquer tecla 2) Proteção 3) Mens. multimídia

Comentários – Os erros de ortografia fazem com que as palavras não sejam reconhecidas pelo sistema e sejam mantidas em inglês. Poderíamos também considerar “anykey” e “keyguard” como neologismos ou alterações decorrentes da restrição de espaço, como é o caso da abreviação “msgs.”, para “mensagens”. Nesse caso, elas entrariam na categoria “fontes de erros intencionais” (do autor).

Solução – Uma opção para identificar os erros de grafia, abreviações e neologismos seria usar o corretor ortográfico no texto original e já listar ou alterar as palavras desconhecidas durante a pré-edição para inclui-las no dicionário, se necessário.

c) Variante lingüística

Exemplos:

Original – 1) dialled 2) Colour

Tradução automática – As palavras foram mantidas em inglês

Tradução do fabricante – 1) discado 2) Cores

Comentários – Aqui temos casos de diferença de grafia entre o inglês britânico e o inglês norte-americano. É provável que no dicionário conste a grafia “color”, por isso a forma “colour” não foi reconhecida.

Já no caso de “dial” e as variações “dials”, “dialled” e “dialling”, não foi possível compreender exatamente o comportamento do programa. Inicialmente, supomos que a tradução “discar” não constasse no dicionário, já que quase todas as ocorrências foram mantidas em inglês. Porém, verificando o manual completo, encontramos a tradução “discar” para “dialing”, na única ocorrência com grafia americana (com apenas um “l”). Supomos, então, que o dicionário tivesse ignorando a grafia inglesa, com dois “l”, mas como “dial” e “dials” também não foram reconhecidas, essa hipótese não se confirmou. Outra estranheza nesse caso é a tradução de “dials” como “dial” em “Speed dials”, traduzido como “Dial de velocidade”. As outras palavras que não constavam no dicionário foram mantidas exatamente como no original, então não conseguimos compreender por que o “s” final foi retirado da tradução nesse caso.

Solução – Uma opção seria usar o corretor ortográfico no texto original e já alterar as palavras de grafia britânica (caso o texto tenha predominantemente grafia norte-americana, ou vice-versa) durante a pré-edição.

1.5.2 Características do autor

Na área de localização, como em muitas áreas técnicas, a questão da autoria, como é considerada de forma tradicional na literatura e nos estudos da tradução, passa a ser secundária. Os textos são produzidos por redatores técnicos, engenheiros, especialistas e em geral não são assinados, o que dificulta as considerações sobre o autor. Em alguns contextos corporativos, a figura do autor poderia ser associada à imagem e ao discurso veiculados por uma empresa, como num website ou num portfólio. Contudo, não consideramos que esse seja o caso dos manuais técnicos de forma geral. Assim, optamos por excluir as subcategorias relacionadas ao autor.

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

O sistema de tradução automática utilizado neste trabalho pertence a uma agência de localização de software multinacional que possui escritório no Brasil. Por motivos de confidencialidade, a empresa nos concedeu acesso somente ao produto final do programa, ou seja, o texto traduzido. Tínhamos também o texto original, que foi selecionado por nós.

Algumas outras informações fornecidas sobre o tradutor automático foram: o sistema possui dicionários que podem ser ampliados e atualizados pelo usuário; é possível criar dicionários para domínios específicos, que são selecionados no momento da tradução; e é possível criar regras gramaticais para resolver as estruturas sintáticas. Com base nessas informações, podemos supor que este seja um sistema baseado em regras, exclusivamente em conhecimento lingüístico, sem análise de contexto, sem recursos estatísticos baseados em *corpora* nem outras fontes de conhecimento extralingüístico, como enciclopédias e bases de conhecimento.

As regras gramaticais parecem ser organizadas em algum tipo de hierarquia. Parece também que elas são desenvolvidas com base em casos específicos, apresentando soluções *ad hoc* que podem solucionar uma questão sintática pontual, mas gerar problemas em outras estruturas.

A partir da observação dos dados, não foi possível inferir a ordem de prioridade das regras nem a relação entre as regras lexicais e sintáticas. Várias palavras possuem mais de uma tradução no dicionário, já que foram traduzidas diferentemente em situações distintas. É provável que a opção do dicionário seja determinada pela estrutura sintática em que a palavra ocorre.

Podemos supor que esse sistema inicialmente tenha sido desenvolvido para a linguagem geral, dada a capacidade de inclusão de regras e dicionários, mas a intenção da empresa é utilizá-lo para fins específicos, como a tradução na área de localização. Fica clara, então, a relevância de procurar identificar questões

específicas dessa área e desses tipos de texto para contribuir para a melhoria da qualidade do sistema.

2.1.1.1.4 Modelos com memória de tradução

Como vimos no capítulo 2, as ferramentas de memória de tradução desempenham um papel fundamental no mercado de localização e a sua integração ao sistema de TA é uma prioridade para viabilizar e justificar sua aplicação no mercado de localização. Alguns sistemas já oferecem algum nível de integração às ferramentas de memória de tradução, como o LOGOS e o SYSTRAN que são compatíveis com o *Trados Translator's Workbench* (Esselink, 2000:395). Neste processo, o computador busca uma tradução para o segmento na memória de tradução. Caso não encontre, será necessária a intervenção do tradutor para solicitar a tradução do segmento ao sistema de tradução automática, editá-la e, então, atualizar a memória com o novo segmento (Esselink, 2000:395). Existem também sistemas de TA desenvolvidos especificamente para serem integrados a memórias de tradução, como o PROMT, desenvolvido para a *Trados* pela empresa PROMT, da Rússia (Shadbolt, 2003:9). É provável que esse tipo de sistema seja mais adequado para o mercado de localização, mas não há relatos de que estejam sendo usados no mercado brasileiro.

Com base nas características básicas do sistema, como a capacidade de processar somente texto em formato txt, podemos supor que o tradutor automático utilizado neste trabalho não foi desenvolvido para ser integrado às ferramentas de memória de tradução. Essa possibilidade de integração dependerá de uma avaliação interna do sistema, a ser realizada pela equipe técnica especializada no sistema em conjunto com a equipe de engenharia de software, e foge ao escopo do nosso trabalho.

Dentro da definição dessa categoria no *website*, encontramos o seguinte comentário dos autores: “A incorporação da memória de tradução às plataformas tradicionais de tradução automática é um campo de estudo relativamente recente e ainda pouco explorado”.⁸

⁸ “The incorporation of translation memory into traditional machine translation platforms is a relatively new and under-represented field of study, although a few examples do exist.” (FEMTI: <http://www.isi.edu/natural-language/mteval/>).

2.1.1.2 Modelos

2.1.1.2.2 Transferência

O tradutor automático da pesquisa é baseado em tradução indireta por transferência.

2.1.2 Recursos lingüísticos

2.1.2.1 Idiomas

Sabemos que o sistema é multilíngüe, mas a nossa pesquisa envolveu somente o par inglês-português. Como o sistema é baseado em regras, a direção da tradução é relevante, nesse caso, do inglês para o português.

2.1.2.2 Dicionários

Fomos informados de que o sistema possui dicionários gerais e dicionários específicos para domínios diferentes, como informática, medicina e direito ou até mesmo para subdomínios, como hardware, cardiologia e direito societário, por exemplo. Os dicionários podem ser ampliados e revisados e podem ser criados novos dicionários conforme necessário. Nos dicionários, é possível incluir informações sintáticas, como a transitividade dos verbos.

2.1.3 Características do processo

2.1.3.1 Atividades de preparação da tradução

Tais atividades deverão ser planejadas pela empresa de localização, de acordo com a capacidade interna do sistema para aceitar diferentes formatos de arquivos. Como vimos, atualmente o sistema só aceita o formato txt, então seria necessária uma etapa de conversão dos outros formatos normalmente utilizados nesse mercado. Pode haver também a necessidade de pré-edição do texto, para resolver ambigüidades, fazer a revisão ortográfica (para evitar que o sistema erre a tradução por erros de ortografia do original) ou alterar formatações que comprometam a tradução gerada, por exemplo.

2.1.3.2 Atividades posteriores à tradução

Como vimos no capítulo 3, a maioria dos sistemas de TA ainda produz textos que precisam ser submetidos à pós-edição para atender aos padrões de qualidade geralmente aceitos. A pós-edição, então, deverá ser uma das principais atividades posteriores à tradução. Outras atividades poderão ser a conversão do texto para o formato de arquivo original e a formatação do arquivo.

A respeito da pós-edição, vale notar que o cálculo de produtividade de um revisor no mercado de localização é de aproximadamente 5.000 palavras por dia, revisando um texto produzido por um tradutor humano. As questões que se colocam, então, são como conseguir que a máquina produza um texto com qualidade mais próxima possível do texto do tradutor profissional, para que a produtividade do revisor não seja afetada, e se esses esforços irão efetivamente representar redução de tempo e custo. Além disso, os revisores deverão ser treinados em novas habilidades, já que muitos dos erros da máquina serão de natureza diferente dos erros dos tradutores humanos. Contudo, antes que esse tipo de avaliação seja feito, é preciso investir na melhoria da qualidade da tradução com ajustes básicos, sobretudo de regras e dicionários, como aqueles sugeridos nas pesquisas apresentadas no capítulo 3 e os discutidos nas próximas seções.

2.1.3.3 Atividades interativas

Acreditamos que esse sistema não ofereça opções de interação durante a tradução.

2.1.3.5 Gerenciamento do processo

Caberá à empresa e aos gerentes de projeto fazer as adaptações necessárias ao processo de localização para a integração do sistema de TA, incluindo e planejando as etapas necessárias para formatação, conversão, pré-edição, pós-edição etc.

2.2 Características externas do sistema

2.2.1 Funcionalidade

Parte desta categoria se concentra na avaliação lingüística do sistema. Como explicado na categoria 1.5 Características da entrada, os exemplos comentados foram retirados das traduções geradas pelo sistema para a lista de menus do telefone celular, apresentada no manual do usuário. Vale lembrar que as subcategorias lingüísticas, indicadas com marcação alfabética, não fazem parte da taxonomia do FEMTI e foram incluídas por nós para classificação e discussão dos problemas específicos ao nosso *corpus*. Por exemplo, na categoria Consideração dos fenômenos específicos ao *corpus*, as subcategorias a) Emprego do modo imperativo, b) Alteração de classe gramatical e c) Tratamento de siglas foram elaboradas por nós e não fazem parte da taxonomia do FEMTI.

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 e

2.2.1.1.1.4 Coesão

A taxonomia oferece várias sugestões de métricas para avaliação dessas quatro categorias: legibilidade, inteligibilidade, coerência e coesão especificamente no texto traduzido. Nessa avaliação inicial, apresentaremos algumas impressões gerais, sem aplicar nenhuma métrica específica.

A lista de menus é composta por itens isolados. Uma leitura de outras passagens do manual traduzido revela que essas quatro categorias ficam bastante comprometidas devido aos erros, de natureza variada, no nível sentencial. Acreditamos que uma avaliação desses aspectos textuais só será possível depois que alguns ajustes básicos de regras e dicionários forem realizados para melhorar a qualidade da tradução no nível sentencial.

2.2.1.1.2 Contrastivo/interlingual

Nesta categoria, o foco não está somente no texto traduzido, mas no processo de tradução. Avalia-se principalmente a capacidade do sistema de lidar com as diferenças mais comumente reconhecidas entre as línguas em questão. Costumam ser usados conjuntos de testes, em geral produzidos pelos próprios fabricantes, que contêm textos especificamente criados ou selecionados para testar fenômenos pontuais, como tempos verbais, homógrafos ou o uso de pronomes, por exemplo (Maia&Barreiro, no prelo). Essas questões são consideradas nas próximas duas subcategorias, 2.2.1.1.2.1 e 2.2.1.1.2.2.

2.2.1.1.2.1 Consideração dos fenômenos específicos ao *corpus*

De acordo com a definição, nesta categoria é avaliada a capacidade do sistema de lidar com os desafios específicos apresentados pelo *corpus* de interesse. Um *corpus* formado por material da área de localização apresenta muitas especificidades. Algumas já foram comentadas anteriormente, como a variedade de formatos de entrada e as características das opções de interface com

o usuário. Outra característica que não é de forma alguma exclusiva deste mercado, mas que faz parte do próprio conceito de localização, é a adaptação. Os exemplos clássicos são aqueles mais pontuais e que ocorrem no próprio texto, como convenções de hora e data e símbolos de moedas, mas as adaptações podem afetar aspectos mais gerais, como a reorganização, a inclusão ou a exclusão de informações. Contrastando a versão original do manual e a versão traduzida disponibilizada pelo fabricante na internet, podemos observar que várias adaptações mais estruturais foram feitas. Por exemplo, a versão original do manual tem 97 páginas enquanto a tradução tem 63 páginas, portanto deve ter havido corte de conteúdo. Enquanto no original as informações sobre bateria, acessórios e cuidados e manutenção estavam distribuídas em seções independentes, na tradução essas informações foram reunidas numa mesma seção, intitulada “Referências”. Foi também incluída uma seção intitulada “Termos comuns”, com uma pequena lista de siglas, seguidas pelo extenso em inglês e a tradução. Obviamente, não podemos esperar que o sistema de TA faça essas adaptações. Elas, em geral, são determinadas pelo cliente final, da área técnica ou de marketing do fabricante do produto. Contudo, consideramos importante registrar que as adaptações não deixarão de ser necessárias caso o uso de TA passe a fazer parte do processo de localização.

A respeito das opções de interface com o usuário, que são o objeto da nossa análise lingüística, é importante observar que o reconhecimento desses itens como “entidades”, com uma identidade que precisa ser preservada tanto no aparelho quanto nas referências em manuais ou outro material de referência, é fundamental para a qualidade do produto. A partir da lista de menus do telefone, pudemos observar que o sistema não está preparado para reconhecer essas entidades como tal nem para lidar com suas peculiaridades lingüísticas da forma como elas são tratadas no mercado de localização, como veremos nos exemplos apresentados.

As próximas três categorias foram criadas por nós com base em exemplos retirados do *corpus*: a) Emprego do modo imperativo, b) Alteração de classe gramatical e c) Tratamento de siglas. Apesar de as duas primeiras serem questões gramaticais e a segunda ser abordada nos guias de estilo no mercado de localização, optamos por não considerá-las como erros do sistema. As soluções

não estavam lingüisticamente erradas, mas foram inadequadas nesse contexto dos itens de interface, e, portanto, são fenômenos específicos ao *corpus*.

a) Emprego do modo imperativo

Exemplo:

Original – Create message

Tradução automática – Crie mensagem

Tradução do fabricante – Criar mensagem

Comentários – Neste caso, o problema é a opção pelo modo imperativo para o verbo, quando o usual na área de localização de software é o emprego do infinitivo (como os menus “Editar”, “Exibir” e “Inserir” do *Microsoft Word*, por exemplo). O uso do infinitivo justifica-se porque estes verbos indicam as ações que o usuário *pode* realizar e não aquelas que ele deve realizar obrigatoriamente. O uso de imperativo se justificaria se os verbos indicassem instruções a serem seguidas em vez de opções, mas no caso dos itens da lista de menus, a forma usada seria o infinitivo.

Observe, contudo, que não podemos fazer uma generalização do uso do infinitivo nos itens de interface, já que as mensagens exibidas na tela em geral contêm instruções e, portanto, irão exigir o emprego do modo imperativo.

Além dos itens de interface especificamente, os próprios manuais estão repletos de instruções para os usuários, apresentando uma alta frequência do uso do imperativo. Assim, o emprego do infinitivo para as opções de interface não poderá ser determinado por uma regra geral.

Solução – Neste caso, será necessário analisar, em um contexto mais amplo, outras características das opções de interface, que possam contribuir para a criação de uma regra para o emprego do infinitivo nesses casos específicos. Uma opção que podemos sugerir seria associar o uso do infinitivo a estes sintagmas verbais mais curtos, muitas vezes sem sujeito ou sem outros complementos, como “Create message” e “Delete folder”.

Outra hipótese é que já haja no sistema uma regra que determine o uso do imperativo em frases sem sujeito. Caso tal regra exista de fato, ela precisaria ser alterada.

b) Alteração de classe gramatical

Exemplo:

Original – Copy

Tradução automática – Cópia

Tradução do fabricante – Copiar

Comentários – Esse é um caso em que a relevância do contexto fica patente para solucionar a ambigüidade provocada pelas formas homógrafas “Copy” como verbo (“Copiar”, “Copie” ou “Copiem”) e “Copy” como substantivo (“Cópia”). Nas opções de interface, em geral encontraremos entidades e ações a serem realizadas com essas entidades. Só o contexto poderá determinar se “Copy” é uma entidade ou uma ação. Podemos supor que num telefone celular, a ação “Copiar” (um número de telefone, um contato, uma mensagem) seja mais freqüente do que a entidade “Cópia”, provavelmente mais recorrente numa máquina de fotocópias, por exemplo.

Solução – Fazer um estudo de freqüência da palavra “Copy” no *corpus* para definir se a prioridade no dicionário deverá ser “Cópia” ou “Copiar”.

c) Tratamento de siglas

Exemplos:

Original – 1) SIM card 2) GPRS connection 3) PIN code

Tradução automática – 1) cartão de SIM 2) conexão de GPRS 3) código de ALFINETE

Tradução do fabricante – 1) cartão SIM 2) conexão GPRS 3) código PIN

Forma por extenso e tradução das siglas:

GPRS – General Packet Radio Service, Serviço geral de rádio em pacotes (tradução do fabricante, disponível no manual)

SIM – Subscriber Identity Module, Módulo de identificação do assinante (tradução do fabricante, disponível na seção de perguntas freqüentes do *website* corporativo)

PIN - Personal Identification Number, Número de Identificação Pessoal (tradução do fabricante, disponível na seção de perguntas freqüentes do *website* corporativo)

Comentários – Siglas representam um problema também para tradutores humanos. Em geral, elas são utilizadas para referência a entidades como

tecnologias, serviços, produtos padrão do mercado ou próprios da empresa. As dificuldades para tradução são decorrentes principalmente da não-determinação do gênero em inglês, que precisa ser identificado para ser explicitado em português. Como nem sempre a forma por extenso é apresentada, o tradutor humano precisa buscar outras fontes para determinar o gênero.

Outra questão determinada pelo uso no domínio é a tradução ou não da sigla original. Podemos observar uma tendência maior ao emprego das formas em inglês, mas algumas siglas, como PSTN (Public Switched Telephone Network) tiveram suas traduções consagradas em português: *RTPC*, de *Rede telefônica pública comutada*. (Tradução disponível no glossário online da Agência Nacional de Telecomunicações – Anatel).

Podemos supor que o emprego dessas formas sem preposição tenha sido determinado pelo uso e pela referência à tecnologia subjacente ao produto ou serviço em si, já que, se pensarmos nos termos por extenso, a sintaxe das formas que foram consagradas em português não se justificam, como “conexão GPRS” e “cartão SIM”. Numa construção típica em português, os sintagmas “conexão serviço geral de rádio em pacotes” e “cartão módulo de identificação do assinante” seriam agramaticais.

Solução – Uma opção a ser verificada seria incluir uma regra que restrinja o uso de preposição antes de siglas, observando-se as conseqüências dessa decisão em um *corpus* mais extenso e com ocorrência de outras abreviações.

Outra solução mais pontual seria a inclusão das expressões completas no dicionário, como “GPRS connection”, “GPRS data”, “PIN code” e “SIM card”.

Uma opção para evitar que a sigla seja confundida com uma palavra homônima, como PIN, que foi traduzido como “alfinete”, seria estabelecer uma regra que interpretasse como sigla palavras isoladas em caixa alta. Essa solução requer uma verificação mais detalhada em um *corpus* mais extenso, visto que é comum em manuais em inglês o emprego de caixa alta com a função de ênfase em títulos e palavras tais como “WARNING”, “ATTENTION”, “DO NOT” etc.

2.2.1.1.2.2 Estilo

Segundo a definição, esta categoria avalia mais especificamente a adequação da tradução em termos de registro. Assim como no caso das categorias legibilidade, inteligibilidade, coerência e coesão, consideramos que a adequação

de registro somente poderá ser avaliada quando a qualidade da tradução no nível sentencial tiver sido aprimorada.

2.2.1.2 Precisão

2.2.1.2.1 Fidelidade

A própria definição da categoria reconhece que ela seja subjetiva e faz referência a vários estudos que tentam definir uma métrica para avaliar a fidelidade. Os autores comentam também que a fidelidade está diretamente ligada à inteligibilidade. Como comentado anteriormente, consideramos que essas categorias textuais somente poderão ser consideradas com a qualidade da tradução tiver melhorado no nível sentencial.

2.2.1.2.2 Uniformidade

Nessa taxonomia, a uniformidade, ou padronização, diz respeito à capacidade do programa de gerar sempre a mesma saída para uma mesma entrada. Na lista de menus, não foi possível observar nenhum caso dessa categoria.

2.2.1.2.3 Terminologia

Como os menus do telefone celular contêm grande parte da terminologia específica, teremos muitos exemplos do *corpus* para essa categoria. Eles estão divididos em três subcategorias sugeridas por nós: a) Vocabulário do domínio, b) Vocabulário ambíguo no próprio domínio e c) Vocabulário padrão do mercado de localização. Vale ressaltar que neste trabalho empregamos a palavra “termo” conforme é feito na área de localização e não de acordo com definições da área de terminologia. No mercado de localização, “termo” é usado tanto para palavras do jargão, que designam conceitos bem-definidos em determinada área do conhecimento, quanto para palavras usadas frequentemente ou cuja tradução precisa ser padronizada para evitar ambigüidade ou para sempre remeter ao mesmo conceito ou função (do programa, por exemplo), mesmo que não seja estritamente um termo científico. Um exemplo é o caso das palavras “Selection” e “Choice”, que será discutido no item b a seguir.

a) Vocabulário do domínio

Exemplos:

Original – 1) List of menu functions 2) Screen saver timeout

Tradução automática – 1) Lista de funções de cardápio 2) Esconda intervalo de poupador

Tradução do fabricante – 1) Lista de funções do menu 2) Tempo limite do descanso de tela

Comentários – Conforme discutimos em Specia *et al.* (2006), as variações de sentidos, ou usos, de uma palavra numa língua não necessariamente se mantêm ou são traduzidas por uma única palavra em outra língua, o que representa um desafio para aplicações bilíngües ou multilíngües, como a tradução automática.

Dependendo da estrutura interna dos dicionários, pode ser difícil ou mesmo impossível definir mais de uma tradução para a mesma palavra original. Mesmo que seja possível incluir entradas diferentes para a mesma palavra a fim de atribuir traduções diferentes a cada uma (por exemplo: menu – sentido 1 = cardápio e menu – sentido 2 = menu), a dificuldade reside em como o sistema escolherá entre uma das opções na interpretação do original, ou seja, em como ele fará a desambiguação de sentido. Esse é um desafio especialmente complexo para sistemas de TA voltados para uso geral, que precisam lidar com os vários usos de uma mesma palavra em diferentes contextos.

Uma opção para contornar esse problema é a delimitação do domínio. Como observam Ide&Véronis (1998), a determinação do domínio é fundamental para a desambiguação lexical do sentido. Delimitado o domínio, é possível criar um dicionário específico, em que constem apenas as traduções mais freqüentes naquele contexto. Nos domínios de informática ou telecomunicações, é muito pouco provável que seja necessário utilizar a palavra “cardápio”, portanto não há necessidade de manter essa opção no dicionário do domínio. Outros exemplos são “cela”, como tradução de “cell”, em vez de “célula” ou “celular”; “colocações”, como tradução de “settings”, em vez de “configurações”; e “ligações”, como tradução de “links”, em vez de manter “links”.

Por fim, o fato de o termo “screen saver” (“Proteção de tela”) não constar no dicionário fez com que cada palavra fosse interpretada e traduzida isoladamente, “screen” como “esconder” e “saver” como “poupador”.

Solução – Inclusão da tradução mais freqüente no domínio no dicionário. No caso de “screen saver”, a solução seria incluir esse termo no dicionário. Entre a opção do fabricante, “Descanso de tela”, e a tradução padrão da *Microsoft*, “Proteção de tela” sugerimos a inclusão da tradução da *Microsoft*, por ser mais utilizada no mercado de localização.

b) Vocabulário ambíguo no próprio domínio

Exemplos:

Original – 1) Operator selection 2) Home 3) time

Tradução automática – 1) Seleção de operador 2) Casa 3) tempo

Tradução do fabricante – 1) Escolha de operadora 2) Homepage ou Sistema local e 3) hora

Comentários – Aqui se repete o problema de usos diferentes de uma mesma palavra no original exigirem traduções diferentes. Nesses casos, contudo, existe uma ambigüidade de termos dos próprios domínios de telefonia celular e informática, a saber “operador” (o atendente do *call center*, por exemplo) ou “operadora” (a empresa de telefonia que disponibiliza o acesso à rede), como traduções de “operator”, e “homepage” (página principal de um *website*) ou “sistema local” (sistema ou área em que o telefone celular está registrado), como traduções de “home”. Também incluímos nessa categoria a palavra “time”, que, apesar de não ser específica de um domínio, é muito freqüente em informática e telefonia celular.

Em alguns contextos, como “operator logo”, é fácil para os tradutores humanos supor que a tradução seja “operadora”. Já no caso do termo “operator selection”, o leitor ou o tradutor podem considerar que haja ambigüidade entre “seleção de operadora” ou “seleção de operador”.

No caso de “home” e “time”, em geral o contexto esclarece a ambigüidade para o tradutor humano. A tradução “casa” precisaria ser mantida porque a maioria dos aparelhos oferece a opção de gravação de vários números para um mesmo contato, associando-se cada número a uma referência, como “Casa”, “Trabalho” ou “Fax”.

Solução – No caso de “operator”, assim como nos casos anteriores, a solução pode ser manter “operadora” como primeira opção no dicionário, se possível, ou então manter apenas essa opção, a fim de reduzir o número de correções na pós-edição. Outra alternativa é incluir sintagmas mais longos no dicionário, ou seja, uma entrada para “operator selection” e outra para “operator logo”. Esta última opção parece ser a melhor também para os termos “home” e “time”, fazendo com que “home” isoladamente tenha como primeira tradução “homepage” e “home system” seja traduzido como “sistema local”. A tradução

“casa” pode ser incluída como segunda opção se o sistema tiver uma forma de considerar o contexto (da agenda de contatos, no caso) ou pode não ser incluída, já que esse caso deve ser bem pontual e poderia ser alterado durante a pós-revisão.

Já para “time”, uma opção seria verificar a frequência de cada uso para priorizar as traduções no dicionário, além de incluir no dicionário expressões ou colocações, como “date and time”.

Caso haja alguma ambigüidade mais complexa (não identificada nos itens de interface analisados neste trabalho), uma solução seria avaliar o contexto para tentar determinar o uso específico ou consultar o cliente ou um especialista para definir o termo e incluí-lo no dicionário.

Novamente, uma análise baseada na frequência de cada uso pode ser útil para priorização no dicionário e determinação de outras características que favoreçam a seleção da tradução mais adequada. No caso deste programa de TA, tais informações poderiam ser relevantes para a elaboração de regras gramaticais que ajudassem a determinar a tradução a ser empregada.

c) Vocabulário padrão do mercado de localização

Exemplos:

Original – 1) Memory status 2) PIN code request

Tradução automática – 1) Estado de memória 2) Pedido de código de ALFINETE

Tradução do fabricante – 1) Status da memória 2) Solicitação de código PIN

Comentários – Na área de localização, em que predominam as traduções determinadas pela *Microsoft*, “status” é traduzido como “status”, e não “estado”, e “request” é traduzido como “solicitação”, e não como “pedido”.

Acreditamos que esta categoria mostre a diferença entre *domínio*, como informática ou telefonia celular, e *processo de localização*. Enquanto as subcategorias a e b discutiram termos específicos dos domínios de informática e telefonia celular, esta subcategoria exemplifica uma padronização imposta pelo processo de localização. A diferença básica é que manter “cardápio” como tradução de “menu” seria um erro no domínio da informática, ao passo que “estado” ou “pedido” *a priori* seriam traduções corretas. As traduções padrão no mercado de localização foram determinadas arbitrariamente e foram impondo-se ao longo do tempo, podendo ser consideradas preferenciais. No entanto, como

vimos no capítulo 2, o uso de glossários e a padronização são fundamentais na área de localização e, por isso, a não-utilização da terminologia padrão é considerada um erro grave mesmo para tradutores humanos.

Outro aspecto interessante neste caso é o fato de os telefones celulares poderem ser incluídos tanto no domínio da informática como da telefonia, devido às tecnologias utilizadas. Assim, no processo de localização, os dicionários e glossários acabam precisando combinar o vocabulário da informática e do domínio específico do produto, seja um telefone celular, uma câmera digital, um aparelho de MP3 ou um forno de microondas.

Uma última questão a ser comentada é que nem sempre a localização dos telefones celulares é realizada por empresas especializadas no processo, o que resulta em algumas variações em relação aos termos utilizados no mercado de localização. Alguns exemplos são as traduções “Marcadores” em vez de “Favoritos”, para “Bookmarks”, e “Escolha” para “Selection”, em lugar de “Seleção”. No mercado de localização, em geral “Choice” é traduzido como “Escolha” e “Selection”, como “Seleção”.

Solução – Aqui se aplica a mesma solução do item anterior, já que para o sistema o fato de ser um consenso de uso ou uma preferência do mercado não seria relevante. Portanto, bastaria alterar a tradução da palavra “status” de “estado” para “status” e de “request” de “pedido” para “solicitação” no dicionário. Essa solução pode ser aplicada a outros termos com base em frequência no *corpus*.

2.2.1.3 Boa formação

2.2.1.3.1 Pontuação

Temos dois exemplos do *corpus* para essa categoria.

Exemplos:

Original – 1) Multimedia msgs. 2) Organise 'Go to' options

Tradução automática – 1) Msgs de multimídias. 2) Organize 'Go a opções de ‘

Tradução do fabricante – 1) Mens. multimídia 2) Organizar opções de “Ir para”

Comentários – Nesses casos, observamos que a pontuação não foi corretamente interpretada pelo sistema. O ponto que indica a abreviação da

palavra “messages” como “msgs.” foi interpretado como ponto final, sendo erradamente colocado no final do sintagma, depois da palavra “multimídia”. Esse problema acontece em outras ferramentas também, como o *Word*, cuja função de autocorreção altera automaticamente a primeira letra da palavra que se segue a um ponto para maiúscula, o que não procede no caso de abreviações.

Já no caso das aspas, não apenas na lista de menu, mas também em outras ocorrências no manual, observamos que o sistema não as mantém ligadas às mesmas palavras do original. Inicialmente, supomos que o problema estivesse relacionado ao fato de as aspas simples que precedem a palavra “Go” estarem invertidas. Contudo, analisando outras instâncias do manual original, vemos que o problema também ocorre em casos em que o sinal foi usado corretamente.

No exemplo mencionado, vale comentar que as aspas buscam identificar a ocorrência de um item de interface dentro de outro item de interface, a saber “Ir para” dentro da opção de menu “Organizar opções de ‘Ir para’”.

Solução – No primeiro caso, uma opção seria criar regras para reconhecimento de abreviações, para que o ponto fosse mantido junto à abreviação e não interpretado como ponto final. Outra opção seria identificar as abreviações mais frequentes nesse domínio em grandes *corpora* e inclui-las no dicionário.

Uma solução mais abrangente seria criar regras para que o ponto não fosse interpretado como ponto final quando cabível. Para isso, também seria necessário observar em grandes *corpora* em que circunstâncias isso ocorre.

No caso das aspas, uma opção seria indicar que elas devem ser mantidas com os mesmos sintagmas a que estavam ligadas no original. Essa solução, contudo, exigiria que o sistema fosse capaz de reconhecer os sintagmas e as traduções atribuídas a eles, o que é um recurso bastante complexo e em muitos casos ainda não disponível. Enquanto uma solução automática não for identificada, será necessário verificar este item durante a pós-edição.

2.2.1.3.2 Léxico/escolha lexical

Na taxonomia do FEMTI, não encontramos nenhuma categoria para incluir o caso das palavras que não constavam no dicionário e foram mantidas em inglês. Optamos por considerar neste item. As subcategorias propostas por nós são a) Escolha lexical inadequada e b) Palavras que não constam no dicionário.

a) Escolha lexical inadequada

Exemplos:

Original – 1) Speed dialling 2) Keypad 3) Sequence

Tradução automática – 1) Faça andar depressa dialling 2) teclado complementar 3) Sucessão

Tradução do fabricante – 1) Discagem rápida 2) teclado 3) Sequência

Comentários – Podemos supor que o dicionário do programa tenha sido inicialmente atualizado com base em um dicionário bilíngüe de uso geral, não apenas por reconhecer quase todas as palavras do manual, mas especialmente pelo fato de conter traduções de usos pouco prováveis em contexto técnico, como “corcel” para “charger”.

A palavra “speed” consta no dicionário como substantivo, tendo sido traduzida como “velocidade” em várias ocorrências ao longo do manual. No entanto, como verbo, podemos supor que a tradução incluída no dicionário seja “Fazer andar depressa”, por conta da tradução atribuída ao item de interface “Speed dialling”, mencionada acima.

No caso de “keypad”, apesar de, em determinados contextos, haver uma diferenciação terminológica entre “keyboard” (“teclado”) e “keypad” (“teclado numérico”), ambos os termos são traduzidos por “teclado” no uso mais geral e em várias ocorrências dos glossários da *Microsoft*. Portanto, não há a necessidade da palavra “complementar” nesse caso, especialmente no caso de um telefone celular, em que há apenas um teclado.

Já o termo “sequence” foi sempre traduzido como “sucessão”, apesar de a tradução “seqüência” ser a mais freqüente.

Solução – As soluções pontuais seriam retirar a palavra “complementar” da tradução de “keypad” no dicionário, alterar a tradução de “speed” como verbo de “fazer andar depressa” para uma opção mais concisa e genérica, como “agilizar” e adotar “seqüência” como primeira ou única opção de tradução para “sequence”. Essas opções poderiam ser verificadas em *corpora* e em glossários da área de informática.

b) Palavras que não constam no dicionário

Exemplos:

Original – 1) connectivity 2) inbox

Tradução automática – As palavras foram mantidas em inglês.

Solução – No caso das palavras pertinentes, uma solução seria gerar listas de termos a partir de grandes *corpora* do domínio para atualizar os dicionários.

2.2.1.3.3 Gramática/sintaxe

Com base nos exemplos do nosso *corpus*, sugerimos duas subcategorias sintáticas: a) Sintagmas nominais e b) Sintagmas verbais. Os sintagmas nominais estão subdivididos em a.1) Alteração de classe gramatical, a.2) Não-diferenciação entre participípio e pretérito perfeito, a.3) Ordem das palavras no sintagma, a.4) Emprego do artigo definido e a.5) Concordância de gênero e número. Nos sintagmas verbais, temos a subdivisão b.1) Regência verbal.

a) Sintagmas nominais

a.1) Alteração de classe gramatical

Exemplos:

Original – 1) Activation *key* list 2) Service numbers 3) Ringing volume 4) Extras 5) Phone language 6) call divert

Tradução automática – 1) Ativação lista fundamental 2) Conserte numbers 3) Volume tocando 4) Suplementares 5) Telefônico idioma 6) Chame desvie

Tradução do fabricante – 1) Lista teclas ativ. 2) Nos. de serviço 3) Volume do toque musical 4) Extras 5) Idioma do telefone 6) Desvio de chamadas

Comentários – Em alguns casos, aqui se repete o problema de sentidos diferentes de uma mesma palavra no original exigirem traduções diferentes em português, porém com o agravante de as traduções poderem provocar mudança de classe de palavras e, portanto, terem implicações sintáticas. Em outros casos, mesmo sendo usado um único sentido no original, o programa não reconhece a classe em que ele está sendo empregado e gera traduções inadequadas (p.ex.: “multimídia”).

No exemplo 1, observamos que consta no dicionário a tradução “fundamental” para “key”, o que faz com que o sistema gere a tradução “lista fundamental”, sem conseguir estabelecer relação entre o sintagma “key list” e a palavra “activation”. A interpretação de “key” como adjetivo, apesar de possível,

foge do sentido mais freqüente nos domínios de informática e de telefonia, que é “tecla”. Observe ainda que este item apresenta uma característica bastante específica do mercado de localização: a restrição do número de caracteres na tela do aparelho faz com que seja necessário abreviar a tradução, como demonstra a tradução do fabricante “ativ.”. Além da abreviação da palavra “ativação”, a restrição de espaço gera um sintagma *malformado* em português, sem uso de preposição entre as palavras “lista” e “teclas”.

Outro exemplo é o da palavra “multimídia”. O item de interface “Multimedia msgs.” foi traduzido pelo programa como “Msgs de multimídias” e pelo fabricante como “Mens. multimídia”. Provavelmente, a palavra “multimedia” consta no dicionário como substantivo, apesar de ser usada muito mais freqüentemente como adjetivo neste domínio.

Já o exemplo 2 representa um problema bastante freqüente neste *corpus*: muitos sintagmas nominais não foram reconhecidos como tal e foram traduzidos como sintagmas verbais. Esse é um problema especialmente interessante no contexto de itens de interface, especialmente em menus, em que há uma freqüência alta de sintagmas nominais desse tipo.

Este caso exemplifica duas suposições que podemos fazer sobre o sistema a partir da observação do manual traduzido:

- as palavras não-reconhecidas pelo sistema são interpretadas como substantivos, como aconteceu com “numbers”, que não foi reconhecida nesta ocorrência porque havia o algarismo “1” junto a ela, indicando uma nota de rodapé no original;
- se a palavra que antecede ou que sucede um substantivo puder ser um verbo, o sistema emprega esta opção. Como “numbers” não foi reconhecido e foi tomado como substantivo, “service” assumiu a forma verbal “consertar”.

Outros exemplos provocados por palavras que não constavam no dicionário são: “Anykey respondem” (no original, “Anykey answer”), “Dialled numera” (“Dialled numbers”), “Conserte *inbox*” (“Service inbox”).

Já “Chame duração” (“Call duration”) e “Exiba brilho” (“Display brightness”) são exemplos da priorização da forma verbal de palavras que constam no dicionário, mas podem ser empregadas em mais de uma classe

gramatical. Se elas ocorrem com substantivos, parece haver uma priorização pela forma verbal, em vez de o sistema assumir que seja um sintagma nominal do tipo “substantivo + substantivo”.

O exemplo 3 trata das ocorrências das formas com o sufixo “-ing”. Esse sufixo polissêmico em inglês apresenta dificuldades de tradução até mesmo para tradutores humanos, portanto sua formalização para a máquina também não é simples. Mais uma vez, consideramos que o ideal seja analisar as ocorrências deste sufixo em grandes *corpora* deste domínio para observar seu comportamento a fim de criar regras com base nas constatações.

Nas ocorrências da forma “ing” no *corpus*, a maioria dos casos foi interpretada pelo programa como gerúndio, como no exemplo “Volume *tocando*”. Isso parece indicar que haja uma priorização desta forma verbal. No entanto, observamos um predomínio do caráter atributivo nestes sintagmas nominais com a forma “-ing + substantivo”, que normalmente são traduzidas como locução adjetiva (“Volume de toque musical” como tradução do fabricante para “Ringing volume”) ou adjetivo (“alerta vibratório” como tradução do fabricante para “Vibrating alert”) em português.

Já “call waiting” foi traduzido pelo programa da mesma forma que encontramos no manual do fabricante: “Espera de chamada”. Isso provavelmente ocorreu porque “waiting” como substantivo é uma forma dicionarizada (Random House Dictionary, 1996).

O exemplo 4, “Extras”, parece indicar que o dicionário tem algum tipo de prioridade em relação a determinadas regras gramaticais do programa. A tradução “Suplementares” indica a preferência pela forma adjetiva, o que parece pouco usual para uma palavra isolada. É provável que “extra” esteja primeiramente classificado como adjetivo no dicionário do programa, seguindo a classificação dos dicionários tradicionais em português (Dicionário Houaiss, 2001; Dicionário Aurélio; 1999).

No exemplo 5 e observando outras ocorrências no manual, parece haver uma regra que prioriza a tradução de “phone” como adjetivo quando ele coocorre com substantivos. Contudo, o uso mais freqüente não é como adjetivo e sim como substantivo, na locução “de/do telefone” (como “Idioma do telefone”), empregado como sinônimo de “aparelho”.

Consideramos também curioso que o adjetivo “telefônico” tenha sido colocado antes do substantivo “idioma”, o que contradiz a regra geral do português de posposição do adjetivo ao substantivo, aplicada em outras traduções do programa.

No exemplo 6, observamos o emprego, no texto original, de uma palavra numa classe gramatical não-usual, ainda não-dicionarizada em inglês, como “divert”, em “call divert”. Este termo foi traduzido pelo fabricante como “desvio de chamada” e é bastante recorrente em outros textos deste domínio encontrados na internet. Como “divert” está dicionarizado somente como verbo, o programa não consegue resolver o sintagma satisfatoriamente e o traduz como “chame desvie”.

Solução – O não-reconhecimento da classe gramatical costuma ser entendido como um problema de análise sintática (conhecida como *parser* na área de processamento de linguagem natural). No caso deste tradutor automático, os problemas de *parsing* são resolvidos com a elaboração de regras. Como não temos acesso às regras gramaticais do programa, não podemos propor soluções mais abrangentes, mas podemos observar que a priorização de regras do tipo “verbo + substantivo” ou “substantivo + verbo”, por exemplo, gera problemas em relação às palavras que têm a mesma forma em inglês para classes gramaticais diferentes, como mostram os exemplos anteriores.

É possível que uma abordagem mais voltada para o léxico, com a determinação da frequência de determinadas palavras no domínio, fosse útil também nesse caso. A palavra “key” é mais frequentemente traduzida como o substantivo “tecla” do que como o adjetivo “fundamental” neste contexto, assim como “service” é mais recorrente como “serviço” do que como o verbo “consertar”, e “multimedia”, como adjetivo do que como substantivo. Outros exemplos são: “endereço” como tradução de “address”, em vez de “dirigir”, e “recebida” (chamada, ou mensagem, recebida) como tradução de “incoming”, em vez de “entrante”. Dessa forma seria possível priorizar os usos mais frequentes no domínio e diminuir o número de correções na pós-edição.

Outras soluções que podemos propor são:

- um estudo do comportamento do sufixo “-ing” para a elaboração de regras mais precisas e para confirmação da nossa hipótese de que este sufixo desempenha mais frequentemente um papel atributivo quando ocorre em sintagmas nominais antes de um substantivo;
- a criação de uma regra para que a forma adjetiva não seja priorizada caso a palavra também possa ser um substantivo e ocorra isoladamente, como “Extras”. (Essa solução também precisa ser verificada em *corpus*.);
- a alteração da prioridade da forma adjetiva de “phone” no dicionário ou até mesmo sua exclusão, caso um estudo de frequência confirmasse nossa hipótese de que essa forma adjetiva é pouco recorrente;
- a inclusão do termo “call divert” no dicionário com a tradução “desvio de chamada”, como um termo específico do domínio.

Em relação à abreviação decorrente da restrição de caracteres, podemos supor duas soluções: a implantação de um sistema automático que elimine os caracteres excedentes, ou a atribuição da abreviação a tradutores ou revisores humanos. Essa última opção provavelmente garantirá melhores resultados, mais inteligíveis para os usuários, além de dispensar a incorporação de mais um sistema para o corte automático dos caracteres excedentes.

a.2) Não-diferenciação entre participio e pretérito perfeito

Exemplo:

Original – All sent data

Tradução automática – Tudo enviaram dados

Tradução do fabricante – (Não disponível)

Tradução sugerida – Todos os dados enviados

Comentários – A coincidência de forma entre o participio e o pretérito perfeito “sent” confunde o sistema, que opta pela tradução do pretérito perfeito “enviaram”. Podemos inferir que esta opção seja determinada por uma regra gramatical que priorize o pretérito perfeito sempre que o verbo seja precedido por

uma forma nominal que possa ser interpretada como sujeito, como é o caso de “all”.

A esse respeito, podemos observar que a ocorrência do pretérito perfeito do indicativo, por ser uma forma verbal mais típica de narrativa (Koch, 1993), será pouco provável num manual, em que podemos observar a predominância do modo imperativo ou do infinitivo (principalmente nas instruções) e das formas do presente do modo indicativo (nas descrições de recursos e características do produto).

Solução – Uma opção para solucionar o problema seria criar regras mais elaboradas para evitar a interpretação do verbo como pretérito perfeito, a partir da análise das formas nominais que a precedem e das palavras que se seguem ao verbo, como outros substantivos ou advérbios, por exemplo. Uma rápida verificação do emprego da terminação “-ram” (terceira pessoa do plural do pretérito perfeito do indicativo, como *armazenaram, enviaram, receberam* etc) ao longo do manual corrobora a suposição de que a presença de uma forma nominal (e também de formas desconhecidas pelo dicionário, que são sempre interpretadas como formas nominais) determina a opção pelo pretérito perfeito. Contudo, em nenhuma ocorrência essa regra gerou uma tradução correta, sendo sempre a forma do particípio que deveria ter sido empregada.

Outra solução mais simples, se aceita pelo sistema, seria priorizar sempre o particípio. Apesar de exigir uma análise mais detalhada, por poder acarretar outros problemas, é provável que nesse contexto essa solução reduzisse o número de correções necessárias durante a pós-edição.

a.3) Ordem das palavras no sintagma

Exemplos:

Original – 1) Closed user group 2) Add new contact 3) Delete recent call lists 4) Time and date settings

Tradução automática – 1) Grupo de usuário fechado 2) Some contato novo 3) Apague recentes listas de chamada 4) Tempo e colocações de data

Tradução do fabricante – 1) Grupo fechado de usuários 2) (Não disponível) 3) Excluir listas chamadas recentes 4) Configurações de data e hora

Tradução sugerida - 2) Adicionar novo contato

Comentários – Neste tópico, podemos observar a complexidade da tarefa de reconhecimento e interpretação de sintagmas nominais um pouco mais complexos em inglês, bem como a dificuldade para determinar a distribuição mais adequada das palavras na tradução.

No primeiro exemplo, a dificuldade está em determinar que “closed” está relacionado a “group” e não a “user”. O sistema parece seguir uma regra de inversão do sintagma do último para o primeiro substantivo, gerando a tradução “usuário fechado”, que, apesar de inteligível, pode provocar estranhamento para o leitor.

No segundo caso, parece ser aplicada a regra de posposição do adjetivo ao substantivo em português. Contudo, como discutimos em Ribeiro&Dias (2005), esta regra geral não se aplica em todos os casos. Alguns adjetivos têm seu significado alterado dependendo da distribuição no sintagma e do substantivo com que são empregados. Alguns exemplos tradicionais são “um grande homem”, em oposição a “um homem grande”, e “uma boa mulher”, em oposição a “uma mulher boa”.

No caso de “Some contato novo”, apesar de sua tradução não constar no manual fornecido pelo fabricante, a análise de outras ocorrências ao longo do manual, como “novo aplicativo”, “novo código PIN”, “novo conteúdo” e “novo participante” reforça a hipótese de que a tradução mais comum neste caso seria “(Adicionar) novo contato”. A preferência pela anteposição do adjetivo, neste caso, pode ser explicada pelo fato de promover uma leitura mais subjetiva, ao passo que “contato novo” sugerir uma interpretação mais objetiva. “Novo contato” indicaria um contato que ainda não consta na agenda, ao passo que “contato novo” poderia ser interpretado como um contato incluído na agenda recentemente.

Incluimos o terceiro exemplo, “Apague recentes listas de chamada”, porque não conseguimos compreender a forma como o sistema resolveu o sintagma “recent call lists”. Além de não adotar a regra geral de posposição do adjetivo, o sistema parece interpretar “call lists” como um sintagma e só depois relaciona-o ao adjetivo “recent”, quando de fato “recent” está relacionado a “call”.

O último exemplo, “Tempo e colocações de data”, demonstra mais uma vez a dificuldade de reconhecimento dos sintagmas e de determinação das relações estabelecidas pelas palavras que o compõem. Se, nos casos anteriores, “closed” estava relacionado somente a “user” e “recent” a “call”, neste exemplo “settings” está relacionado tanto a “time” quanto a “date”.

Solução – Assim como em muitos aspectos do processamento de linguagem natural, neste caso não é possível determinar uma solução pontual. Martins *et al.* (2004) identificam a questão da “[...] ordem dos itens lexicais na sentença (principalmente no interior dos sintagmas nominais) [...]” como um dos principais desafios para o processamento de linguagem natural e os exemplos acima parecem corroborar essa constatação. Somente com base em estudos mais aprofundados sobre sintagmas será possível tentar inferir padrões de comportamento para a elaboração de regras ou para a inclusão de informações ou termos no dicionário.

A identificação de sintagmas mais frequentes no domínio, como aqueles que envolvam a palavra “call”, por exemplo, poderá ser útil para sua inclusão no dicionário, evitando, assim, que cada palavra seja interpretada separadamente.

No caso do adjetivo “novo”, pode ser útil criar uma regra para sua anteposição.

Para uma discussão mais aprofundada sobre o processamento de adjetivos em português, duas referências são Pria (2005) e Di Felippo (2004).

a.4) Emprego do artigo definido

Exemplos:

Original – Data sent in last session 2) All call’s duration

Tradução automática – 1) Dados enviaram em última sessão 2) A duração de todas as chamadas

Tradução do fabricante – (Não disponível)

Tradução sugerida – 1) Dados enviados na última sessão 2) Duração de todas as chamadas

Comentários – O emprego de artigos representa uma dificuldade não só para a tradução automática, mas também para tradutores humanos. O uso é diferente em inglês e em português, variando em termos de referência (genérica ou específica) e de substantivos contáveis ou não-contáveis. Um exemplo simples é a

estrutura “He’s a translator”, que deve ser traduzida como “Ele é tradutor”, sem o artigo (Frota&Britto, inédito).

No primeiro exemplo citado, observamos a falta do artigo definido “a” em contração com a preposição “em” antes do sintagma “última sessão”. Analisando outras ocorrências no manual, pudemos inferir que o artigo só é incluído quando o artigo definido “the” está presente no original também. Um exemplo é “Sticker in the sales package”, traduzido como “Adesivo no pacote de vendas”. Daí, podemos inferir que há regras para fazer a contração da preposição com o artigo e para identificar a referência do artigo, para determinar o gênero e o número.

Já no segundo exemplo, o artigo “a” deveria ser retirado do sintagma “a duração de todas as chamadas” porque no mercado de localização não se costuma usar artigo definido nas opções de interface. Além disso, observando os outros empregos indevidos de artigo no *corpus*, supomos que a inclusão deste artigo definido esteja relacionada a alguma regra referente ao caso possessivo, indicado pela apóstrofe em inglês, já que esse sinal estava presente nas outras ocorrências também.

Solução – Como comentamos, o emprego de artigos apresenta dificuldade para tradutores humanos também, então é pouco provável que haja uma solução pontual ou simples. O mais recomendável seria realizar uma pesquisa detalhada, num *corpus* expressivo, para tentar observar os padrões de tradução dos artigos para estabelecer regras a partir deles.

No segundo caso, primeiramente é preciso verificar se de fato há uma regra determinando a inclusão do artigo definido com o possessivo. Se houver, a regra deve ser alterada.

a.5) Concordância de gênero e número

Exemplos:

Original – 1) All 2) Missed

Tradução automática – 1) Tudo 2) Perdido

Tradução do fabricante – (Não disponível.)

Tradução sugerida – 1) Todas 2) Não atendidas

Comentários – Esta questão morfológica da não-determinação do gênero e do número de certas palavras em inglês representa um desafio não só para a

tradução automática, mas também para o tradutor humano, e tem implicações sintáticas, por conta da concordância.

Sem avaliação do contexto, na tradução de uma lista de menu isoladamente, não é possível resolver a ambigüidade de “All”, para determinar se a tradução deve ser “Tudo”, “Todos” ou “Todas”, e de “Missed”, para optar entre “Perdido”, “Perdidos”, “Perdida”, “Perdidas” ou “Não atendido”, “Não atendidos”, “Não atendida”, “Não atendidas”, por exemplo.

No caso analisado, pudemos verificar no manual que “All” e “Missed” referem-se ao termo “chamada”, por isso sugerimos as traduções “Todas” e “Não atendidas” (alguns aparelhos disponíveis no mercado usam “Perdidas”).

Esse é um caso de ambigüidade específico, especialmente relevante no mercado de localização. Em muitos produtos, encontramos termos como esses aplicados a várias situações diferentes. Em inglês, a palavra “All” poderia ser utilizada em seguida aos menus “Delete calls?” (“Apagar chamadas?”) ou “Organize appointments” (“Organizar compromissos”), por exemplo, ao passo que em português seria necessário ter a tradução “Todas” no primeiro caso e “Todos” no segundo. Se a internacionalização do produto não for planejada, a utilização desse mesmo item de interface em vários contextos representará um problema na localização do produto em outros idiomas.

Solução – Provavelmente o programa que utilizamos não faz análise do contexto, portanto, a determinação da tradução adequada terá de ser feita por um tradutor humano durante a pós-edição.

b) Sintagmas verbais

b.1) Regência verbal

Exemplos:

Original – 1) Listen to voice messages 2) Send my caller ID

Tradução automática – Escute para expressar mensagens 2) Envie para meu visitante ID

Tradução do fabricante – 1) Ouvir mensagens de voz 2) Enviar minha ID de chamada

Comentários – No caso de “listen to”, inicialmente supomos que fosse um erro de dicionário, que apresentasse apenas “listen”, como a tradução “escutar”, fazendo com que a partícula “to” fosse interpretada independentemente do verbo,

como sendo a preposição “para” isolada. Contudo, a verificação das outras ocorrências de “listen to” no manual mostra que a partícula “to” é omitida da tradução sempre que o verbo é seguido por um substantivo que é seu complemento. Um exemplo é “[...] listen to a sound [...]”, traduzido como “[...] escutar um som [...]”. Supomos, então, que a partícula “to” tenha levado o sistema a interpretar que “voice” fosse um verbo, traduzindo-a como “para”. Esse é um caso especialmente complexo, dado que, de fato, poderíamos ter uma estrutura em que “listen” tivesse uma oração como complemento, como “Listen to choose” (“Ouça para escolher”), num menu de opções, por exemplo.

No caso de “send”, apesar de não haver a partícula “to” no original, podemos supor que haja uma regra que identifique se o complemento do verbo é humano e, em caso afirmativo, considere esse complemento humano como objeto indireto. Uma regra desse tipo evitaria que uma frase como “Send the user a message” fosse interpretada como “Envie o usuário uma mensagem”, por exemplo. Essa suposição parece fazer sentido nesse caso, já que “caller” consta no dicionário como “visitante” mas foi traduzida como “chamada” no exemplo. Outra possibilidade é que a presença da sigla “ID” (de “identification”), provavelmente não reconhecida pelo sistema, tenha afetado a interpretação da oração.

Ainda sobre a partícula “to” vale comentar que aparentemente existe uma regra para que ela seja ignorada antes de verbos no início do período. No entanto, a generalização dessa regra gera problemas numa estrutura bastante recorrente em manuais de instruções, exemplificada por “To make another call, press ‘Options’”. Na tradução automática, a partícula “to” foi retirada, resultando na seguinte tradução: “Fazer outra ligação, aperta Opções”.

Solução – Este caso exigiria a análise do comportamento desses e de outros verbos e partículas similares em inglês e de suas respectivas traduções, a fim de identificar mais características que pudessem contribuir para a elaboração de regras mais precisas.

Inicialmente, seria recomendável revisar as regras já existentes no sistema envolvendo os verbos “listen” e “send” e a partícula “to” e observar a possibilidade de fazer alterações pontuais que minimizassem os problemas identificados aqui.

2.2.1.3.4 Morfologia

No *corpus* analisado, não foram encontrados erros de morfologia.

(Reiteramos que as categorias de "Interoperabilidade" a "Coexistência" (2.2.1.4 a 2.2.6.5), referentes ao software, foram retiradas porque requeriam acesso ao software.)

2.2.7 Custo

A taxonomia não apresenta uma definição do que deve ser considerado nas categorias de avaliação dos custos. Comentaremos alguns aspectos que consideramos relevantes.

Talvez mais drasticamente do que em outras áreas de tradução, o fator custo assume um papel decisivo no mercado de localização, com a freqüente pressão dos clientes finais para a redução do custo do processo como um todo e a conseqüente redução dos preços praticados pelas agências para a remuneração dos profissionais. A diminuição dos prazos também seria interessante neste mercado tão competitivo e com produtos com ciclo de vida tão curto, mas sem dúvida a redução de custo é o fator determinante para o investimento em TA.

2.2.7.1 Custo inicial

O investimento na aquisição do sistema já foi realizado pela empresa proprietária do programa. Consideramos que será necessário um investimento prévio relativamente pequeno nos ajustes básicos de dicionário e regras. Esses ajustes já devem representar uma melhoria significativa, possibilitando uma avaliação mais realista do desempenho do sistema e das adaptações necessárias para a integração da TA ao processo de localização.

Outro investimento necessário nessa fase inicial seria a avaliação técnica da capacidade de integração do sistema com as memórias de tradução e da sua compatibilidade com outros formatos de arquivos além de txt.

2.2.7.2 Custo de manutenção

Em relação ao software e ao hardware, não deve haver um aumento significativo de custo com manutenção, já que ela possivelmente poderá ser realizada pelos técnicos atualmente responsáveis pelos outros equipamentos e softwares da empresa.

2.2.7.3 Outros custos

Outro custo que deve ser considerado são os gastos com pessoal. Seja na preparação dos originais, na utilização do sistema de TA, na pós-edição, na elaboração e na atualização dos dicionários e das regras do sistema ou na editoração final, o investimento no treinamento e na remuneração de profissionais, internos ou terceirizados, será um fator decisivo a ser avaliado e comparado aos investimentos feitos no processo como ele é realizado hoje, sem a utilização de TA.

A categoria Custo finaliza a taxonomia do FEMTI. Na medida do possível, é importante que se faça esse tipo de investimento para acompanhar os avanços tecnológicos e as novas tendências do mercado e as demandas dos clientes. Como afirmam Martins&Nunes (2005:23):

Longe de conduzir ao imobilismo, pois, o reconhecimento das limitações da tradução automática tem alimentado os desenvolvedores, certos de que apenas a experimentação poderá produzir, neste caso, o conhecimento necessário para a implementação (futura) de sistemas bem-sucedidos.

4.2. Casos diversos

Nesta categoria, incluímos outros problemas lingüísticos relevantes que ocorreram no *corpus*, cuja natureza não pudemos determinar nesta análise, impedindo que elas fossem comentadas dentro das categorias da avaliação conduzida na seção anterior. Foram sugeridas soluções sempre que possível.

Caso 1 – sigla GPRS

Comentários - Observando as traduções dos três itens de interface em que constava essa sigla, observamos o seguinte comportamento: “GPRS connection” foi traduzido como “Conexão de GPRS”. A interpretação do sintagma está correta, apesar de a preposição “de” não ser usada neste contexto, como discutido a respeito das siglas. No entanto, os itens de interface “GPRS connection timer” e “GPRS data counter” foram traduzidos como “GPRS conexão cronômetro” e “GPRS dados contador”, respectivamente. Não conseguimos elaborar uma hipótese para o fato de, no primeiro item, o sistema ter sido capaz de reconhecer o

sintagma “GPRS connection” corretamente, mas não reconhecer este mesmo sintagma em “GPRS connection timer”. Da mesma forma, teríamos suposto que o sintagma “GPRS data” fosse reconhecido pelo sistema, o que não se confirmou. Como o sistema foi capaz de reconhecer e tratar adequadamente outros sintagmas com três substantivos, mesmo nos casos em que um dos substantivos não constava no dicionário, supomos que a presença da sigla tenha confundido o sistema.

Solução – Novamente, uma solução pontual seria incluir os sintagmas completos no dicionário, como “GPRS connection” e “GPRS data”.

Caso 2 – Outbox

Comentários – Na lista de menus, há apenas uma ocorrência deste termo, mas verificamos que nas outras duas ocorrências dele no manual foi utilizada a mesma tradução: “Exceda em boxe”. Não foi possível concluir por que o sistema interpretou esta palavra como um sintagma verbal, já que as outras palavras desconhecidas foram mantidas em inglês e interpretadas como substantivos. Uma hipótese é que esta tradução conste no dicionário, mas como a tradução não faz sentido em português, achamos pouco provável que ela tenha vindo de algum dicionário ou tenha sido incluída por um tradutor humano.

Solução – Inclusão ou correção da tradução para “Caixa de saída” no dicionário.

Caso 3 – Phone settings

Comentários – Repetimos aqui este item já comentado na discussão sobre classes de palavras porque a tradução “Telefônicas colocações” representa um problema que observamos, mas não conseguimos identificar a causa. Apesar de ter sistematicamente posposto o adjetivo ao substantivo na tradução, neste exemplo o adjetivo fica anteposto, gerando uma estrutura incomum em português ainda que tivesse sido usada a tradução “configurações” para “settings”, mais freqüente neste domínio.

Solução – Seria preciso verificar a regra de colocação do adjetivo. A interpretação de “phone” como adjetivo é muito menos freqüente neste domínio e, portanto, não deveria ser priorizada no dicionário.

Caso 4 – Message alert tone

Comentários – Como o sistema foi capaz de interpretar outros sintagmas nominais com três ou mais substantivos, não entendemos por que neste caso ele

gerou a tradução “Mensagem tom alerta”. Observando outras traduções de “alert”, vimos que a palavra foi interpretada como adjetivo sempre que havia outro substantivo junto a ela, como é o caso de “tone”. Mesmo assim, a regra geral que pudemos inferir geraria a tradução “tom alerta de mensagem” ou “mensagem de tom alerta”, conseguindo estabelecer alguma relação entre “tone alert” e “message”, o que não ocorreu neste caso. A única hipótese que conseguimos imaginar é que o sistema tenha interpretado “alert tone” como o *nome* da mensagem, como se fosse um título. Para tal, seria necessário que ele tivesse uma regra com esta instrução, aplicável a determinados tipos de palavras que pudessem ser seguidas de seus *nomes*, o que parecer pouco provável.

Solução – Alteração no dicionário da tradução de “alerta”, priorizando a forma do substantivo em vez do adjetivo.

Caso haja de fato uma regra para que as palavras que se seguem a “Mensagem” ou determinados substantivos sejam interpretadas como *títulos* ou *nomes*, seria necessário revisar esta regra.

Caso 5 – As palavras “call”, “service”, “display” e “clear”

Comentários – Estas palavras poderiam ter sido incluídas nos tópicos de análise sobre alteração de classe gramatical ou de terminologia do domínio. No entanto, na comparação das traduções atribuídas a elas, observamos alguns problemas para os quais não foi possível supor uma causa. Como esses são termos recorrentes no contexto de telefonia celular, consideramos especialmente relevante discuti-los.

O termo “service” foi traduzido como “serviço” sempre que foi interpretado como substantivo e como “consertar” sempre que foi interpretado como verbo. Já o termo “clear” foi traduzido como “claro”, quando interpretado como adjetivo, e como “clarear”, quando interpretado como verbo.

No caso de “service”, verificamos que nesse *corpus* nenhuma ocorrência era de fato a forma verbal. Contudo, ela sempre foi empregada quando “service” precedia ou sucedia algum substantivo, o que nos faz supor que o sistema não considere a possibilidade de a estrutura ser um sintagma nominal da forma N+N (substantivo com substantivo, como “Service numbers” ou “Info service”), caso alguma das duas palavras também possa ser um verbo. Contudo, essa

possibilidade é a mais freqüente, especialmente no caso dos itens de interface analisados, em que predominam sintagmas nominais.

Apesar do predomínio dos sintagmas nominais na lista de menus, não é possível criar uma regra geral para priorizá-los, já que muitos termos que poderiam ser substantivos ou adjetivos aqui são empregados como verbos, como é o caso de “clear”. O termo foi usado no original como verbo nas três ocorrências, mas foi interpretado como adjetivo pelo sistema em duas delas, “Clear counters”, traduzido como “Contadores claros”, e “Clear timers”, traduzido como “Cronômetros claros”. A ocorrência “Clear the cache” foi traduzida como “Clareie o esconderijo”. Podemos supor que a presença do artigo “the” tenha induzido o programa a reconhecer o sintagma verbal. Já nas outras duas ocorrências, uma hipótese é que a hierarquia de regras do sistema priorize a interpretação como adjetivo em vez de verbo quando ambos forem possíveis. Vale comentar ainda o problema das escolhas lexicais, que nesse contexto deveriam ser “zerar” ou “limpar”, para “clear”, e “cache” para “cache”, em vez de “esconderijo”.

Já os casos de “display” e “call” parecem representar um desafio ainda mais complexo, pela variedade de estruturas em que essas palavras ocorrem. Se nos casos anteriores conseguimos identificar que, pelo menos nesse *corpus*, “service” ocorre somente como substantivo e “clear” somente como verbo, nos casos de “display” e “call” é fácil pensar em várias formas possíveis e muito freqüentes nesse domínio (mesmo que elas não tenham ocorrido no *corpus*), o que torna a formalização mais difícil. Temos sintagmas nominais, como “Call duration”, “Call register” e “Display brightness”, traduzidos pelo fabricante como “Duração da chamada”, “Registro de chamada” e “Brilho da tela”, respectivamente, mas interpretados pelo sistema como sintagmas verbais, traduzidos como “Chame duração”, “Chame registro” e “Exiba brilho”. Contudo, como vimos anteriormente, não é possível priorizar os sintagmas nominais, já que “call” e “display” são freqüentemente empregados em sintagmas verbais nesse contexto, como “Call” (“Chamar”), “Call number” (“Chamar número”), “Display” (“Exibir”) e “Display contact” (“Exibir número”), por exemplo.

Outro caso curioso envolvendo a palavra “call” foi “Last call duration” que foi traduzido pelo programa como “Por último chame duração”. Como “last” foi

interpretado como adjetivo em outras ocorrências e “call” como substantivo, não compreendemos por que o sintagma nominal “last call” não foi reconhecido neste caso. É possível que a ocorrência de “last” junto com uma palavra que possa ser verbo induza a sua interpretação como advérbio.

Solução – Nesses casos não foi possível sugerir soluções específicas. Consideramos necessário que esses termos sejam analisados em *corpora* maiores e contextos variados, a fim de que suas características sejam observadas mais detalhadamente para que seja possível fazer as alterações pertinentes nas regras ou no dicionário do sistema.

4.3. Acertos do programa

Dos 144 itens de interface da lista de menu, 29 foram traduzidos corretamente pelo programa, com a mesma tradução encontrada no manual do fabricante. Desses 29, 18 eram palavras isoladas, o que indica que elas constam no dicionário. As 11 restantes exigiam algum nível de análise sintática, que deve ter sido resolvido pelas regras gramaticais.

Em quatro casos havia total correspondência entre a estrutura sintática em inglês e em português, portanto não houve necessidade de uma análise sintática mais elaborada. Desses, dois envolviam a estrutura “substantivo 1 + preposição + substantivo 2” como “Memory in use” (“Memória em uso”), um continha um pronome adjetivo antes do substantivo 2 (“Duration of all sessions”, cuja tradução sugerida é “Duração de todas as sessões”) e o último tinha a estrutura “pronome + substantivo” (“My numbers”, “Meus números”).

Nos outros sete casos, as alterações sintáticas eram simples e as estruturas envolviam sintagmas nominais com apenas duas palavras. Cinco tinham a estrutura “substantivo 1 + substantivo 2”, como “Security level”, resolvidos satisfatoriamente com uma regra de inversão do tipo “substantivo 2 ‘de’ substantivo 1” (“Nível de segurança”). Um apresentava a estrutura “adjetivo + substantivo”, resolvida também com a inversão “substantivo + adjetivo” (“Personal shortcuts”, “Atalhos pessoais”) e o último tinha a estrutura “particípio + substantivo” resolvida também com a inversão “substantivo + particípio” (“Received calls”, “Chamadas recebidas”).

Com esses casos, foi possível inferir algumas regras básicas que o programa deve possuir, com algum nível de refinamento, como concordância de gênero e número. Contudo, como vimos na análise dos problemas, essas regras nem sempre são suficientes para resolver sintagmas mais complexos. Mais do que isso, elas parecem corroborar o que Martins *et al.* (2004) afirmam sobre outros tradutores automáticos por eles avaliados: “esta aparente eficiência esconde a verdadeira razão do sucesso: a proximidade entre as construções do inglês do português” (p. 42). Como vimos, quanto menos parecidas as estruturas das duas línguas, menos eficiente é uma lógica de tradução mais próxima da tradução direta, praticamente palavra por palavra.

4.4. Comentários finais

Neste capítulo, aplicamos parte da taxonomia do FEMTI para fazer uma avaliação inicial do tradutor automático utilizado na pesquisa.

A respeito dos aspectos contextuais, observamos a relevância das características do texto de entrada e do usuário (os profissionais que já atuam no mercado de localização), que terão de se adaptar à TA como uma nova ferramenta de trabalho.

Já para a análise lingüística, apresentamos exemplos em português retirados da tradução da lista de menus do aparelho apresentada no manual. Observamos que muitos dos problemas discutidos nas tipologias do capítulo 3, como a homonímia, ocorrem nesse *corpus* também e que muitas categorias do FEMTI são semelhantes às aquelas utilizadas no controle de qualidade no mercado de localização de software (apresentadas no capítulo 2), como precisão, uniformidade e terminologia. A diferença nesse último caso está no enfoque e no peso que será dado a cada categoria, tendo em vista que no mercado de localização a avaliação da qualidade é voltada para tradutores humanos e aqui, para o software de tradução automática.

Consideramos que os aspectos discutidos com base no FEMTI e nas subcategorias propostas são relevantes para um estudo inicial das implicações do uso do tradutor automático no mercado de localização. O aprofundamento das questões aqui propostas e a avaliação das categorias que não analisamos

(referentes aos aspectos técnicos do software) poderão ser realizados pela empresa ou em futuras pesquisas acadêmicas, estreitando a parceria entre o mercado de trabalho e a universidade.

Em termos contextuais, podemos concluir que a viabilidade de uso desse tradutor automático no mercado de localização depende de uma análise interna da empresa de como inseri-lo no processo de trabalho, considerando-se especialmente sua relação com as ferramentas de memória de tradução e com os diferentes formatos de arquivos (além de txt).

Em relação à avaliação lingüística, observamos que o sistema ainda apresenta uma qualidade baixa, mas pode ser aprimorado com algum investimento na elaboração ou na atualização de suas regras gramaticais e seus de dicionários. Só será possível fazer uma avaliação de aspectos como legibilidade, inteligibilidade, coerência e coesão depois que alguns ajustes básicos forem feitos para minimizar os problemas no nível sentencial.

Alguns aspectos específicos da localização que precisarão estar refletidos nas regras e nos dicionários são a terminologia (nesse caso, dos domínios de informática e telefonia celular), o uso freqüente de siglas e abreviações, a identificação das opções de interface (para que sejam traduzidas sempre da mesma forma) e as especificidades lingüísticas das opções de interface, como a concisão e o uso do modo infinitivo para os sintagmas verbais. Algumas características do gênero manual, como o modo verbal empregado para as instruções, também poderão ser identificadas e incluídas nas regras.

Na nossa análise, o dicionário desempenhou um papel central em relação à qualidade lingüística do sistema. Como os menus contêm a terminologia principal do telefone celular, grande parte dos acertos e dos erros dependeu do fato de os termos constarem ou não no dicionário. Mesmo para algumas questões sintáticas foi possível sugerir soluções baseadas no dicionário.

Em muitos casos, as soluções sugeridas baseavam-se em análises de freqüência em *corpus*. Com isso, observamos que um recurso empírico, em teoria divergente de um sistema baseado em tradução indireta por transferência, na prática pode contribuir para a busca de resultados eficientes. Soluções baseadas em freqüência em *corpus* podem ser aplicadas para a observação de muitos

padrões lexicais e sintáticos, contribuindo, assim, para a atualização dos dicionários e das regras gramaticais do sistema.

Um exemplo de recurso que pode ser utilizado para a análise lingüística é o Corpógrafo, desenvolvido pela Linguateca (Maia&Sarmiento, 2005). Com essa ferramenta gratuita é possível, por exemplo, gerar listas de *n-gramas* (palavras que coocorrem com frequência) e de candidatos a termo a partir de *corpora*, a fim de incluí-los e priorizá-los no dicionário. Quanto maior for o *corpus* analisado, mais representativa a lista será. Uma opção para a composição do *corpus* na área de localização são as memórias de tradução que as empresas já possuem, que são grandes *corpora* paralelos (textos originais alinhados às respectivas traduções). A relevância do dicionário no nosso estudo mostra que os sistemas podem ser significativamente aprimorados com a devida personalização dos próprios recursos disponíveis para que eles reflitam as características do domínio (ou domínios) de interesse e as necessidades dos usuários e dos consumidores da tradução.