

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS

Fausto Magalhães da Silveira

TERMINOLOGIA E TRADUÇÃO NA LOCALIZAÇÃO DE SOFTWARE:

Insumos para o Processamento da Linguagem Natural

Porto Alegre

2013

Fausto Magalhães da Silveira

TERMINOLOGIA E TRADUÇÃO NA LOCALIZAÇÃO DE SOFTWARE:

Insumos para o Processamento da Linguagem Natural

Dissertação de Mestrado em Teorias Linguísticas do Léxico, apresentada como requisito para obtenção do título de Mestre pelo Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul.

Orientadora: Profa. Dra. Cleci Regina Bevilacqua

Porto Alegre
2013

Silveira, Fausto Magalhães da
Terminologia e Tradução na Localização de Software:
Insumos para o Processamento da Linguagem Natural /
Fausto Magalhães da Silveira. -- 2013.
264 f.

Orientadora: Cleci Regina Bevilacqua.

Dissertação (Mestrado) -- Universidade Federal do
Rio Grande do Sul, Instituto de Letras, Programa de
Pós-Graduação em Letras, Porto Alegre, BR-RS, 2013.

1. Localização. 2. Terminologia. 3. Tradução. 4.
Processamento da Linguagem Natural. I. Regina
Bevilacqua, Cleci, orient. II. Título.

AGRADECIMENTOS

Não dá para captar a energia exigida pelo empreendimento que é um trabalho de conclusão acadêmica sem envolver um bocado de pessoas. Esse tipo de projeto demanda especialmente a boa vontade, a parceria, o conhecimento e – talvez acima de todos os outros aspectos – a paciência de professores, empregadores, colegas de trabalho, amigos, família. E talvez de um ou dois desconhecidos pelo caminho.

É por isso que aqui vão meus agradecimentos a todos aqueles que de alguma forma, direta ou indiretamente, tiveram parte na realização deste trabalho.

Então, meu muito brigadão:

Aos professores do ppgletras, pelas aulas, discussões e avaliações ao longo desse mestrado em Teorias Linguísticas do Léxico na UFRGS.

À Profa. Cleci Bevilacqua – e um tremendo de um agradecimento – pela orientação no mestrado, e antes disso, por ter sido receptiva e bacana nos primeiros contatos rumo ao início dessa etapa. Também pela paciência, pela trabalhadeira com a revisão dessa dissertação, pela solidariedade, por tudo mais. Muito brigadão ☺

Ao horário de trabalho flexível nas empresas, em especial no primeiro ano do curso, que me permitiu assistir às aulas à tarde.

Aos colegas, pelo coleguismo, no trabalho e na pós.

Ao Zílio, por ter me pago um almoço no RU – o primeiro e último (até a publicação deste trabalho) depois de cinco anos afastado do Campus do Vale.

Aos amigos, pela amizade e por não terem virado ex-amigos especialmente depois desse ano de vai-ou-racha na conclusão deste trabalho.

À família, pelo parentesco e por motivos meio que parecidos aos acima, exceto por não ser possível virar ex nesse caso.

Às pessoas fundamentais: à minha mãe, Zulmira; meu pai, Luiz Carlos; meus irmãos Dodo, Nana e Dodinho; e à mãe de minha esposa, Maria de Lourdes.

À pessoa mais fundamental das pessoas fundamentais, Luciane Gonçalves Sarti, amor da minha vida – essa sim por não ter virado ex também☺. Complicado escrever agradecimentos a alguém que segurou tanto a barra durante esses dois anos e uns quebrados. Não é pouco possível que tenha me contagiado com uma fração – bem ínfima – de sua força de vontade, dedicação, inteligência, garra etc. etc. Não exagero em dizer que tu foi a diferença entre estar na estaca zero deste trabalho – e talvez até da vida em geral – e ter concluído tudo aquilo que está escrito e não está escrito a respeito deste trabalho. Esta dissertação é pra ti.

*computador computa
a tradição
da traição
da tradução*

Millôr Fernandes

RESUMO

Este trabalho centra-se no processo de QA (na sigla em inglês para *quality assurance* – ou *garantia da qualidade* em português) que é feito no setor da localização, visando a melhorar o trabalho do tradutor. Localização consiste em um processo e um campo de atuação profissional que visam a adaptar produtos (geralmente de *software*) segundo o idioma e as convenções culturais de determinada localidade com o objetivo facilitar a entrada de um produto ou serviço em um país ou mercado. Com relação ao QA, uma de suas etapas consiste na validação da terminologia de um projeto de tradução. O QA terminológico envolve o uso de um *software* que verifica se a terminologia aplicável é usada na tradução. As ocorrências que o *software* considera incorretas são salvas em uma lista de validação terminológica, que é conferida normalmente por um tradutor ou editor. Itens que o tradutor considerar incorretos são corrigidos na tradução; os demais são descartados. Por ignorar aspectos linguísticos, o *software* gera muito ruído, ou falsos positivos, resultando em listas extensas, que não compensam o tempo dedicado a sua revisão. A fim de prover insumos para solucionar o problema, este trabalho emprega uma abordagem comunicativa, cognitiva e funcional à terminologia e à tradução para analisar uma lista de validação terminológica, em um projeto de localização real, no par de idiomas inglês dos Estados Unidos e português do Brasil. Para tal fim, foi gerada uma lista de validação por meio de um *software* de QA usado na área da localização. Ocorrências dessa lista foram analisadas e classificadas segundo critérios de base fraseológica, variacional e tradutória, além de morfológica e discursiva. O objetivo é oferecer subsídios que norteiem o desenvolvimento de aplicações computacionais linguisticamente motivadas que reduzam a incidência de ruído nestas listas. Os resultados mostram que a maior parte do ruído decorre de fatores linguísticos gerais, como morfológicos e discursivos, indicando também que 1/3 destes coocorrem com fenômenos fraseológicos, variacionais e tradutórios.

Palavras-chave: Localização. Terminologia. Fraseologia da Língua de Especialidade. Variação da Língua de Especialidade. Tradução Especializada. Técnicas de Tradução. Erros de Tradução. Processamento da Linguagem Natural.

SUMMARY

This paper focuses on the process of Quality Assurance (QA) that is undertaken by the Localization industry, aiming at improving the work of translators. Localization consists of a process and a professional field whose purpose is to adapt goods or services (usually software-related) according to the language and cultural conventions of a particular locale in order to facilitate market penetration in a given country or market. One of the QA stages consists of validating the terminology on a translation project. The QA for terminology makes use of software to check if the applicable terminology is used in translation. Occurrences that the software identifies as incorrect are saved in a list for terminology validation. The list is usually reviewed by a translator or an editor. The items considered incorrect by the translator are corrected in the translation, and the remaining entries are discarded. Because the software does not take language aspects into account, a good deal of noise is generated, resulting in large lists that are not cost-effective or time-efficient to review. With the purpose of providing input to solve the problem, this work employs a communicative, cognitive and functional approach to terminology and translation for the analysis of a terminology validation list in U.S. English and Brazilian Portuguese, on a genuine localization project. To complete this task, a list for validation was generated via a well-known QA software product used in the Localization field. Occurrences from the generated list were analyzed and categorized according to phraseological, variational and translational criteria in addition to morphological and discursive criteria. The objective is providing input to drive the development of linguistically motivated computer applications that may reduce the incidence of noise on the lists. Results show that most of the noise is due to general linguistic factors, such as morphological and discourse aspects, also suggesting that 1/3 of that noise occurs simultaneously with phraseological, variational and translational phenomena.

Keywords: Localization. Terminology. Phraseology of Special Language. Variation of Special Language. Specialized Translation. Translation Techniques. Translation Errors. Natural Language Processing.

LISTA DE ILUSTRAÇÕES

Figura 1 - Processos do GILT, sua interdependência e caminho para sua compreensão	25
Figura 2 - Componentes de um típico projeto de localização	28
Figura 3 - Cadeia de produção de <i>software</i>	30
Figura 4 - Aplicação típica da ISO 9001:2000 no setor da localização	31
Figura 5 - Lista gerada por <i>software</i> de QA	33
Figura 6 - Página de produto, versão em inglês	65
Figura 7 - Página de produto, versão em português	65
Figura 8 - Uso de segmento de tradução em contexto diferente, em inglês	66
Figura 9 - Uso de segmento de tradução em contexto diferente, em português	67
Figura 10 - Reaproveitamento de parte de segmento de tradução	68
Figura 11 - Página do Microsoft Office na Internet em diversos idiomas	99
Figura 12 - Consulta à terminologia da Symantec no <i>site Norton Together</i>	119
Figura 13 - Consulta a support no <i>site Norton Together</i>	120
Figura 14 - Tela do Norton AntiVirus 2011	121
Figura 15 - Ajuda do Norton AntiVirus 2011	124
Figura 16 - SDL Passolo 2011 exibindo strings do Norton AntiVirus 2011	128
Figura 17 - SDL Trados WinAlign 2007 alinhando arquivos da ajuda do Symantec Norton AntiVirus 2011	129
Figura 18 - Lista de validação terminológica gerada pelo <i>software</i> ApSIC Xbench	131
Figura 19 - Exerto da lista de validação de QA gerada e exportada via ApSIC Xbench em formato .xls	137
Figura 20 - Exerto da lista de validação adaptada para incluir todos os itens processados pelo <i>software</i> de QA	139
Figura 21 - Tabela de registro dos dados para análise	142
Figura 22 - Diagrama processual da análise feita neste trabalho	147
Figura 23 - Tabela de registro dos dados após a análise	150
Figura 24 - Referência do <i>software</i> Norton AntiVirus	191
Figura 25 - Contexto para segmento de tradução envolvendo risk / risco	193

LISTA DE QUADROS

Quadro 1 - Equacionamento metodológico global do PLN	20
Quadro 2 - Descrição dos principais componentes de um <i>software</i> típico.....	29
Quadro 3 - Variantes por <i>subcorpus</i>	198
Quadro 4 - Variantes com estrutura idêntica	205
Quadro 5 - Variantes com alteração de número	206
Quadro 6 - Variantes com alteração de preposição	207
Quadro 7 - Variantes com alteração de estrutura	208
Quadro 8 - Variantes com redução de extensão	210
Quadro 9 - Variantes monolexicais	211
Quadro 10 - Variantes polilexicais	213
Quadro 11 - Variantes que envolvem diversas alterações complexas.....	215

LISTA DE TABELAS

Tabela 1 - Condições que as unidades terminológicas devem satisfazer para cada componente.....	39
Tabela 2 - Denominações para unidades fraseológicas segundo levantamento de Bevilacqua (1999).....	45
Tabela 3 - Estruturas fraseológicas superficiais	48
Tabela 4 - Aplicabilidade das características fraseológicas	56
Tabela 5 - Relação entre causas e tipo de variação	73
Tabela 6 - Proposta de classificação formal	75
Tabela 7 - Causas para variação	82
Tabela 8 - Definições para tradução	88
Tabela 9 - Oposições para equivalência segundo alguns teóricos da tradução	93
Tabela 10 - Noções de análise tradutória e sua probabilidade de ocorrência neste trabalho ..	105
Tabela 11 - Técnicas de tradução e sua probabilidade de ocorrência neste trabalho	110
Tabela 12 - Critérios para a seleção do <i>corpus</i>	116
Tabela 13 - Elementos textuais típicos em interfaces gráficas.....	122
Tabela 14 - Sumário com informações sobre cada <i>subcorpus</i>	126
Tabela 15 - Volume de palavras por <i>subcorpus</i>	127
Tabela 16 - Relação de entradas simples e compostas no glossário.....	127
Tabela 17 - Ferramentas eletrônicas usadas na coleta e análise de dados.....	132
Tabela 18 - Positivos e falsos positivos para ocorrências que o <i>software</i> inclui na lista de QA.....	153
Tabela 19 - Positivos e falsos positivos para ocorrências que o <i>software</i> exclui da lista de QA.....	153
Tabela 20 - Ocorrências que o <i>software</i> inclui na lista de QA por <i>subcorpus</i>	155
Tabela 21 - Ocorrências que o <i>software</i> exclui da lista de QA por <i>subcorpus</i>	155
Tabela 22 - Ocorrências que o <i>software</i> inclui na lista de QA para o <i>subcorpus</i> de ajuda	156
Tabela 23 - Ocorrências que o <i>software</i> exclui da lista de QA para o <i>subcorpus</i> de ajuda....	156
Tabela 24 - Ocorrências que o <i>software</i> inclui na lista de QA para o <i>subcorpus</i> de UI.....	156
Tabela 25 - Ocorrências que o <i>software</i> exclui da lista de QA para o <i>subcorpus</i> de UI.....	156
Tabela 26 - Ocorrências de falsos positivos por classificação para os itens que o <i>software</i> inclui na lista de QA	157
Tabela 27 - Resultados gerais para UFE	162

Tabela 28 – Relação entre falsos positivos e classificação sintagmática por <i>subcorpus</i> para UFE.....	162
Tabela 29 - Ocorrências para a UFE colocar em quarentena na UI	163
Tabela 30 - Total de incidência de falsos positivos para UFE ao considerar fatores relacionáveis	164
Tabela 31 - Identificação de característica causadora de ruído e representação formal para as UFE do <i>corpus</i>	166
Tabela 32 - Unidades potencialmente fraseológicas	171
Tabela 33 - Erros de tradução e ruído tradutório.....	178
Tabela 34 - Técnicas tradutórias envolvidas em ocorrências do QA	179
Tabela 35 - Técnicas causadoras de ruído por <i>subcorpus</i>	180
Tabela 36 – Positivos e falsos positivos para instant messenger / <i>software</i> de mensagens instantâneas.....	181
Tabela 37 - Ocorrências para erros de tradução	189
Tabela 38 - Grafias de on-line ao longo do tempo no <i>site</i> da Symantec	195
Tabela 39 - Ruído decorrente de variação por <i>subcorpus</i>	197
Tabela 40 - Ocorrências variacionais no <i>corpus</i> segundo a tipologia de Freixa.....	199
Tabela 41 - Ocorrências por subtipo de variação	202
Tabela 42 - Resultados para ocorrências linguísticas por <i>subcorpus</i>	217
Tabela 43 - Classificação para as ocorrências consideradas erro pelo <i>software</i> de QA.....	218

LISTA DE ABREVIATURAS E SIGLAS

CAT	<i>Computer-Assisted Translation</i>
EN-US	inglês dos Estados Unidos
FLE	Fraseologia da Língua de Especialidade
GALA	<i>Globalization and Localization Association</i>
GILT	<i>Globalization, Internationalization, Localization, Translation</i>
LISA	<i>Localization Industry Standards Association</i>
MLV	<i>Multi-Language Vendor</i>
MT	Modelo Transformacional
NE	Núcleo Eventivo
NT	Núcleo Terminológico
PLN	Processamento da Linguagem Natural
PT-BR	português do Brasil
QA	<i>Quality Assurance</i>
AS	Sintagma Adjetival
SLV	Single-Language Vendor
SN	Sintagma Nominal
SP	Sintagma Preposicional
SV	Sintagma Verbal
TCT	Teoria Comunicativa da Terminologia
TEP	<i>Translation, Editing, Proof-reading</i>
TGT	Teoria Geral da Terminologia
UC	Unidade de Conhecimento
UCE	Unidade de Conhecimento Especializado
UF	Unidade Fraseológica
UFE	Unidade Fraseológica Especializada
UI	<i>User Interface</i>
UNICOME	Unidade de Comunicação Especializada
USE	Unidade de Significação Especializada
UT	Unidade Terminológica

SUMÁRIO

1	Introdução.....	17
2	Localização.....	22
2.1	O que é localização.....	22
2.1.1	GILT e outros termos comuns à área.....	24
2.1.2	Projetos de localização.....	26
3	Revisão da Literatura.....	34
3.1	Terminologia.....	34
3.1.1	Teoria Geral da Terminologia.....	35
3.1.2	Teoria Comunicativa da Terminologia.....	37
3.1.3	Síntese.....	42
3.2	Fraseologia.....	43
3.2.1	Denominações.....	44
3.2.2	Definições.....	46
3.2.3	Critérios de caracterização.....	48
3.2.4	Síntese.....	54
3.3	Variação.....	57
3.3.1	Causas da variação denominativa.....	60
3.3.2	Pressupostos para análise: autovariação e heterovariação.....	73
3.3.3	Tipos de variação denominativa.....	74
3.3.4	Síntese.....	82
3.4	Tradução.....	85
3.4.1	Concepções de tradução.....	87
3.4.2	Paradigmas da tradução.....	91
3.4.3	Noções centrais de análise tradutória.....	104
4	Procedimentos metodológicos.....	115
4.1	Materiais e ferramentas.....	115
4.1.1	<i>Corpus</i>	115
4.1.2	Ferramentas.....	127
4.2	Metodologia.....	132
4.2.1	Coleta de dados para o <i>corpus</i>	133
4.2.2	Execução do QA terminológico.....	135
4.2.3	Preparação e organização dos dados da lista de QA.....	140

4.2.4	Análise da lista de validação terminológica	144
4.2.5	Quantificação e filtragem dos resultados.....	151
5	Análise dos dados e resultados	152
5.1	Análise global dos resultados	152
5.1.1	Acertos e erros do <i>software</i>	153
5.2	Alinhamento.....	158
5.3	Glossário	159
5.4	Análise fraseológica.....	160
5.4.1	Fraseologia em números	161
5.4.2	Sumário das ocorrências em relação às características das UFE.....	165
5.4.3	Outros resultados da análise	171
5.4.4	Futuros caminhos para essa classificação.....	175
5.5	Análise tradutória.....	177
5.5.1	Tradução em números	177
5.5.2	Classificação tradutória e sua influência no ruído	178
5.6	Análise variacional	196
5.6.1	Variação em números	196
5.6.2	Ocorrências de variação e sua influência no ruído	203
5.7	Análise linguística.....	216
5.7.1	Análise linguística em números.....	217
5.7.2	Ocorrências linguísticas.....	217
6	Conclusões.....	238
6.1	Considerações finais	248
7	Referências	253

1 INTRODUÇÃO

“Localizar”. De uns 20, 30 anos para cá, essa forma lexical passou a designar algo um tanto quanto diferente das acepções mais comuns – ou pelo menos daquelas que se encontra em dicionários gerais, como Houaiss e Aurélio. A expansão do mercado da informática, e, por extensão, da tradução de produtos computacionais, trouxe à tona uma série de problemas de tradução que não envolvia exclusivamente o texto das UI (na sigla em inglês para *user interface*, ou interface de usuário em português), mas também convenções específicas a um país, como diferentes sistemas de medida, e direção do texto na interface, como no caso do português em comparação com árabe, para citar alguns exemplos. A partir daí, a indústria de *software*, como qualquer atuante em setores afins faria, passou a desenvolver e aperfeiçoar processos, culminando na criação de entidades que consolidam a localização como um conjunto de práticas industriais, a exemplo da associação GALA.

Assim, está montado o cenário que consolida a nova noção. A indústria cria e regula processos e depois se mune de mecanismos que lhes dão legitimidade. Localização começa a designar, então, uma atividade industrial, a qual consiste basicamente em traduzir *software* adaptando-o a idiomas ou mercados diferentes. Para interligar tudo, surge a Internet que modifica totalmente às relações de trabalho e a forma como as pessoas consomem textos e *software* no meio digital.

Com isso, hoje, as práticas industriais e metodologias desenvolvidas para localização recebem adesão cada vez maior fora desse âmbito. No que se refere ao empresariado, veículos de mídia e imprensa internacionais são exemplos disso. Em relação ao prestador de serviços, o tradutor *freelancer* que atua na área normalmente recebe divisas de fora, em dólar americano e euro – aqui no Brasil mesmo. Só que, ao mesmo tempo, esse profissional é deixado de lado em boa parte do processo tradutório como um todo. Como os funcionalistas sugeriam, o tradutor participa de projetos; o detalhe é que, na localização, as decisões ficam praticamente todas para outras pessoas tomarem; pessoas “de dentro”, funcionários de empresas ou agências atuantes no setor. Na opinião de alguns, ao tradutor resta o trabalho de converter texto entre idiomas; nada mais. Mas talvez as coisas não sejam bem assim, como será visto em *Localização* na página 96).

Nesse cenário, surge um campo totalmente novo a ser explorado, profissionalmente, é evidente, mas também academicamente. Alguns autores já deram seus passos, como Dunne (2006, p. 1-11). Outros deram muitos passos, como Pym (2003, 2004 e 2010), que suscita

muitas perguntas a respeito das implicações da localização para aspectos como ética e tradução.

Dado o contexto anterior, este trabalho se insere no rol de produções acadêmicas sobre o tema, com o modestíssimo objetivo de:

- contribuir para a discussão do tema nos Estudos Linguísticos no Brasil, em especial no que diz respeito a estudos de Terminologia;
- fazer um estudo centrado no tradutor;
- analisar um aspecto bem pontual, e totalmente aplicado, da prática da localização.

Com relação ao primeiro ponto, de fato, não há muitos trabalhos voltados ao tema em âmbito nacional. Alguns autores publicaram textos sobre o assunto no Brasil, como o próprio Pym (2008), que tem artigo publicado em português sobre competência tradutória e uso de ferramentas eletrônicas (artigo que, diga-se, tangencia a localização, mas não a aborda diretamente). Alguns centram mais na tecnologia usada em localização, embora não exclusiva a ela. Com relação à Terminologia, em um âmbito internacional, trabalhos como os de Pym não chegam a entrar em discussões aprofundadas sobre o tema. Agora, e já tocando o segundo ponto, até onde se pôde observar, não há muitos trabalhos que discutem o assunto na perspectiva “de quem está dentro”, diga-se, voltada ao tradutor e que analise práticas bem particulares, específicas a atividades dentro da localização.

Por fim, sustentando a ideia de que a pretensão é modesta, o trabalho analisa a extremidade mais inferior possível em todo o processo. Trata-se da atividade de QA terminológico, envolvendo o par de idiomas português brasileiro (PT-BR¹) e inglês americano (EN-US). O capítulo *TEP QA em detalhes* (p. 31) traz mais informações sobre o tema, mas para sintetizar, o QA terminológico é uma atividade pontual dentro do QA. QA é sigla para o inglês *quality assurance* – garantia da qualidade – que é um processo executado nas etapas finais das tarefas que cabem ao tradutor no processo de localização. Para simplificar, mas alertando que os passos podem variar para cada caso, depois de traduzir, revisar a tradução e revisar o texto traduzido, o tradutor faz o QA do material localizado.

No QA, são analisados aspectos pontuais da tradução, sejam eles linguísticos ou funcionais. Por exemplo, verifica-se que todo o texto traduzível foi traduzido, verifica-se a orto-

¹ Para se referir aos idiomas, serão usados os códigos aqui indicados entre parênteses. Por padrão, o setor da localização segue o seguinte formato para indicar o idioma: <código ISO 639-1:2002>-<código ISO 3166-1-alpha-2>. A norma ISO 639-1 (ISO, 2013a) define o código do idioma, ao passo que a ISO 3166-1 (ISO, 2013b) especifica o código do país.

grafia, a pontuação. Verifica-se também a terminologia, no chamado QA terminológico. Nessa atividade, como em tudo que envolve localização, há uma presença marcante de tecnologia computacional. Existem tipos de *software* que ajudam no trabalho, processando material bilíngue em busca de problemas como os indicados acima. Com relação ao QA terminológico, esse processamento consiste em verificar se as traduções listadas em um glossário ou base de dados terminológica foram efetivamente usadas no material localizado. Durante essa comparação, o *software* registra as ocorrências que não correspondem em uma lista, a qual é apresentada para o usuário (tradutor). Tendo essa lista em mãos, o tradutor passa a revisá-la, item por item, para validar se as ocorrências anotadas pelo *software* precisam realmente ser corrigidas na tradução.

Até essa etapa não há maiores problemas ou dificuldades. O problema começa quando se observa o que esse *software* coloca na lista. Segue um breve exemplo:

Entrada do glossário	hidden / oculto
Segmento original	Always keep this feature turned on to prevent any hidden programs that affect the operating system of your computer.
Segmento traduzido	Mantenha sempre este recurso ativado para impedir que programas ocultos afetem o sistema operacional do computador.

Um simples ajuste na superfície para fazer concordar o número do adjetivo *oculto* faz com que o *software* retorne a ocorrência como possível erro. Esse tipo de caso é muito frequente. A mesma entrada do glossário, por ser adjetiva, implica que o uso do gênero feminino também seria colocado na lista para o tradutor validar. Esses pequenos aspectos, bem como outros mais complexos (como será visto mais adiante), resultam em uma lista repleta de ruído, ou seja, de itens desnecessários, os quais o tradutor simplesmente ignora². E o problema pode atingir proporções extremas, em que o volume de texto da lista de QA chega a superar o tamanho da tradução (cf. *Acertos e erros do software* na página 153).

Assim, o tradutor investe um grande esforço desnecessário para validar a lista. Além disso, constata-se que faltam no mercado programas elaborados com base em uma fundamentação linguística, que ajudariam a diminuir o problema, reduzindo o ruído nessas listas ao levar em consideração características particulares de cada idioma, como morfologia, variações

² O inverso também é verdade. No momento da detecção das unidades, o *software* também não considera aspectos linguísticos. Sendo assim, pode deixar de anotar alguma ocorrência que estaria de fato incorreta. Embora também seja um aspecto importante, o silêncio, ou seja, os itens que o *software* deixa de marcar, mas que deveria anotar, está fora do escopo deste trabalho.

sináticas e relações anafóricas, variação lexical, aspectos textuais como o uso de elipses e anáforas. É nesse ponto que este trabalho se insere. Por tratar do tema com base nos resultados gerados por processamento computacional para problemas linguísticos, este trabalho dialoga com áreas como o Processamento da Linguagem Natural (PLN).

Nesse sentido, tem grande importância para o desenvolvimento deste trabalho o “equacionamento metodológico global” proposto por Dias da Silva, o qual é “(...) essencial para o desenvolvimento nessa área [do PLN]” (DIAS-DA-SILVA, 2006, p. 106). Segundo o autor, isso envolve uma metodologia que tem três etapas (linguística, linguística-computacional e computacional), conforme o seguinte quadro³:

Quadro 1 - Equacionamento metodológico global do PLN

Tarefas	Resultados
Explicitação do conhecimento	Descrições linguísticas precisas
Representação formal do conhecimento	Representações linguístico-computacionais
Codificação computacional do conhecimento	Implementação computacional

Fonte: Dias da Silva (ibid, p. 123).

Este trabalho enfocará exclusivamente a primeira tarefa, ou seja, explicitar o conhecimento linguístico envolvido na geração de ruído nas listas de QA, com o objetivo de:

- colocar esse conhecimento em evidência e abrir caminho para sua representação formal e subsequente implementação computacional;
- ajudar a melhorar as ferramentas de QA terminológico, tornando-as linguisticamente motivadas.

Para tanto, este trabalho se propõe a analisar uma lista de QA, usando como base teorias que tenham relação direta com o problema. Para sintetizar, a questão envolve:

- Terminologia: glossários digitais estão em jogo;
- Tradução: o conteúdo do glossário é contrastado com o material textual localizado;
- Computação: um *software* executa essa comparação.

Isso implica uma análise com bases distintas: uma terminológica, uma tradutória e outra computacional. Para ficar em linha com a metodologia do PLN, este trabalho centra-se nas duas primeiras, com a intenção de prover insumos que permitam desenvolver *software* de QA

³ Este quadro é uma reprodução daquele encontrado em Dias da Silva (2006, p. 123).

terminológico que seja linguisticamente motivado. A intenção do trabalho é, portanto, fixar a base linguística para o desenvolvimento desse tipo de *software*.

Para tal fim, é necessário que essa base seja sustentada por alguma teoria que seja coerente com aquilo que os dados apresentam, com aquilo que acontece nas traduções no setor da localização. Para isso, as abordagens à Terminologia e à Tradução adotadas neste trabalho são comunicativas, funcionais, cognitivas, pragmáticas. Embora esteja em oposição àquilo que a indústria da localização adota para a Terminologia, ou seja, uma abordagem prescritivista, normativa (nos moldes da primeira teoria moderna sobre o assunto, como se vê em *Teoria Geral da Terminologia* na página 35) e no discurso da localização (cf. *Localização* na página 22), essas abordagens encontram amparo na prática, no cotidiano da localização, ou seja, na linha de frente, prescritividade e normatividade nem sempre são a regra, como se verá na apresentação dos resultados (páginas 152-235).

Nesse aspecto, o objetivo deste trabalho é analisar as ocorrências de ruído em listas de QA, atento à variação terminológica (cf. *Análise variacional* na página 196 para os resultados baseados em critérios relacionados ao assunto), à fraseologia (cf. *Análise fraseológica* na página 160) e à tradução (cf. *Análise tradutória* na página 177), bem como a questões morfológicas e a aspectos textuais (cf. *Análise linguística* na página 216).

Enfim, o percurso seguido é longo e requer um leitor indulgente no quesito “quantidade de matérias diferentes para tratar de um só tema”. O trabalho começa, no Capítulo 2 (p. 22), com uma breve introdução ao mundo da localização, falando de suas origens e processos. O Capítulo 3 (p. 34) apresenta uma revisão teórica, em que são discorridos brevemente os temas centrais à Terminologia, como a oposição entre teorias comunicativas e teorias normativas, passando a tópicos tratados com maior densidade, como a fraseologia (cf. *Fraseologia* na página 43), a variação (cf. *Variação* na página 57) e a tradução (cf. *Tradução* na página 85). Estes tópicos formam a base teórica que sustenta a análise de dados. No Capítulo 4 (p. 115), é apresentada a metodologia aplicada à análise das informações encontradas na lista de QA, bem como os passos necessários para gerá-la. No Capítulo 5 (p. 152), é apresentada a análise dos dados, incluindo alguns números ou estatísticas que têm a função de, nortear, futuramente, o desenvolvimento de *software* linguisticamente motivado ou ajudar a priorizar os problemas linguísticos que devem ser atacados primeiro. Finalmente, as conclusões são apresentadas no Capítulo 6 (p. 238).

2 LOCALIZAÇÃO

Antes de falar do paradigma da localização na revisão teórica e em sua relação com a tradução e os Estudos de Tradução, é necessário contextualizar a área como um todo, descrevendo brevemente sua história, seus conceitos básicos e seus processos. A necessidade de posicionar a área está na constatação de que, embora represente um setor industrial que cresceu em escala astronômica nos últimos anos, a localização consiste em um fenômeno pouco conhecido e compreendido fora de seu círculo mais direto (DUNNE, 2006, p. 1). Além disso, uma visão objetiva sobre o tema facilitará os questionamentos que são feitos sobre esse paradigma.

2.1 O que é localização

Em poucas palavras, o termo *localização* pode definir um processo e um campo de atuação profissional⁴. Como processo, diz respeito à adaptação de produtos (geralmente de *software*), segundo o idioma e as convenções culturais de determinada localidade ou país. Como campo de atuação, consiste em uma atividade que tem o objetivo facilitar a entrada de um produto ou serviço em um país ou mercado. Trata-se do setor industrial relacionado à tradução que provavelmente mais cresce e que tem maior envergadura em termos de giro de capital.

Segundo afirmam alguns trabalhos – como Dunne (2006) e Esselink (2000) – sobre o setor, nas últimas décadas, a localização teve uma expansão meteórica, passando de um trabalho conjunto entre umas poucas empresas de *software* para um setor profissional multibilionário. Um relatório recente (KELLY e DEPALMA, 2012) sobre os números aponta para um capital de quase US\$ 30 bilhões. Em uma cartilha publicada em 2003 por organizações vinculadas à localização, falava-se em um mercado que fazia girar números em torno de US\$ 3,7 a

⁴ A esse respeito, Pym (2010, p. 127) problematiza que muitas vezes se designa localização como o uso de tecnologia, como ferramentas CAT e bases terminológicas. Na verdade, a localização diz respeito mais a uma metodologia do que ao uso concreto de ferramentas. Em última análise, conforme o autor argumenta, é igualmente possível usar essas ferramentas para traduzir outros tipos de materiais.

5 bilhões. Em um intervalo de 10 anos, isso representa um crescimento na casa dos 1000%, e as projeções apontam para um crescimento anual na casa dos 12%. Esse crescimento demonstra a importância desse setor e também a de se estudar alternativas para melhorar as diferentes etapas implicadas no processo de localização.

O desenvolvimento da área, conforme relatado por Esselink (2000) e Dunne (2006), passa por pelo menos quatro momentos:

- no início, as iniciativas de localização estavam confinadas ao âmbito da tecnologia da informação e a grandes empresas de tecnologia. A localização consistia na tradução de *software* para comercialização em mercados internacionais;
- até meados dos anos 90, localização significava a tradução das interfaces de usuário (UI) e da documentação de *software*, bem como no trabalho de reengenharia necessário para garantir a exibição correta das informações na UI e manter a funcionalidade do produto nas versões localizadas.
- com a popularização da internet na metade dos anos 90, o escopo da localização mudou drasticamente, deixando de ser exclusividade de grandes corporações globais. A internet passa a ser uma espécie de oitavo continente, habitado por milhões de consumidores ao redor do globo (DEPALMA, 2002). Com isso, o alcance a públicos globais passa para as mãos de muitos;
- a explosão do comércio eletrônico fomentou a proliferação de uma série de conteúdos localizáveis, em formatos dos mais variados e para um conjunto de usuários e localidades cada vez mais diversificado, todos em franca expansão.

Como afirma Dunne, e conforme se depreende da introdução de Esselink ao tema, a localização consiste em uma série de processos interdependentes e não pode ser compreendida totalmente sem estes. Esses processos são chamados coletivamente de GILT (Globalização, Internacionalização, Localização e Tradução). A seguir é apresentada a definição de cada um desses processos tal qual especificado por representantes do setor. Em referências anteriores a 2011, autores que tratam do tema (como Esselink, Pym e Dunne) faziam referência à extinta associação de padronização do setor da localização, conhecida como LISA⁵. Hoje, o setor recorre a outra associação, a GALA⁶, como mantenedora de seus padrões e definições.

⁵ Sigla para Localization Industry Standards Association.

⁶ Sigla para Globalization & Localization Association.

2.1.1 GILT e outros termos comuns à área

Antes de falar dos detalhes envolvidos em um projeto de localização, em seus processos e nos papéis que cada agente desempenha em sua cadeia de produção, é necessário esclarecer alguns termos comuns à área. Os parágrafos a seguir trazem breves definições para esses termos, levando em conta as informações contidas no *site* da GALA⁷. As definições originais estão em inglês e foram traduzidas pelo autor deste trabalho.

Globalização: diz respeito às atividades de negócio associadas a lançar um produto em escala global. Na globalização de produtos altamente tecnológicos, isso compreende integrar a localização em uma empresa, após a internacionalização e o planejamento do produto. Também envolve *marketing*, vendas e suporte (assistência técnica) em um mercado global.

Internacionalização: é o processo de generalizar um produto de tal maneira que ele possa absorver múltiplos idiomas e convenções culturais sem a necessidade de incorrer alterações em seu projeto. Esse processo ocorre durante o planejamento do *software* e da redação de sua documentação. No que diz respeito ao texto, a internacionalização visa à redação de um texto de partida simplificado, sem marcas idiomáticas, a fim de facilitar a tradução para qualquer idioma. É uma atividade que constrói textos *a priori* descontextualizados e projeta o inglês⁸ a um patamar de língua universal.

Localização: consiste no processo que torna um produto linguisticamente e culturalmente apropriado para a localidade de chegada, na qual o produto será usado e comercializado. Visto de outra maneira, designa parte de um processo que tem o objetivo de vender cada vez mais *software* (ou qualquer produto) em um número cada vez maior de locais e de idiomas (PYM, 2004). Localização também designa a área ou o setor em que esse processo é realizado.

Tradução: é o processo de converter todo o conteúdo textual ou as palavras do idioma original para o de destino. É necessário determinar o contexto ou sentido do idioma original para transmitir a mesma mensagem no idioma de destino.

Localidade⁹: refere-se ao mercado que se quer atingir dentro de um processo de localização, englobando aspectos como idioma e convenções culturais.

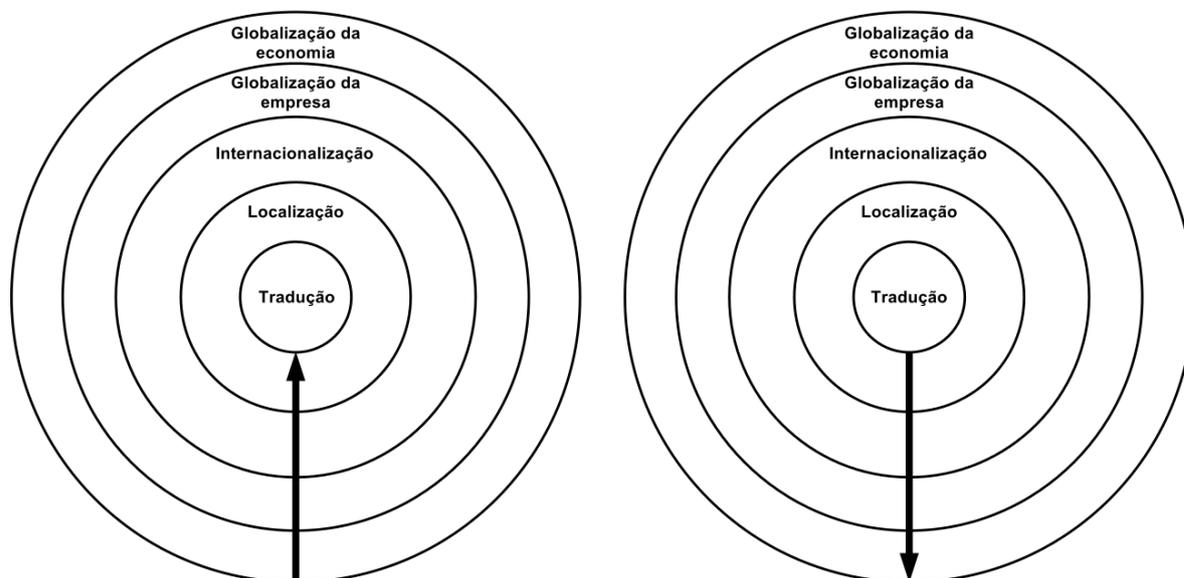
⁷ Os originais se encontram em: <http://www.gala-global.org/taxonomy/term/35>

⁸ Não se trata do inglês americano ou britânico, mas sim de um inglês artificial e genérico que não está atrelado a nenhuma cultura.

⁹ O termo usado originalmente em inglês é *locale*.

A seguinte figura é uma reprodução traduzida de Dunne, indicando a interdependência dos processos do GILT (esquerda) e o caminho para se entender suas dependências (direita):

Figura 1 - Processos do GILT, sua interdependência e caminho para sua compreensão



Fonte: Adaptado de Dunne (2006, p. 5).

Para o autor, a relação de interdependência parte da globalização da economia, passando por vários processos até chegar à tradução. Ao mesmo tempo, é a partir dela tradução que se consegue compreender com mais facilidade os demais processos. Nesse sentido, partindo do diagrama da direita, em termos cronológicos, a sucessão de processos segue essa ordem:

- as primeiras tentativas de *traduzir software* e a respectiva documentação de um idioma para outro levou à conclusão de que havia mais questões em jogo além de tradução¹⁰, dando origem à *localização*;
- depois disso, a análise sobre trabalhos de localização revelou algumas medidas prévias que poderiam facilitar o processo, originando a *internacionalização*;
- a necessidade de reorganizar todo o ciclo de desenvolvimento para dar suporte à distribuição simultânea de várias versões de um produto em idiomas diferentes levou as empresas de *software* a adotarem estratégias globais com o objetivo de facilitar todos os aspectos implicados no processo inteiro.

¹⁰ Aqui, tradução é tomada na definição mais simplista possível, muito longe, por exemplo, daquilo que as teorias do escopo colocavam. De fato, a tradução, vista como projeto no paradigma do escopo, é tratada como uma pequena parte do processo. Para a localização, o termo *localização* está mais próximo daquilo que era definido como tradução no paradigma do escopo.

Para Dunne, essa mudança estratégica de planejamento e trabalho, que faz o foco passar da tradução e localização para a internacionalização, consiste na *globalização*, ou seja, o processo mais abrangente de integração econômica, tecnológica e financeira em nível internacional.

Embora as definições em geral sempre apontem para a cadeia de produção de *software*, hoje a localização tende a englobar mais do que isso. Qualquer produto é passível de ser internacionalizável ou localizável. Um exemplo são as agências de notícias, que contam com processos muito semelhantes aos da localização de *software* para a distribuição de conteúdo em vários idiomas e localidades. Outro exemplo é a indústria automobilística, que usa processos semelhantes para a distribuição dos manuais técnicos de seus produtos, bem como parte de seu material de *marketing*.

2.1.2 Projetos de localização

Existem várias metodologias aplicáveis a um projeto de localização. As variantes envolvem desde os agentes, passando pelos componentes que fazem parte do processo, até as etapas específicas aplicadas. Um projeto clássico, conforme Esselink (2000) descreve, envolve os seguintes agentes:

- cliente: a empresa que faz o produto;
- MLV: da sigla em inglês para *multi-language vendor*. Trata-se de uma empresa que provê serviços de localização para vários idiomas. Normalmente, consiste em uma empresa que concentra a tarefa de preparar o material para tradução, distribuir para tradução terceirizada e consolidar o material traduzido;
- engenharia: empresa/profissional terceirizado ou departamento do MLV encarregado de preparar o material para tradução ou consolidar o material traduzido em uma versão localizada do *software*. No caso de um *software*, a engenharia se encarrega de extrair o conteúdo linguístico para tradução;
- tradução: empresa/profissional terceirizado ou departamento do MLV encarregado de traduzir o material. Como um MLV em geral não tem condições de manter funcionários nativos de todos os idiomas com os quais trabalha, essa tarefa fica mais a cargo de terceiros;

- editoração eletrônica: empresa/profissional terceirizado ou departamento do MLV encarregado de corrigir erros de diagramação em material textual para publicação, como arquivos .pdf ou imagens.

Essa formação, contudo, nem sempre é a que realmente ocorre. Por vezes, o próprio cliente concentra boa parte desses processos, como a engenharia e a editoração eletrônica. Nesse tipo de formação, a tradução seria a única etapa conduzida fora da empresa que localiza seus produtos. Outro caso possível é que a tradução fique a cargo de um SLV (sigla em inglês para *single-language vendor*), designação para empresas que prestam serviços de tradução para um idioma ou, para ser mais preciso, para um número reduzido de idiomas¹¹. O SLV, por sua vez, recorreria a terceiros para prover traduções que são repassadas ao MLV.

Seja qual for a formação, essas diferenças dificilmente representam mudanças drásticas na forma como as atividades são desempenhadas, mas sim na entidade sob a qual estas recaem e na quantidade de trabalho (e retorno financeiro) designado a cada um. Como observação, vale dizer que, quanto menos agentes intermediários existirem no processo, melhor será para um tradutor. A título de exemplo, em um exercício de matemática simples, se o cliente pagasse R\$ 0,20 a palavra do original traduzida e os intermediários (MLV e SLV) ficassem com 40% do montante total, à tradução restaria 60% desse valor inicial, a ser dividido entre tradutor (65% ou R\$ 0,08), editor¹² (25% ou R\$ 0,03) e revisor textual (10% ou R\$ 0,01). Em um relacionamento direto, esse valor seria respectivamente R\$ 0,13, R\$ 0,05 e R\$ 0,2. Isso dá uma diferença de R\$ 80 em um dia de trabalho.

As próximas seções tratam dos componentes e da cadeia de produção da localização. Como o tipo de material que é passível de localização depende do projeto em questão, para ficar com o exemplo mais clássico da área, estas seções descrevem um projeto de localização de *software*. A premissa é que isso seja suficiente para inteirar o leitor que não conhece localização, bem como para dar suporte aos capítulos de metodologia e análise de dados, já que o *corpus* usado está constituído de material produzido para localização de *software*.

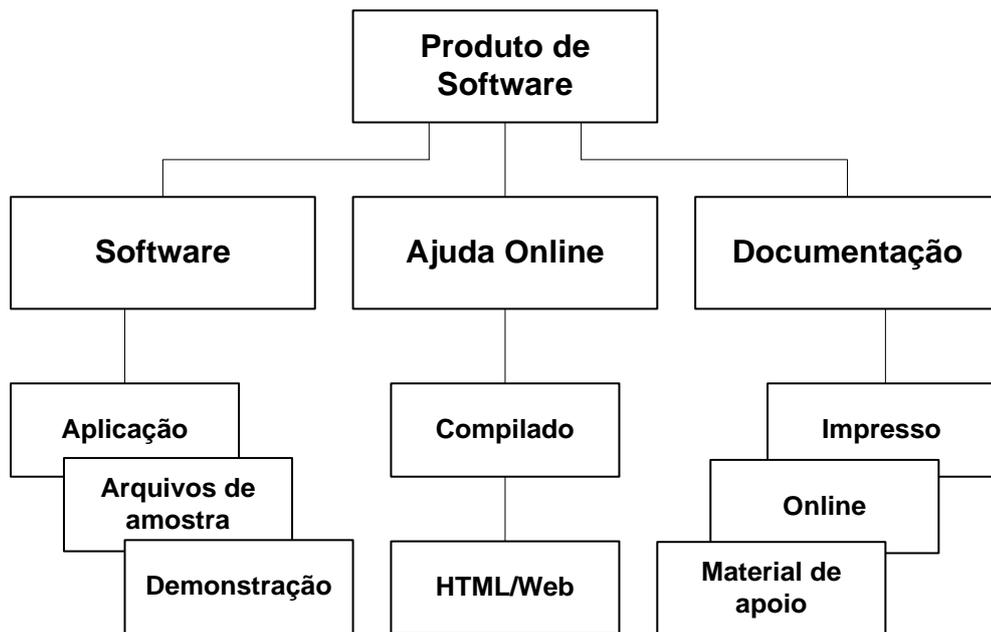
¹¹ O mercado Brasil-Argentina é um exemplo disso. A formação mais comum encontrada é de empresas de tradução baseadas na Argentina que passam a contratar tradutores no Brasil, ou que montam escritórios no Brasil e contratam tradutores *in-loco* para suprir uma crescente demanda por traduções em português brasileiro.

¹² No original, trata-se de *translator*, *editor* e *proof-reader*. No mercado brasileiro, a denominação mais comum para *editor* é “revisor”. No entanto, “revisor” em português também denomina o *proof-reader*, ou o profissional que revisa o texto só no idioma da tradução, sem comparar com o original. Por esse motivo, este trabalho optou por adotar a denominação “editor” para o *editor* do inglês.

2.1.2.1 Componentes de um projeto

Um projeto de localização de *software* envolve mais que a codificação do *software* em si. A documentação de apoio também é pensada e trabalhada desde o início. A figura a seguir traz um resumo dos componentes geralmente envolvidos, conforme apresentado por Esselink (2000):

Figura 2 - Componentes de um típico projeto de localização



Fonte: Adaptado de Esselink (2000, p. 10).

Como se vê, um *software* típico é constituído de três componentes principais: o *software* em si, a ajuda e a documentação. O quadro a seguir descreve brevemente cada um:

Quadro 2 - Descrição dos principais componentes de um *software* típico

Componente	Descrição
<i>Software</i>	Consiste em programa computacional que hoje em dia é disponibilizado com mais frequência via Internet ou intranet. É composto pelos arquivos de programa (como a UI), pela ajuda <i>on-line</i> e por arquivos informativos (como as notas da versão ou leia-me).
Ajuda	Trata-se do material de ajuda que acompanha o <i>software</i> . Geralmente consiste em um arquivo que encapsula vários arquivos que são exibidos em formato .html ao usuário. Podem estar instalados localmente (como um arquivo .chk ou .hlp no Windows) ou ser acessíveis <i>on-line</i> no <i>site</i> de suporte da desenvolvedora do <i>software</i> . Centra-se mais no uso direto do <i>software</i> e é acessível diretamente da UI.
Documentação ¹³	Constitui o material que descreve o <i>software</i> em linhas mais gerais, centrando-se em aplicações de uso, por exemplo. Também pode incluir material de apoio a vendas, como folhetos de <i>marketing</i> transfigurados em referência técnica.

Fonte: Esselink (2000, p. 10-12); traduzido pelo autor.

2.1.2.2 Cadeia de produção

A localização está situada dentro da cadeia de produção de um *software*, a qual envolve o desenvolvimento de seu código, a redação da respectiva documentação, montagem, empacotamento, distribuição, entre outros. Esta seção traz uma imersão em três diagramas com o objetivo de situar a parte exata no processo em que se insere a validação terminológica tratada neste trabalho. O seguinte diagrama situa a localização dentro da cadeia de produção de um *software*, nos moldes apresentados pela organização LISA.

¹³ Na definição de Esselink (2000, p. 12). Também é possível pensar em documentação como o material que descreve o *software* em linhas específicas, como documentos de arquitetura e especificações técnicas.

Figura 3 - Cadeia de produção de *software*



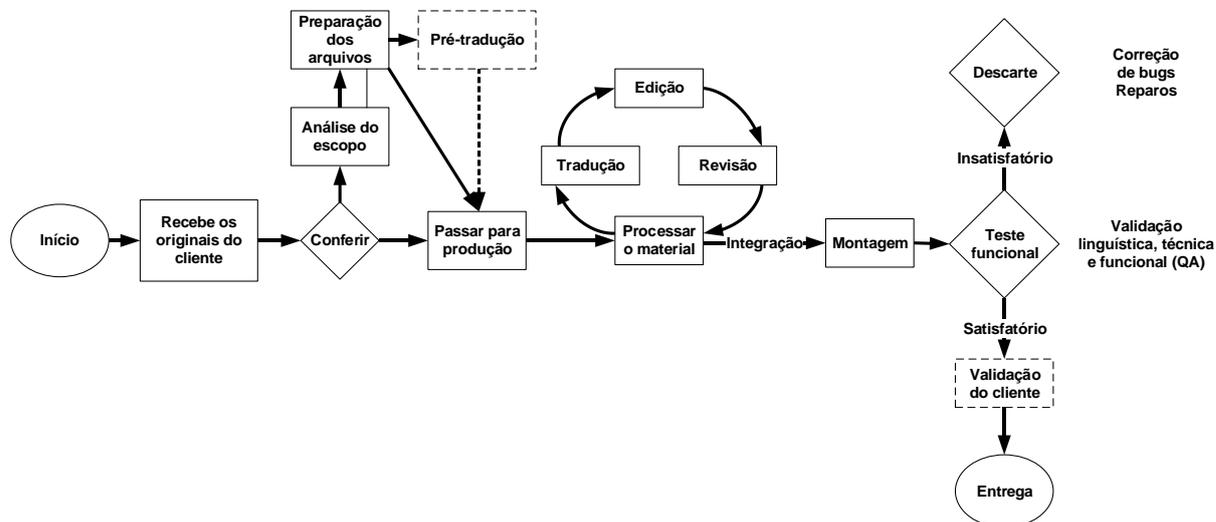
Fonte: Adaptado de Lommel e Ray (2007, p. 19).

É necessário esclarecer algo a respeito desse diagrama. Nele, existem duas etapas de QA: uma em localização e outra em internacionalização. Na verdade, o QA tratado neste trabalho não é nenhum desses; ele está dentro de *Localização do produto*, que será tratada a seguir. As etapas de QA no diagrama acima são funcionais, ou seja, servem para validar que o *software* e sua documentação estão de acordo com os padrões de determinada localidade e que sua apresentação é consistente. Para os padrões, por exemplo, verifica-se se o *software* mostra o indicador de casa decimal correto. Para a apresentação, valida-se que o *software* não apresenta texto truncado ou cortado na UI.

Aproximando-se um pouco mais da parte que envolve o tradutor diretamente, a cadeia de produção da localização (tipicamente, baseada na norma ISO 9001:2000), conforme descrevem Esselink (2000) e Dunne (2006), está representada no diagrama a seguir. A etapa cí-

clica ao centro representa o TEP QA (cf. abaixo na seção *TEP QA em detalhes*, página 31). Esse é o ponto que tem envolvimento direto do tradutor.

Figura 4 - Aplicação típica da ISO 9001:2000 no setor da localização



Fonte: Adaptado de Dunne (2006, p. 101).

2.1.2.3 TEP QA em detalhes

TEP + QA são duas siglas que designam a cadeia de processos da qual o tradutor faz parte, consistindo em *translation, editing, proof-reading* e *quality assurance* – respectivamente, tradução, edição (revisão da tradução comparada ao original), revisão textual (somente do texto na língua de chegada) e garantia da qualidade (QA). Conforme atenta Pym (2011, p. 93), a boa ideia por trás do processo é que envolveria três pessoas: o tradutor, o editor e o revisor textual.

Dentro do TEP, o QA consiste em uma etapa de validação pontual e final da tradução. Normalmente, é feito pelo tradutor ao concluir seu trabalho e sucessivamente pelo editor, ao conferir o trabalho do tradutor. A execução do QA passa por dois processos. Em um primeiro momento, o profissional executa um *software* de computador que verifica automaticamente uma série de aspectos na tradução, tais como:

- segmentos¹⁴ sem tradução;

¹⁴ Na área da localização, usa-se o termo *unidade de tradução* para se referir a segmentos de tamanho variável que são armazenados paralelamente em dois ou mais idiomas nas memórias de tradução ou em arquivos bi ou multilíngues. Contudo, para não confundir com a noção de *unidade de tradução* (cf. *Unidade de tradução* na

- consistência na tradução de segmentos idênticos;
- presença de segmentos cujo original é idêntico à tradução;
- espaços duplos entre palavras;
- palavras repetidas;
- uso incorreto de *tags*¹⁵;
- ortografia;
- terminologia.

Após essa verificação, o *software* retorna uma lista de validação, a qual o profissional deve conferir para validar que as traduções estão corretas ou, caso contrário, para corrigir eventuais erros. Como a lista acima mostra, um dos pontos analisados é a terminologia. No QA de terminologia, o computador compara, mediante cruzamento de padrões, o conteúdo de um glossário bilíngue com as traduções contidas em um texto bilíngue. Os itens que não casarem são registrados na lista de validação para que o tradutor ou editor faça a devida conferência¹⁶. A figura a seguir ilustra uma lista gerada por um *software*¹⁷ de QA:

página 107) usada neste trabalho, será usado o termo *segmento de tradução* ou apenas *segmento* para se referir a estes textos paralelos.

¹⁵ Denominação emprestada do inglês, *tag* consiste em um elemento formal que identifica determinado elemento em um programa de computador ou uma linguagem de programação.

¹⁶ Para exemplos de *software* que executa esse QA, bem como de lista de validação, consulte *Execução do QA terminológico* (p. 135).

¹⁷ A imagem se refere ao *software* ApSIC Xbench (cf. *ApSIC Xbench 2.9* na página 130).

Figura 5 - Lista gerada por *software* de QA

The screenshot displays a QA software interface with the following components:

- Check Group:** Basic (checked), Content (checked), Checklists (checked).
- List of Checks:** Untranslated Segments (checked), Inconsistency in Source (checked), Inconsistency in Target (checked), Target same as Source (unchecked).
- Options:** Only New Segments (unchecked), Exclude ICE Segments (unchecked), Case-sensitive Inconsistencies (unchecked), Ignore Tags (unchecked).
- Filter Issues:** Show All (selected), Show Marked (unchecked), Hide Marked (unchecked).
- Buttons:** Check Ongoing Translation, Run Project Checklists, Run Personal Checklists, Clear All Marks.

Error Description	%	Source	Target
Inconsistency in Target			
Symantec - Documentation.tmx (83)	0%	A portal is a collection of links, content, an...	Portal é um conjunto de links, conteúdo e ...
Symantec - Documentation.tmx (184)	0%	A portal is a collection of links, content, an...	Portal é um conjunto de links, conteúdo e ...
Symantec - Documentation.tmx (1741)	0%	A portal is a collection of links, content, an...	Portal é um conjunto de links, conteúdo e ...
Inconsistency in Target			
Symantec - Documentation.tmx (1742)	0%	The reputation details indicate whether th...	suportados.
Symantec - Documentation.tmx (84)	0%	The reputation details indicate whether th...	Os detalhes de reputação indicam se o a...
Inconsistency in Target			
Symantec - Documentation.tmx (1743)	0%	You can use these details to decide the a...	Os detalhes de reputação indicam se o a...
Symantec - Documentation.tmx (85)	0%	You can use these details to decide the a...	Você pode usar esses detalhes para de...
Symantec - Documentation.tmx (1815)	0%	You can use these details to decide the a...	Você pode usar esses detalhes para de...
Inconsistency in Target			

Detail

- **2: Inconsistency in Target**
- A portal is a collection of links, content, and services that are designed to guide users to information they are likely to find interesting.
- Portal é um conjunto de links, conteúdo e serviços desenvolvido para direcionar os usuários a informações que eles possam achar interessantes. ~~suportados.~~
- Portal é um conjunto de links, conteúdo e serviços desenvolvido para direcionar os usuários a informações que eles possam achar interessantes. suportados.

Fonte: Elaborado pelo autor.

Para terminologia, no entanto, em geral essa função se limita a um cruzamento de padrões básicos, ou seja, busca apenas a expressão exata no original e na tradução. Qualquer divergência, como um *s* de plural, já faz com que um possível item a verificar seja ignorado pelo *software*. Isso significa, não é possível fazer essa tarefa sem uma boa dose de trabalho braçal.

3 REVISÃO DA LITERATURA

3.1 Terminologia

Terminologia hoje é um tópico com uma série de designações distintas e relacionáveis: conjunto de práticas e métodos, teoria, disciplina, vocabulário de algum campo específico, campo de estudos. Os avanços na formação de um campo teórico sobre o assunto (uma ou várias Teorias da Terminologia) deu contornos claros à distinção entre a prática terminológica e os Estudos Terminológicos.

Historicamente, a Terminologia como campo de estudos encontra sua origem nas práticas terminológicas. Nas práticas desenvolvidas com a expansão da indústria a partir de meados do séc. XX, origina-se o trabalho considerado marco da Terminologia moderna: o estudo seminal de Wüster (1998). Esse trabalho serviu de base para várias perspectivas teóricas, inclusive as que o contrapõem e que deram origem a métodos mais abrangentes para lidar com o fenômeno terminológico, especialmente com base nos avanços das Ciências Cognitivas. Seus fundamentos ainda recebem ampla adoção na indústria, inclusive na localização. Como este trabalho se propõe analisar dados encontrados na prática terminológica e tradutória desse mercado, é preciso refletir sobre a evolução da disciplina em suas principais vertentes para encontrar aquela que possa ser aplicada com mais efetividade no nosso trabalho.

As próximas subseções trazem uma síntese das duas principais perspectivas dos Estudos Terminológicos. A primeira apresenta a Teoria Geral da Terminologia (TGT), que é considerada a teoria tradicional da Terminologia e que é amplamente seguida em setores industriais e na localização. A seção traz ainda uma breve síntese dos desdobramentos dessa teoria na atualidade. A segunda introduz a Teoria Comunicativa da Terminologia (TCT), que consiste em uma das muitas contrapropostas à teoria tradicional e que talvez seja a mais representativa entre estas. Essa vertente apresenta um fundo teórico abrangente, que a permite incorporar diversas perspectivas sobre terminologia.

3.1.1 Teoria Geral da Terminologia

A Teoria Geral da Terminologia (TGT) foi desenvolvida pela Escola de Viena em meados do séc. XX com base nas propostas de Eugen Wüster, engenheiro e terminólogo austríaco, considerado o fundador da Terminologia moderna. Trata-se da proposta terminológica mais conhecida, tendo inspirado a maioria dos trabalhos feitos até hoje, seja seguindo seus fundamentos, seja os contrapondo. Segundo Cabré (1998a, p. 7), algumas de suas características mais relevantes são o enfoque onomasiológico e a orientação para a normalização de termos e conceitos, tomando por base a lógica, a busca de uma língua universal e a uniformidade da comunicação.

A TGT tem sua origem na expansão do conhecimento técnico e científico, com a necessidade que profissionais e pesquisadores têm de padronizar os conceitos e denominações de suas disciplinas e práticas para facilitar a comunicação profissional e a transferência de conhecimento. Segundo Cabré, a TGT considera a Terminologia um tópico interdisciplinar, porém autônomo, que está a serviço de disciplinas científicas e técnicas. A Terminologia não é um fim em si, e a prática terminológica não se restringe à elaboração de compilações de conceitos e dos nomes correspondentes: a Terminologia está a serviço da ciência, tecnologia e comunicação. Por isso, estaria limitada a prestar serviço a outras disciplinas.

A TGT estabelece alguns fundamentos que até hoje, datadas décadas de sua existência, ainda formam a base de muitos outros trabalhos sobre Terminologia, mesmo aqueles que se opõem com mais veemência aos postulados dessa teoria. Seguem alguns dos fundamentos de maior destaque da TGT, inter-relacionáveis em algum grau, dependendo da perspectiva projetada sobre eles: separação entre conceito e termo (denominação); o conceito é visto como uma unidade supralinguística e, portanto, interlinguística;

- perspectiva onomasiológica: parte dos conceitos no lugar da unidade lexical;
- univocidade: a razão entre unidade lexical e conceito deve ser de 1:1;
- enfoque lexical: entre os níveis de análise linguística, à Terminologia interessa apenas o nível do léxico; morfologia ou sintaxe ficam de fora;
- normatividade: a Terminologia deve ser necessariamente prescritiva; a linguística, descritiva;
- interdisciplinaridade: a Terminologia pressupõe intercâmbio com ciência e tecnologia;
- prevalência da forma gráfica sobre a fônica.

Entre essas características, nenhuma livre de reservas na literatura que trata do tema, a interdisciplinaridade, a perspectiva onomasiológica e o enfoque sincrônico são exemplos notáveis de fundamentos seguidos por diversas vertentes teóricas em Terminologia. Entre as características mais contestadas estão seu caráter prescritivo atrelado à univocidade, bem como o enfoque lexical. Como a Terminologia tem de necessariamente ditar como a comunicação técnica e científica deve ser conduzida, usando conceitos com contornos nítidos e sua relação transparente com uma única unidade lexical, a TGT não dá conta de uma série de necessidades que uma teoria da Terminologia teria de suprir. Por polarizar traços distintivos marcantes entre língua geral e terminologia, entre Linguística Geral e Teoria da Terminologia, a TGT afasta do escopo dos Estudos Terminológicos uma série de objetos de análise e características que também podem ser do interesse de grupos que se dedicam à área e procuram contribuir para trabalhos teóricos e práticos neste campo.

Segundo Cabré (2003), os postulados da TGT se mostram insuficientes para dar conta das novas necessidades comunicativas que surgem no final do séc. XX. Recorrendo à exposição de vários ângulos distintos – recurso que se materializa em sua teoria das portas –, a autora exemplifica esses requisitos em suas múltiplas dimensões, em termos de estudos sociais, linguísticos e cognitivos:

- para as ciências sociais, surge a necessidade de tratar da terminologia em situações reais de comunicação, o que requer mais que sua normatização;
- para as ciências linguísticas, é necessário trabalhar a terminologia como linguagem natural, dentro de modelos gramaticais que contemplem a diversidade de variedades linguísticas e comunicativas;
- para as ciências cognitivas, é necessário ir além da idealização do conceito para abarcar sua complexidade cultural e social.

Nesse ponto, segundo a autora, é necessária uma ramificação ou desvio de percurso para superar essas limitações e assim suprir as novas necessidades surgidas na comunicação especializada. Uma opção é ampliar os limites da teoria wüsteriana; outra é buscar uma base que origine uma nova Teoria da Terminologia que tome por base os fundamentos da linguagem e seu caráter sociocultural. Caminho este que a autora segue ao propor sua Teoria Comunicativa da Terminologia e que não é necessariamente excludente em relação à TGT, pois há situações em que pode caber a normatização.

Seja adotando ou não um caminho novo, as propostas que sucedem devem em algum grau sua fundamentação à TGT, seja por contrapor alguns dos principais postulados da teoria wüsteriana, seja por ampliá-los.

Investigações recentes¹⁸ na linha wüsteriana – Myking (2007), Picht (1990), Roche (2012) – apontam para uma revisão de alguns temas centrais da TGT. Caem os limites bem definidos entre termos e léxico comum, o que permite, por exemplo, incorporar uma análise fundamentada em Estudos Linguísticos dentro dessa perspectiva teórica. Com base nesse pressuposto de indefinição, essa vertente também passa a reconhecer que a terminologia pode ser descritiva; o enfoque prescritivista ainda predomina, contudo.

3.1.2 Teoria Comunicativa da Terminologia

A Teoria Comunicativa da Terminologia (TCT) foi elaborada por Cabré (1998a, 1998b, 1999a, 1999b, 1999c, 2001 e 2003) nas últimas décadas do séc. XX, partindo de uma revisão da TGT para cobrir necessidades que esta não pôde suprir, segundo a autora, em função de seu caráter idealista e reducionista. Cabré (2001, p. 19) enumera algumas das características do léxico especializado que a teoria wüsteriana não daria conta:

- poliedricidade denominativa, cognitiva e funcional;
- dupla função, representativa e comunicativa;
- variação inerente à comunicação, em função de características como emissor e receptor, contexto sociocultural, situação comunicativa;
- diversidade aplicada.

Outra força motriz para se propor uma nova base teórica provém da metodologia, visto que a TGT cobriria apenas determinados tipos de trabalhos (prescritivos, de normalização). Cabré cita três fatores que demandariam a revisão da metodologia proposta pela TGT e condensados em normas como a ISO:

- diversificação das necessidades terminológicas atuais, partindo das fronteiras da indústria para outras áreas, como a de planejamento de linguagem (tendo sido o caso da Catalunha um forte motivador para Cabré ter buscado uma nova proposta terminológica);
- tendência de adequar o planejamento de trabalhos e suas aplicações às características de cada aplicação, contexto e situação;

¹⁸ A estes desenvolvimentos baseados na TGT, Cabré (2003) chama “Extended general theory” (ou “teoria geral expandida” em tradução livre).

- introdução de novas tecnologias da informação e comunicação no trabalho terminológico

Dadas as necessidades motivadas por esses fatores e das limitações da TGT, Cabré propôs um modelo teórico mais aberto, que pudesse descrever o léxico especializado tendo em conta sua complexidade, além de uma metodologia que fosse coerente com esse modelo.

Para tanto, a TCT está fundamentada em dois pilares:

- uma perspectiva linguística, que concebe a linguagem ao mesmo tempo como sistema gramatical de significação e denominação e como sistema de cognição e comunicação ancorado em uma realidade sociocultural; nessa perspectiva, as unidades passam a ser tratadas como signos de linguagem, seguindo portanto as mesmas regras gramaticais que o léxico comum e passando a assumir aspectos como a variação;
- uma visão de texto como hábitat natural das unidades especializadas em comunicação natural¹⁹.

Cabré parte de duas premissas para fundamentar sua proposta. A primeira trata do caráter multifacetado da Terminologia como conjunto de necessidades, práticas e campo de conhecimento. A segunda coloca as unidades terminológicas no centro de qualquer análise e aplicação que se fizer em Terminologia.

Sobre estes dois pilares e com base nas duas premissas, Cabré (2001, p. 23-25) enumera alguns fundamentos para o aporte teórico da TCT. Segue uma síntese.

3.1.2.1 Campo interdisciplinar

A Terminologia é um campo interdisciplinar que bebe de uma teoria do conhecimento, uma teoria da comunicação e uma teoria da linguagem. Cada uma dessas perspectivas tem seus próprios propósitos e objetos de análise:

- a primeira explica como retratar a realidade em conceitos, os tipos possíveis de conceituação e a relação dos conceitos entre si e suas possíveis denominações. A unidade mínima é a unidade de conhecimento (UC), e a subunidade específica à Terminologia é a unidade de conhecimento especializado (UCE);
- a segunda descreve os tipos de situações em que podem surgir, a correlação entre o tipo de situação e o tipo de comunicação e as características, possibilidades e limites

¹⁹ Essa visão encontra-se originalmente em Krieger (2006).

dos sistemas de expressão de um conceito e suas unidades. Nesta perspectiva, as unidades se chamam unidades de comunicação especializada (UNICOME);

- a terceira trata das unidades de significação especializada (USE) na linguagem natural, considerando a gramática das línguas naturais sem perder de vista o caráter especializado dessas unidades, e explica como esse caráter é ativado na comunicação. Essas unidades incluem as unidades terminológicas (UT).

3.1.2.2 Objeto de estudo

O objeto de estudo da TCT são as unidades de significação especializada (USE). A TCT as descreve como unidades autônomas que formam um léxico especializado ou como unidades denominativo-conceituais, capazes de fazer referência e exercer funções distintas, como referencial, expressiva e conativa, entre outras. Ainda que autônomas, estas unidades fazem parte do léxico geral e, portanto, obedecem às regras da gramática de cada língua. A TCT não faz uma descrição mais detalhada de outros tipos de USE (por exemplo, as unidades fraseológicas), mas assinala que os mecanismos gramaticais dos outros tipos ainda estão por ser explorados.

3.1.2.3 Termo

Os termos são unidades lexicais que, a exemplo das unidades lexicais de uso geral, têm múltiplas facetas, sendo ao mesmo tempo unidades de conhecimento, unidades de linguagem e unidades de comunicação. A diferença está no conjunto de condições restritas que a unidade terminológica²⁰ (UT) deve satisfazer em cada um de seus três componentes (cognitivo, gramatical e pragmático). Para Cabré, tratam-se de unidades que manipulam com mais eficiência o conhecimento sobre determinado tópico.

A tabela a seguir lista as condições propostas pela autora:

Tabela 1 - Condições que as unidades terminológicas devem satisfazer para cada componente

Componente	Condições
-------------------	------------------

²⁰ Termo que será empregado neste trabalho.

Componente	Condições
Cognitivo	<ul style="list-style-type: none"> • dependem de um contexto temático; • ocupam um lugar preciso em uma estrutura conceitual; • seu significado é determinado pelo lugar que ocupa nessa estrutura; • seu significado é expressamente fixado; • seu significado é considerado uma propriedade da unidade; • são fixadas, reconhecidas e disseminadas com a ajuda da comunidade especializada.
Lexical	<ul style="list-style-type: none"> • são unidades lexicais, por esta ser sua origem ou por ter passado por um processo de lexicalização; • podem ter estrutura lexical ou sintática; • como estruturas lexicais, exploram todos os mecanismos de formação de palavras e processos para a aquisição de novas unidades; • formalmente, podem coincidir com unidades que pertencem ao discurso geral; • quanto a classe gramatical, ocorrem como substantivos, verbos, adjetivos ou advérbios ou como estruturas nominais, verbais, adjetivais ou adverbiais; • pertencem a uma das categorias semânticas mais gerais: entidades, eventos, propriedades ou relações; essas categorias, com suas próprias subcategorias, não são excludentes entre si e não devem ser consideradas valores semânticos; • seu significado é discreto em uma área ou tópico específico; • seu significado é extraído do conjunto de informações que formam uma unidade lexical; • sua combinabilidade sintática é regida pelos princípios combinatórios de todos os itens lexicais de uma língua.
Comunicativo	<ul style="list-style-type: none"> • ocorrem em discurso especializado; • formalmente, adaptam esse tipo de discurso segundo suas características temáticas e funcionais; • compartilham o discurso especializado com unidades que pertencem a outros sistemas semióticos; • são adquiridas mediante um processo de aprendizado e são manipuladas pelos especialistas da área em questão; • são basicamente denotativas (o que não exclui conotações).

3.1.2.4 Conceito

O conceito consiste em um conjunto de traços que corresponde a um nó cognitivo de uma estrutura conceitual em um contexto especializado. Assim como os demais fundamentos, é possível observá-lo na perspectiva cognitiva, linguística e sócio-comunicativa. Os conceitos especializados são precisos e denotativos, ao contrário dos conceitos de língua geral, que são polissêmicos, têm muitas conotações e semanticamente subjetivos. As relações entre os conceitos são mais abrangentes que as encontradas na TGT; além das lógicas e ontológicas, há uma ampla gama, embora Cabré não diga quais são especificamente, assinalando que se trata de um tema ainda pouco explorado.

3.1.2.5 Valor

Importado da teoria saussuriana, o valor é um elemento de metalinguagem que compõe o signo, ao lado do significante (denominação) e do significado (conceito). Nessa tradição, o valor representa as possíveis relações entre os constituintes do signo. Ele reúne e soma todas as outras características da língua não abarcadas por significado e significante. Isso tem uma consequência metodológica importante: é preciso considerar o todo, e não elementos separados (NORMAND, 1990). Talvez não seja exagero afirmar que este é o fundamento mais básico da TCT. Por ampliar o escopo de análise, o valor é o componente da língua que permite observar as unidades especializadas fora do par denominação-conceito²¹.

A título de exemplo, na TCT, este fundamento tem consequência sobre a visão de:

- conceito: pode fazer parte de mais de uma estrutura com valores iguais ou diferentes;
- termo: não pertence a uma área específica, mas é usado nessa área com um valor específico e singular; também pode ser usado em mais de uma área, apresentando valor idêntico ou com algum grau de diferença.

²¹ A TGT também se baseia no modelo saussuriano, mas deixa de lado a noção de valor.

3.1.2.6 Abordagens à Terminologia

A Terminologia, como campo de conhecimento e trabalho, pode ser abordada como teoria, descrição e aplicação. Como teoria, o objetivo é descrever as unidades terminológicas em termos formais, semânticos e funcionais, explicar como obtêm seu valor terminológico e descrever suas relações com outros tipos de signos. Como aplicação, o principal objetivo é compilar as unidades que têm valor terminológico em determinado contexto e estabelecer suas características.

3.1.3 Síntese

Ao se considerar as características estabelecidas para textos produzidos para localização, observa-se um alinhamento com a TGT; características dessa vertente teórica podem ser associadas àquelas encontradas na localização em uma razão de 1:1, como univocidade (consistência) e prescritividade. Contudo, o que se observa em textos originais escritos para localização e em suas respectivas traduções não corrobora totalmente com esse alinhamento. Por exemplo, ao contrário de univocidade e consistência, a variação (denominativa e conceitual) é um aspecto que ocorre com frequência, andando na contramão do discurso e da propaganda do setor da localização; a prescritividade, apesar de endossada, reforçada e imposta em muitos casos, também é relaxada e até ignorada em outros.

Outra questão está relacionada ao tipo de léxico que figura nos glossários encontrados no setor da localização e sua relação com a tradução. Para a TGT, os termos pertencem basicamente à classe gramatical dos substantivos; em sua versão expandida, as denominações passam a incorporar também verbos e fraseologias. Em glossários do setor, encontra-se todo tipo de léxico, inclusive adjetivos e até frases inteiras. Em situação real de uso, não é raro encontrar traduções que reformulam a sintaxe em comparação com o original e, com isso, acabam por não usar a forma arrolada para determinado termo, mas outra representação para o esquema conceitual subjacente a este.

Enfim, fundamentar este trabalho com a TGT poderia resultar em uma descrição irreal ou inadequada dos fenômenos que se pretende descrever. Embora a TGT represente o fundo teórico que é amplamente aplicado no setor da localização, o aporte teórico proposto pela

TCT permite uma análise mais realista e abrangente, sem, no entanto, deixar de lado o componente prescritivista da terminologia dessa área.

3.2 Fraseologia

Assim como outros grandes temas de investigação dos Estudos Linguísticos, a Fraseologia recebe tratamentos diversos que dependem de questões como a perspectiva adotada e os objetivos de cada estudo, em uma visão que passou a ficar clara para os Estudos Linguísticos a partir dos questionamentos saussurianos. Por esse motivo, este trabalho apresenta um breve panorama sobre o tema para mostrar os critérios que serão adotados para analisar esse tipo de fenômeno, o objeto de estudo e seus limites, bem como a denominação que será empregada.

Alguns autores²² apontam uma grande diversidade de concepções para a *fraseologia*. Essa diversidade passa por sua aceção, por sua designação, pela delimitação e pelas características de seu objeto de estudo, pelas metodologias e critérios para seu tratamento.

Em linhas gerais, segundo alguns pesquisadores desse fenômeno, a fraseologia pode designar o estudo das unidades fraseológicas (BEVILACQUA, 1999), o conjunto desse tipo de unidade (AGUADO DE CEA, 2007; BEVILACQUA, 1996 e 1999; FIALA, 1987; GOUADEC, 1994) ou as próprias unidades (AGUADO DE CEA, 2007; BEVILACQUA, 1996). Além dessa abrangência conceitual, contribui para a diversidade o fato de a fraseologia combinar gramática, léxico e semântica, bem como outros aspectos, como a morfologia, que normalmente são tomados como uma característica implícita ou conhecimento pressuposto, mas que precisam ser descritos de maneira mais explícita para fins de processamento computacional.

Ainda nessa esteira, Bevilacqua (1999) apresenta dois fatores hipotéticos e correlatos para essa diversidade de definições e delimitações. Sobre o primeiro fator, a autora fala dos distintos pontos de vista sobre os quais se observa o fenômeno, como a perspectiva da lexicografia, da terminologia, da linguística computacional, da tradução, da produção textual, entre outras; sobre o segundo fator, aponta para a falta de uma proposta robusta e abarcadora que reúna os mais variados critérios, como os linguísticos, pragmáticos e quantitativos, e que permita uma análise mais ampla e completa sobre a fraseologia.

²² Aguado de Cea (2007), Bevilacqua (1996, 1999 e 2004), Blais (1993), L'Homme (2000).

Nos próximos tópicos, este trabalho abordará a questão sob o ponto de vista da Terminologia, ou seja, na perspectiva da Fraseologia da Língua de Especialidade (FLE).

3.2.1 Denominações

A diversidade de denominações para tratar do fenômeno é extensa. Bevilacqua (1996 e 1999), em suas investigações sobre a fraseologia jurídico-ambiental, relaciona uma ampla e distinta gama de termos empregados para cobrir o tópico, como, por exemplo:

- colocação das línguas de especialidade (L'HOMME, 1998);
- coocorrente (LAINÉ, PAVEL, BOILEAU, 1992; PESANT E THIBAUT, 1993);
- entidade fraseológica (GOUADEC, 1994);
- locução fraseológica (BALLY, 1951);
- unidade fraseológica (GOUADEC, 1994 e BEVILACQUA, 1996);
- unidade fraseológica especializada (CABRÉ, LORENTE e ESTOPÀ, 1996).

Como a autora argumenta, sua lista não é exaustiva. Na literatura, é possível encontrar outras denominações, como *expressão multilexêmica*²³ (GUENTHNER e BLANCO, 2004) *expressão fixa, congelada* ou *cristalizada*²⁴ (GROSS, 2004) e *multiword expression* ou *multiword unit* (Heid, 2007), estas últimas associadas ao tratamento do fenômeno no âmbito do PLN ou da linguística computacional.

Especificamente com relação às línguas de especialidade, Bevilacqua (1999) faz um levantamento dos termos empregados na literatura que se referem à fraseologia especializada. Em sua relação, subdivide as denominações em três grupos segundo características comuns aos trabalhos analisados. A seguinte tabela sintetiza os achados da autora:

²³ Do inglês, *multi-lexemic expression*.

²⁴ Estas três denominações são traduções do francês *expression figée*. Observa-se que a tradução também pode ter participação nessa diversidade denominativa.

Tabela 2 - Denominações para unidades fraseológicas segundo levantamento de Bevilacqua (1999)

	Perspectiva sobre o tema	Terminologia usada
Grupo 1 (L'HOMME, 2000)	Define as unidades fraseológicas como colocações	<i>colocação, colocação lexical especializada, combinatória, coocorrência, comportamento sintagmático, fraseologismo, unidade fraseológica</i>
Grupo 2 (CABRÉ, LORENTE e ESTOPÀ, 1996; BEVILACQUA, 1999 e 2004)	Trata as unidades como sintagmas nominais, verbais, adjetivais ou pronominais	<i>coocorrente, frasema, fraseologismo, LSP phrase, unidade fraseológica, unidade fraseológica especializada</i>
Grupo 3 (GOUADEC, 1994)	Engloba unidades sintagmáticas e fórmulas	<i>coocorrente, entidade fraseológica, expressão formulaica, fragmento de frase, fraseologismo, unidade fraseológica</i>

Fonte: Elaborado pelo autor com base em levantamento feito por Bevilacqua (1999).

No entanto, mais importante é observar que, conforme a autora indica, algumas denominações são comuns aos três grupos, demonstrando que a mesma denominação não corresponde necessariamente a determinado objeto de análise entre os diversos trabalhos encontrados na literatura sobre o tema. Há, portanto, variação conceitual, isto é, uma mesma denominação pode referir-se a diferentes conceitos de unidade fraseológica.

Este trabalho usará a denominação *Unidade Fraseológica* (UF) para se referir às unidades observadas sobre esse fenômeno no âmbito da língua geral e *Unidade Fraseológica Especializada* (UFE) no âmbito das línguas de especialidade. A escolha por esses termos são baseadas principalmente no alinhamento deste trabalho com a proposta de Bevilacqua (2004) para a identificação e descrição dessas unidades, bem como na adesão da definição proposta pela autora.

3.2.2 Definições

Seguindo o agrupamento apresentado na seção anterior, Bevilacqua (1999) aponta para três tendências para o tratamento da fraseologia. A primeira define as UFE como colocação. Nesta perspectiva, a UFE é definida pela combinação de duas unidades lexicais, em que uma é o núcleo (terminológico ou não) e a outra é o colocado (p.ex.: *ruído infernal*). A segunda tendência define UFE como as combinações sintagmáticas que apresentam uma UT em seu núcleo terminológico (p.ex.: *baixar a febre* e *causar degradação ambiental*). A terceira tendência considera as UFE como formulações próprias de um domínio discursivo; são unidades que, além de se enquadrar na delimitação proposta pelas duas primeiras, podem chegar ao nível de uma frase completa (exemplos nos parágrafos seguintes).

Com relação a sua caracterização, é possível identificar alguns pontos em comum entre as três tendências, como o caráter semifixo e a não-composicionalidade. A primeira característica é definida como a possibilidade de uma UFE acolher comutações em sua composição (por exemplo, *firmar um contrato* e *assinar um contrato*). A segunda característica consiste na impossibilidade de deduzir o significado da UFE pela interpretação individual dos elementos que a compõem (por exemplo, *chutar o balde*).

A segunda tendência trabalha muito com a noção de estabilidade, em que as unidades podem apresentar diferentes graus de fixação. Esta noção pode ser entendida como a capacidade de inserir elementos entre aqueles que compõem uma UFE (estabilidade sintática) e a possibilidade de comutação entre seus elementos (estabilidade semântica).

Já a terceira tendência expande a delimitação de UFE das anteriores; Gouadec (1994), por exemplo, considera em sua análise até mesmo o nível da frase e atribui às UFE um caráter matricial. O autor usa principalmente os critérios de estereotipia e frequência aliados à noção de variabilidade e invariabilidade para caracterizar as UFE. Nessa tendência, a UFE se apresenta como um conjunto de caracteres estabilizados e recorrentes no qual alguns de seus elementos são fixos e outros variáveis, formando unidades matriciais. Por exemplo, na área de informática, é comum o padrão:

Clique em [variável 1] para [variável 2].

Neste padrão, os elementos *Clique em* e *para* são fixos, ao passo que [variável 1] pode se referir a *Abrir* e *Fechar* e [variável 2] a *selecionar um documento de trabalho* e *sair do programa*. Essa relação recorrente entre os elementos da frase compõem um tipo de cadeia de

caracteres que, por critérios pragmáticos, assume valor especializado em determinado domínio.

Conforme Bevilacqua (1999) argumenta, os autores da primeira tendência também definem UFE como combinações sintagmáticas; a principal diferença está em uma restrição maior quanto aos padrões sintagmáticos que caracterizam as UFE no primeiro grupo, os quais se restringem à combinação de duas unidades. As estruturas sintáticas propostas para o segundo grupo em geral correspondem a sintagmas nominais, verbais ou preposicionais – embora isso não seja consensual entre os autores. Nessa tendência, a pesquisadora aponta para trabalho, do qual partilha autoria, que atribui às UFE um caráter eventivo e significado relacional.

Destas tendências, a última é distintiva em termos de abrangência, visto que as duas primeiras situam suas análises no nível do sintagma. A esse respeito, embora os estudos dessa tendência sejam eficazes para identificar estruturas típicas de um domínio, esse tipo de formulação prototípica geralmente não é registrado em glossários terminológicos de empresas de tecnologia que empregam metodologia de localização, apesar de se observar casos de valência – como no exemplo de matriz apresentado acima, em que está representada a regência do verbo (*clicar em*), bem como alguns aspectos que indicam a natureza dos elementos variáveis (item de UI com o qual o usuário pode interagir e ação resultante, respectivamente).

A definição adotada neste trabalho para UFE toma por base aquela encontrada em Lorente, Bevilacqua e Estopa (2002) e expandida por Bevilacqua (2004), a qual caracteriza a UFE como unidade de significação especializada sintagmática que é formada por um núcleo terminológico (NT), o qual pode ser uma UT simples ou sintagmática, e por um núcleo eventivo (NE), que pode ser um verbo, um nome derivado de verbo ou um particípio derivado de verbo. Nessa definição, a UFE apresenta as seguintes características:

- representa as atividades e os processos específicos de um âmbito específico, logo, depende de uma área temática;
- tem determinado grau de fixação interna;
- apresenta uma presença recorrente nos textos de um âmbito especializado.

Na observação de Bevilacqua (2004), essa definição contrasta com as propostas de reconhecimento e descrição feitas para essas unidades, as quais se concentram principalmente na UT. Nesta, os núcleos compõem uma estrutura profunda e operam como dois polos, em função das relações semânticas que um estabelece com o outro. Dessa estrutura, é possível

derivar três estruturas superficiais básicas de caráter sintagmático²⁵. A seguinte tabela apresenta as três estruturas acompanhadas da notação usada pela autora:

Tabela 3 - Estruturas fraseológicas superficiais

Tipo de estrutura	Representação formal
Verbal	[[NE] _v + [NT] _N]
Nominal (nome derivado de verbo)	[[NE] _{Ndev} + [NT] _{SP}]
Adjetival (particípio derivado do verbo)	[[NT] _N + [NE] _{PartAdj}]

Fonte: Elaborado pelo autor com base em Bevilacqua (2004).

Essa proposta é semelhante ao modelo transformacional²⁶ (MT) proposto por Desmet (1995-1996) (segundo menção da autora). Assemelha-se também à Gramática Transformacional de Noam Chomsky (1965). Em linguística computacional e PLN, também há estudos que tratam desse tema; alguns tomam por base o modelo de Chomsky, como em Zhang e Patrick (2005), ao passo que outros dão como pressuposto tal conhecimento linguístico, como em Marton e Katz (2002) e Strzalkowski *et al.* (1999).

Bevilacqua (2004) demonstra que o grau de fixação ou a estabilidade dessas unidades ocorre mais por meio dessas relações do que pela fixação morfossintática observada em outras propostas. Os resultados obtidos pela autora corroboram para que este estudo adote essa definição de UFE, visto que ele trabalha com a hipótese de que parte do ruído gerado nas listas de validação pode ser explicada em função desse tipo de relação semântica, o qual decorreria da atividade tradutória.

3.2.3 Critérios de caracterização

Na literatura, são encontrados diversos critérios para caracterizar as UFE. Sob a perspectiva da TCT, muitos dos critérios observados por fraseologistas da língua comum passam a ser considerados em investigações sobre fraseologia especializada. Em geral, os autores apontam o caráter sintagmático das unidades como um critério evidente. Nesse sentido, qualquer

²⁵ Entre os trabalhos que fundamentam a proposta da autora está Faber e Mairal (1999), o qual apresenta o Modelo Lexemático Funcional, baseado na Gramática Funcional de Simon Dik e na Lexemática de Coseriu.

²⁶ Para fins de simplificação, este trabalho usará a denominação *modelo transformacional* e *MT* para se referir a essa propriedade.

unidade linguística que tenha duas unidades lexicais ou mais é passível de ser caracterizada como UF e, por extensão, UFE.

Para Bevilacqua (1999), é possível classificar os critérios delimitadores em linguísticos (especialmente sintáticos e semânticos), pragmáticos e quantitativos. Aguado de Cea (2007), outra autora que procura traçar um panorama sobre os critérios, emprega uma divisão semelhante. Em sua pesquisa, a autora agrupa os critérios em morfossintáticos, semânticos, pragmáticos, estatísticos (que se enquadrariam nos quantitativos de Bevilacqua) e conceituais.

A maior parte dos critérios usados, contudo, são de base linguística. Em sua tese de doutoramento, Bevilacqua (2004) apresenta uma síntese dos critérios mais recorrentes na literatura²⁷. Entre os critérios que a autora lista, apenas três não são de ordem linguística: os critérios de inclusão de UT e de uso em âmbito específico (pragmáticos) e o de frequência (estatístico). As subseções a seguir apresentam um resumo desses critérios, os quais serão considerados para análise neste trabalho.

3.2.3.1 Caráter sintagmático

Esta característica determina que faz parte da natureza das UFE serem formadas por mais de um elemento linguístico. Trata-se de uma condição intrínseca das UFE. Contudo, mesmo que seja uma propriedade necessária, esta não é exclusiva dessas unidades, visto que uma UT nominal também pode ser sintagmática.

Há várias tentativas de delimitação sintagmática das UFE segundo este aspecto, nenhuma das quais é exaustiva a ponto de cobrir todas as formações possíveis. Hausmann (1989) propõe os seguintes padrões sintáticos para caracterizá-las:

- N+N
- N+V
- N+Adj
- Adj+Adv
- V+Adv

Esses padrões são seguidos estreitamente pelo autor para identificar essas unidades. L'Homme e Bertrand (2000) apresentam uma padronização semelhante para as UFE:

²⁷ Na revisão dos critérios, a autora se baseia em Béjoint e Thoiron (1989), Blais (1993), Cohen (1992), Desmet (1995-1996), Gouadec (1994), Heid (1992, 1998), Kjaer (1990), L'Homme (1998), L'Homme e Bertrand (2000), Lorente, Bevilacqua e Estopà (2002), Martin (1992) e Roberts (1994-1995).

- N+N
- N+V
- N+Adj
- Adj+N
- V+N

Os resultados de sua análise, contudo, apontam para uma série de elementos implícitos a esses padrões, como a valência no caso dos verbos:

- V+(Prep) +N
- N+(Prep)+N

Nenhum trabalho, no entanto, se propõe a apresentar uma delimitação que chegue ao máximo de detalhamento possível, por exemplo, indicando a quantidade de elementos que podem ser inseridos entre as unidades lexicais que compõem a UFE.

A delimitação desse aspecto é bastante discutida em trabalhos sobre linguística computacional ou PLN, especialmente para a detecção de unidades sintagmáticas compostas, ou *multiword units/expressions* – para citar termos comuns a essas áreas para tratar do fenômeno. Uma definição comum que é usada para tratar desse tipo de unidade (que englobam UFE) é a de *palavras com espaços*²⁸, em uma acepção de *palavra* que normalmente está baseada em critérios puramente formais (em linhas gerais, mas não exaustivas, consiste em sequência de caracteres separados por: espaços, pontuação ou marcas de início ou fim de linha).

Em Estudos de Fraseologia, esta propriedade é partilhada entre as UT sintagmáticas e as UFE. Em função disso, alguns trabalhos procuram estabelecer limites entre UT e UFE, traçando características distintivas entre elas. Entre os autores que fazem essa distinção estão Blais (1993) e Gouadec (1994). Em geral, a diferença está na designação de cada um desses tipos de unidades especializadas; as UT designam objetos e conceitos, enquanto as UFE designam uma combinação de conceitos ou noções (Blais) ou expressam um conteúdo próprio de uma área (Gouadec). Apesar das tentativas de diferenciá-las, há certo consenso entre alguns autores (Blais, 1993; Pavel, 1993; Peasant, 1993) de que nem sempre é fácil estabelecer a distinção. Isso pode ser observado também em Gouadec, que propõe elementos intermediários entre as UT e as UFE.

Mesmo que exista essa dificuldade, a distinção pode ser importante para fins de processamento computacional, na medida em que determinaria o tipo de comportamento esperado para cada caso, já que a UT tende a apresentar um maior grau de compactação conceitual e

²⁸ Do inglês *words-with-spaces* (SAG, BALDWIN, *et al.*, 2002).

designa um objeto ou conceito, ao passo que a UFE aceita mais elementos em seu interior e designa uma ação ou processo.

3.2.3.2 Grau de fixação (ou estabilidade)

Essa característica é definida pela invariabilidade sintagmática e paradigmática da unidade. Em geral, é determinada por duas outras características comuns atribuídas às UFE: estabilidade sintática e estabilidade semântica. Quanto menos variações houver no interior da unidade, maior será seu grau de fixação.

Essa gradação apresenta uma escala imprecisa e genérica, que vai das unidades mais fixas às mais flexíveis. Dentro dessa escala, alguns autores buscam situar onde exatamente as UFE estão. Por exemplo, Desmet (1995-1996) caracteriza as UFE exclusivamente como semifixas. Uma definição mais abarcadora, como a de Bevilacqua (1996), as colocaria entre as mais fixas e as semifixas. A autora também considera como UFE expressões fixas, como *para efeitos desta lei*.

Essa característica está associada a uma metodologia de base estruturalista. Pavel (1993) apresenta o método mais comum para a identificação desse grau, o qual implica na comutação de um ou mais elementos de uma UFE; se a comutação for observada como possível, conclui-se que a unidade não apresenta o grau máximo de fixação. Por exemplo, para as unidades *prevenção contra intrusões* e *armazenamento de dados no corpus* usado neste trabalho (cf. *Corpus* na página 115), a primeira não comporta a inserção de nenhum elemento entre seus componentes; não seria possível dizer *prevenção máxima contra intrusões*, por exemplo. Já no segundo caso, é possível inserir elementos, como em *armazenamento externo de dados*.

Divisões semelhantes são encontradas também em PLN. Nessa área, aspectos morfológicos normalmente são considerados para definir essa característica. Sag *et al.* (2002) separaram as *multiword expressions* da língua inglesa entre expressões fixas, que incluem unidades totalmente imutáveis, como *in short* e *ad hoc*; expressões semifixas, que incluem por exemplo expressões não decomponíveis, como *chutar o balde*²⁹; expressões sintaticamente flexíveis, como *look up* (em que é possível a inserção de elementos entre *look* e *up*). A divisão, no en-

²⁹ Nesse caso, há estabilidade sintática, visto que outros elementos não seriam inseríveis dentro da expressão, há estabilidade semântica, já que sua nominalização ou transformação em voz passiva também não é comum, mas não há estabilidade morfológica, já que está prevista a flexão do verbo *chutar*.

tanto, classifica nomes compostos como expressões semifixas exclusivamente por critérios morfológicos³⁰.

3.2.3.3 Inclusão de (pelo menos) uma UT

Para muitos autores, a UT é o ponto de partida para identificar as UFE, é seu núcleo de valor especializado e o elemento que confere valor especializado à unidade. O critério é pertinente para a análise realizada neste trabalho, visto que os glossários e as bases terminológicas produzidas para localização geralmente não apresentam qualquer tipo de identificação para unidades fraseológicas, sejam elas gerais ou especializadas. Nesse sentido, o critério pode ser usado para distinguir as UFE das UF que tenham sido incluídas em glossário.

3.2.3.4 Semicomposicionalidade, composicionalidade e não-composicionalidade

Especialmente na fraseologia da língua geral, os autores costumam tratar do aspecto oposto, ou seja, da não-composicionalidade. A não-composicionalidade ocorre quando o significado de uma UF é dedutível apenas a partir do conjunto. A *composicionalidade*, por sua vez, se dá quando o sentido da combinação pode ser deduzido do significado de cada um dos elementos. Apesar de antagônicas, as características não são anuláveis, ou seja, é possível que uma expressão seja compreendida tanto por um sentido global dedutível quanto pela interpretação de suas partes individuais (p.ex.: *bater as botas*).

A fraseologia especializada traz ainda a noção de *semicomposicionalidade*. Esta ocorre quando um dos elementos, o coocorrente, adquire um sentido novo, embora ainda aluda a aspectos característicos (p.ex.: *banco* em *banco de dados*). Essa noção está fortemente enraizada em uma visão lexicográfica que procura agrupar unidades de significação segundo sua forma, classificando-as como homógrafas (homônimas ou polissêmicas). Tomando por base uma noção saussuriana de signo, não caberia esse tipo de agrupamento; cada unidade de significação existe por si só, e as formas se repetem. Em geral, unidades com forma idêntica parti-

³⁰ Por particularidades do idioma inglês, na visão do autor, não seria necessário investigar mais a fundo a característica de inserção em sintagmas nominais compostos, que se limitaria a marcação de plural ou inserção de artigo definido.

lham traços conceituais da mais básica (pressupondo-se que seja possível determinar qual seria a mais básica). O próprio caso de *banco de dados* mostra isso. Um banco de dados consiste em um programa de computador que armazena dados. Disso partilha de alguns traços de banco como instituição financeira, mas trata-se de um signo totalmente diferente.

Para fins de processamento computacional, a não-composicionalidade é um aspecto importante, visto que as UFE que não apresentam composicionalidade são geralmente fixas. É o caso da unidade inglesa *round robin*; isoladamente, poderia se referir a um pequeno pássaro de penugem avermelhada no peito (*robin*) e roliço (*round*), mas a maior parte de seu uso corrente se refere a algum tipo de petição pública (como abaixo-assinados) ou a competições de pontos corridos. Essa unidade ainda figura como termo, designando um tipo de mecanismo usado por servidores DNS para compartilhar e distribuir a demanda por recursos de rede.

3.2.3.5 Uso em âmbito específico

Essa característica identifica o uso de uma UFE como própria de um âmbito, ou seja, dependente de uma área temática. Herdada da perspectiva teórica em Terminologia, essa característica parte do componente comunicativo.

Na aplicação computacional investigada neste trabalho, talvez esse critério seja útil na medida em que implica em uma maior probabilidade de que as entradas terminológicas de um glossário específico encontrem correspondentes entre a língua de partida e a de chegada. Maior probabilidade, contudo, não significa garantia total. Na medida em que o computador basearia sua identificação na forma de uma unidade glossariada, não há garantias de que uma denominação seja usada em uma área específica de um jeito especializado em todas as ocorrências de sua forma. Logo, a forma de uma unidade no glossário pode se referir a uma unidade de uso geral ou ainda a outra UFE.

Ao considerar o porte das empresas que empregam metodologias de localização, que costumam atuar em áreas diversas (mesmo que dentro de um mesmo domínio), observa-se que não é raro terminologias de áreas distintas sejam armazenadas na mesma base terminológica sem qualquer tipo de indicação sobre o âmbito de cada unidade. Enfim, trata-se de um critério de análise aplicável computacionalmente, mas que depende de uma boa manutenção das bases terminológicas usadas.

3.2.3.6 Presença recorrente

Essa característica tem por função ajudar a identificar determinada unidade como própria de um âmbito. Quanto maior for a frequência de ocorrência da unidade, maior será essa probabilidade.

A fundamentação teórica adotada para o trabalho influi na aplicação dessa característica. Por exemplo, L'Homme e Bertrand (2000) desconsideram a variação denominativa (ou *sinonímia*, na noção relativamente equivalente aplicada naquele estudo) para contabilizar a frequência de coocorrentes a fim de identificar novas colocações léxicas especializadas com base em colocações já catalogadas. Ao aplicar a frequência em sua análise das UFE eventivas, Bevilacqua (2004) aplica um modelo que agrupa três estruturas superficiais (verbo, nome verbal e particípio) como derivadas de uma mesma estrutura subjacente (conceitual). Com isso, todas as manifestações superficiais seriam contabilizadas como ocorrências da mesma UFE eventiva. Apesar das diferenças associadas ao emprego metodológico que é dado à característica, o critério raramente se basta por si só, conforme aponta L'Homme (2000).

Para os propósitos deste trabalho, acredita-se que a aplicação de tal critério, por exemplo, na identificação de falsos positivos nas listas geradas pelo *software* de QA pode ajudar a nortear o desenvolvimento de novas aplicações computacionais dessa natureza, indicando que tipos de fenômenos são os possíveis maiores causadores de erros. Além disso, trata-se de um critério que não se aplica apenas às UFE, mas também aos outros fenômenos investigados no trabalho (como variação e as UT).

3.2.4 Síntese

Para a investigação proposta neste trabalho, será adotada a definição de Bevilacqua (2004) sobre o fenômeno da fraseologia especializada. Essa definição caracteriza as UFE como unidades sintagmáticas de significação especializada que são formadas por dois núcleos (um terminológico e outro eventivo), representam as atividades e os processos específicos de um âmbito específico, têm determinado grau de fixação interna e apresentam uma presença recorrente nos textos de um âmbito especializado. Além disso, seus núcleos compõem uma estrutura profunda, o que faz com que estas se manifestem superficialmente como unidades verbais, nominais (derivadas de verbo) e adjetivas.

Um dos motivos para adotar a perspectiva da autora está na conceituação de que as UFE estão subordinadas a uma estrutura básica. Essa abordagem está alinhada com a concepção que alguns estudos sobre tradução, como Hurtado Albir (2001), têm a respeito dos procedimentos utilizados pelo tradutor ao fazer seu trabalho. Isso é observável na técnica de transposição (cf. *Técnicas de tradução* na página 109), originalmente introduzida por Vinay e Darbelnet (1995, p. 55), em que a classe gramatical de determinada unidade do original é trocada na tradução³¹.

Com relação às características das UFE e sua aplicação aos propósitos deste trabalho, a seguinte tabela apresenta uma síntese, com comentários a respeito de sua aplicabilidade para esta pesquisa:

³¹ De fato, como foi observado anteriormente (p. 47), essa característica está atrelada a um aspecto mais fundamental e não é exclusivo da Fraseologia.

Tabela 4 - Aplicabilidade das características fraseológicas

Característica	Aplicabilidade
Caráter sintagmático	Critério mais básico, aplicável na medida em que qualquer unidade composta tem potencial de ser UFE.
Grau de fixação	Embora seja indicativo da presença de alterações formais nas UFE, talvez seja melhor segmentar o critério, fazendo uso em seu lugar da estabilidade sintática, estabilidade semântica e estabilidade morfológica.
Inclusão de pelo menos uma UT	Embora seja mais relacionado à identificação e coleta, pode ser aplicado para identificar unidades não especializadas que tenham sido glossariadas.
Não-composicionalidade	Este aspecto pode ajudar a determinar a estabilidade das UFE, visto que as unidades não composicionais tendem a ser fixas do ponto de vista formal.
Uso em âmbito específico	Depende da gestão terminológica aplicada ao glossário. Pode ser útil, mas talvez dependa de outros aspectos também (não é porque uma denominação é usada em uma área específica de um jeito especializado que todas as suas ocorrências serão especializadas).
Presença recorrente	Neste caso, a relevância está em determinar as características mais recorrentes na ocorrência de falsos positivos.

Fonte: Elaborado pelo autor.

3.3 Variação

A variação, como aponta Freixa (2002), é um dos problemas terminológicos que o tradutor enfrenta na tradução especializada. Segundo a autora, o uso real dos termos não raro contradiz as afirmações contrárias à variação – as quais se abrigam nas metodologias de caráter prescritivo, imperantes na esteira da TGT e, por extensão, da localização, que propõem que se evite a sinonímia e a polissemia com objetivos diversos, como consolidar marcas comerciais (por exemplo, mediante a promoção de acesso mais rápido à aquisição de conhecimento) e facilitar a comunicação entre colaboradores e especialistas.

Talvez não seja exagero dizer que a variação tenha relação com um dos tópicos mais discutidos e polemizados nos Estudos de Terminologia e que este seja um dos primeiros fenômenos a despertar o interesse por propostas alternativas à TGT. O fenômeno, ainda observado por uma visão lexicográfica mais tradicional, aponta para uma falta de consenso entre especialistas – por exemplo, na determinação de critérios como homonímia e polissemia (ancorado na problemática noção de *palavra*).

No estudo de língua de especialidade, fazendo oposição entre os trabalhos desenvolvidos com base na TGT e os trabalhos desenvolvidos sob outras vertentes, como a TCT, o fenômeno é discutido no âmbito da *biunivocidade*, segundo a qual cada conceito é atribuído a uma única denominação (WÜSTER, 1998, p. 182 e SUÁREZ, 2004, p. 9). O fenômeno é reconhecido, mas tratado de maneiras distintas. Na TGT, de perspectiva essencialmente prescritivista, a biunivocidade é tida como condição ideal: a visão é de que os conceitos não sofrem variação, por exemplo, entre idiomas ou entre comunidades de especialistas, e o objetivo é minimizar a ocorrência de variantes denominativas, ou evitar o uso denominações (quase-) sinônimas, para facilitar a comunicação internacional. Na TCT, que conta também com estudos de caráter descritivo, trabalhos³² demonstram que na prática a variação (tanto conceitual quanto denominativa) ocorre por ser um fenômeno natural na língua, tanto em âmbito geral quanto nos de especialidade. Isso vai na contramão da consistência terminológica que é amplamente propagandeada na localização, consistindo em um dos argumentos de venda mais recorrentes e demonstrando uma fundamentação baseada na TGT.

Ao introduzir o tema da variação em Terminologia, Freixa (2002, p. 53) cita Jean Corbeil (1988, p. 57), autor que, ao traçar um panorama da Terminologia na França, apontava

³² Por exemplo, Freixa (2005) e Suárez (2004) supracitados.

para duas vias distintas para a variação: por um lado, observava que várias denominações pareciam corresponder mais ou menos à mesma noção e, em contrapartida, uma mesma denominação correspondia a várias noções (total ou parcialmente) diferentes. Dessa observação, Freixa faz sua proposta para a análise da variação terminológica, assentada sobre dois grandes aspectos: a *variação denominativa* e a *variação conceitual*.

Em linhas gerais, a variação denominativa é definida como um fenômeno no qual várias denominações correspondem à mesma noção ou – melhor dizendo – a uma noção equiparável em determinados níveis de análise. Por exemplo:

Por exemplo, um programa shareware ou freeware que você tenha **obtido por download** poderá utilizar adware para manter o seu preço baixo.

Quando você **fizer download** de um arquivo, o Norton AntiVirus analisa o arquivo com base nos detalhes de reputação.

No par de frases acima, existem as UFE *fazer download* e *obter por download*. É possível considerá-las variantes denominativas na medida em que ambas estão inseridas na mesma situação comunicativa e fazem parte da mesma estrutura conceitual. Há uma diferença de perspectiva, visto que, na primeira frase, o foco está no resultado da ação, em que o usuário (indicado por *você*) obtém algo mediante um download. Já na segunda, o foco está na ação em si, de o usuário fazer ou executar o download. Em termos de estruturação sintagmática, na primeira a UFE é complemento de um sintagma nominal e, na segunda, é núcleo de um sintagma verbal, mas seria igualmente possível usar “shareware ou freeware do qual você tenha feito download”.

Esse exemplo serve para demonstrar que, nesse entendimento sobre variação denominativa, é possível determinar que duas unidades sejam variantes sem se comprometer com a ideia de que sejam sinônimos perfeitos. Assim, a noção de variação denominativa, como aponta Suárez (2004), tem vínculo direto com a de sinonímia – tema de investigação recorrente nos Estudos Linguísticos, com destaque para a Lexicografia. Dentro da linha da Lexicografia, tanto Freixa quanto Suárez apontam para uma tendência de diferenciar entre variantes e sinônimos. Nessa linha, alguns autores usam critérios lexicológicos para definir que variações são consideradas sinônimos e critérios morfossintáticos para determinar aquelas que são consideradas variantes.

Com relação a essa oposição entre variantes e sinônimos, ambas as autoras concordam que não há na literatura argumentos e critérios suficientes para fazer tal distinção. Este trabalho usará a denominação *variante* para se referir ao fenômeno.

Esse tipo de variação tem relação direta com outros objetos de estudo deste trabalho, como as UFE. A estabilidade sintática ou semântica de uma unidade, por exemplo, pode ser definida pela presença de uma variante em textos especializados. É o que se vê nos dois exemplos a seguir:

Os itens removidos são **colocados na quarentena**, podendo ser restaurados a qualquer momento através do Histórico de segurança.

Remove o risco à segurança do computador e o **coloca em quarentena**.

No caso da UFE acima, é possível observar sua estabilidade sintática pela inclusão do artigo *a* (em *na quarentena*) na primeira frase. Ao mesmo tempo, os dois exemplos também são explicáveis via variação; ao contrastar as duas unidades, verifica-se um caso de variação morfossintática.

A análise da variação também pode ser aplicada a unidades menores ou a outros tipos de unidades, como as UT. Por ter relação direta com o nível formal das unidades, a variação denominativa influi diretamente no objeto de análise deste trabalho. Em função disso, os tipos de variação denominativa e as causas apontadas para ela serão apresentadas em mais detalhes a seguir.

A variação conceitual, por sua vez, não é definida pela contraparte da denominativa. Ou seja, em vez de estabelecê-la como um fenômeno em que várias noções estão vinculadas a uma mesma denominação, Freixa move a análise para a definição dos conceitos. Nesse caso, esse tipo de variação ocorre dentro de um mesmo conceito e não está vinculado à denominação. Logo, a análise deve observar um agrupamento de vários conceitos que compartilham de vários aspectos, mas que divergem em outros.

A título de ilustração, este trabalho trata de várias acepções para o objeto de análise da FLE (cf. *Denominações* na revisão teórica sobre Fraseologia, na página 44). O objeto é tratado como unidade polilexical que inclui pelo menos um termo, mas as noções divergem quanto a seus limites (colocação binária, unidade sintagmática ou frase inteira). A denominação associada a cada uma dessas noções é muito variada, embora às vezes seja partilhada – caso do termo *unidade fraseológica*. Nesse sentido, observa-se que o estudo da variação conceitual

parte do pressuposto de que existe algo de comum entre todas as acepções possíveis ou apresentadas, algo que permite estudá-las em conjunto e que não depende da denominação usada por cada autor.

Como a análise proposta neste trabalho tem forte relação com o caráter formal das unidades investigadas, a revisão teórica sobre esse tópico se firmará basicamente na variação denominativa. Apesar das particularidades das investigações feitas neste trabalho, somadas ao caráter e às premissas sobre equivalência na tradução de textos e sobre Terminologia na localização (mais informações na revisão teórica sobre tradução, na página 96), não se descarta que análises sobre variação conceitual cheguem a resultados robustos e projetem novas perspectivas sobre essas noções nessa área industrial.

Suárez (2004) faz um levantamento sobre estudos que abordam a variação denominativa, apontando dois ângulos de observação. De um lado, há estudos como o de Freixa (2002), os quais, além de dar visibilidade ao fenômeno, demonstram como a variação denominativa se manifesta em um texto especializado. O outro ângulo³³ demonstra que os textos especializados recorrem naturalmente ao uso de variantes com o objetivo de evitar a redundância no texto, contrastando diretamente com o caráter prescritivista dos textos produzidos para fins de localização, os quais em geral buscam essa redundância.

3.3.1 Causas da variação denominativa

Freixa (2002, p. 124-161) discorre a respeito das possíveis causas para a variação denominativa em textos especializados. O assunto não havia sido muito esmiuçado até então, especialmente em função do objetivo de trabalhos anteriores sobre Terminologia de perspectiva prescritivista, os quais buscavam uma comunicação biunívoca.

Em seu levantamento, além de apresentar causas prévias para a variação, como a redundância linguística, a arbitrariedade do signo linguístico e a possibilidade de ocorrer variação nas línguas (FREIXA, 2002, p. 125) a autora agrupa as causas em cinco grandes grupos, subdivididos em tipos secundários. Segue uma síntese dessas causas, bem como dos respectivos subtipos:

³³ A autora cita os trabalhos de Myers (1991) e Fernández Polo (1999).

3.3.1.1 Causas dialetais

3.3.1.1.1 Variação geográfica

Própria de línguas faladas em diversos países e regiões. Freixa aponta esse subtipo de variação como sendo expressamente reconhecido pela TGT. Em função disso, essa afirmação encontra eco na área de localização. Em geral, cientes (em algum nível) das diferenças entre as variantes de um mesmo idioma, empresas globais geralmente optam por uma dessas variantes para padronizar sua produção textual e evitar a ocorrência desse tipo de variação. Para empresas com escopo de atuação maior, a opção geralmente é pelo inglês americano, mas empresas com atuação basicamente no mercado europeu costumam optar pela variante britânica.

Exemplos: *analise* e *anализe*³⁴ para alterações gráficas, *lorry* e *truck*³⁵ e *current account* e *checking account*³⁶ para alterações mono e polilexicais (os exemplos estão de acordo com a tipologia apresentada pela autora [2002, p. 281-282]).

3.3.1.1.2 Variação cronológica

Esse subtipo de variação apresenta uma complexidade particular por ter relação com a variação conceitual. Pode ser provocada pela evolução do conhecimento, já que em dado momento pode haver um período de convivência entre um termo antigo e outro atual, mas também pelo fato de um referencial ser novo e apresentar uma falta de estabilização do léxico relacionado. O motivo para tal ausência de estabilidade, conforme aponta a TGT, pode ser a falta de conhecimento da terminologia correta por parte de profissionais ou tradutores ou a existência de ramificações em nível profissional e universitário (FELBER e PICHT, 1984, p. 215), ou ainda, como resume Freixa (2002, p. 132), falta de coordenação entre especialistas ou órgãos de normalização.

³⁴ Denominação para *anализar* em inglês britânico e americano, respectivamente.

³⁵ Denominação usual para *caminhão* em inglês britânico e americano, respectivamente.

³⁶ Termos usados para *conta corrente* em inglês britânico e americano, respectivamente.

Exemplo: *sistema digestivo* e *sistema digestório*, em que o último foi imposto pelo órgão de normalização do campo da Anatomia no Brasil.

3.3.1.1.3 Variação social

Na comunicação especializada, se refere às variantes usadas por diferentes grupos de especialistas. Em uma perspectiva socioterminológica, isso implica que a diversidade denominativa é causada por condições de produção, práticas sociais, restrições de enunciação, fatores históricos, falantes, divisão de trabalho, concorrência comercial, entre outros (GAMBIER 1991, p. 47 *apud* FREIXA, 2002, p. 133).

Em uma empresa global que aplica metodologias de localização, grupos de profissionais se encontram distribuídos em diversas localidades ao redor do mundo e fazem uso instrumental de um idioma³⁷. Contudo, em sua interação cotidiana, cada grupo individual exerce suas atividades tanto no idioma local quanto no instrumental. Dentro dos grupos individuais, é possível estratificar os colaboradores, por exemplo, por seu nível de domínio do idioma instrumental (desenvolvedor de *software* versus redator técnico) ou pelo nível de conhecimento sobre o âmbito do produto localizado (especialista em finanças versus desenvolvedor de *software* e redator técnico, para um *software* de finanças). Para exemplificar, termos de domínio sobre o idioma, essas diferenças podem resultar em variação denominativa – como no uso do par *interest* / *interesse* (referindo-se a *juros bancários* em português) ou *principal* / *principal* (referindo-se a *capital financeiro* em português).

Isso também pode ser levado para uma escala maior, nas diferenças entre empresas ou ainda entre grupos da mesma empresa situados em países distintos, dependendo do nível de análise que se quer chegar.

³⁷ Normalmente a língua inglesa.

3.3.1.2 Causas funcionais

3.3.1.2.1 Adequação ao registro linguístico

Este subtipo decorre do grau de formalidade da situação comunicativa, o qual é determinado especialmente pela relação que os interlocutores mantêm. Freixa afirma que a formalidade tende a oscilar menos em textos especializados.

3.3.1.2.2 Adequação ao grau de especialização

Este subtipo de variação tem sua origem na relação entre o emissor e o receptor de um texto. Dependendo do caso, um dado emissor pode incorrer reformulações para que o receptor compreenda determinado conceito. Freixa (2002, p. 144) aporta esse subtipo principalmente nos trabalhos de Ciapuscio (1993, 1997a, 1997b, 1998), os quais classificam essa reformulação em pelo menos três modalidades: *expansão* (por exemplo, por meio de paráfrases explanatórias); *redução* por omissão ou compactação da informação; e *variação* (nesse caso, em nível de *performatividade* e de *léxico*) (CIAPUSCIO, 1993, p. 113-114).

Nessa esteira, Ciapuscio (2003, p. 95-104), em sua proposta de tipologia textual, apresenta um nível, o situacional, que cobre a relação entre diversos graus de especialização para a interlocução (no eixo emissor-receptor). Assim, a variação denominativa estaria condicionada a aspectos como: a comunicação interna ou externa à área ou disciplina especializada; a relação entre os interlocutores (especialista-especialista, especialista-semileigo, especialista-leigo, semileigo-semileigo, semileigo-leigo) e a (as)simetria dessa relação; o número de interlocutores (monólogo, diálogo, grupo pequeno, grupo numeroso) e parâmetros espaço-temporais (comunicação pessoal, gráfica, televisiva, virtual, entre outras).

Essa relação pode ter influência na análise de variação em textos produzidos para localização, na medida em que o tipo de comunicação também pode variar. Embora a terminologia nesse caso seja específica de determinada área, isso não significa que seu uso esteja restrito a um público especializado. Por exemplo, um *software* como o Microsoft Word serve a um público que não pode ser definido facilmente, mas tem sua terminologia própria, que está acessível a seus usuários, como termos de tipografia (fonte, tabela, parágrafo e quebra de li-

nha). Outro exemplo semelhante é um *software* antivírus como o Symantec Norton AntiVirus, que traz termos específicos da área de segurança de dados eletrônicos (como vírus, infecção, *firewall* e quarentena), mas está igualmente voltado a um público heterogêneo. Nesses dois exemplos, de maneira geral e simplificada, a interlocução vai de um emissor especialista em *software* antivírus (pelo menos em comparação com o usuário final) para um receptor leigo ou semileigo.

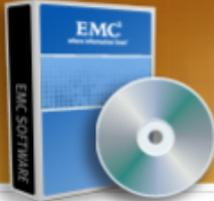
3.3.1.3 Causas discursivas

3.3.1.3.1 Evitar a repetição

Esta causa advém do emprego de recursos estilísticos. A questão pode ser colocada da seguinte maneira: um texto com repetição excessiva de determinada denominação pode resultar mais preciso e menos ambíguo, mas, em contrapartida, também será mais entediante e terá um estilo mais pobre.

Metodologias aportadas na localização não raro defendem o contrário: o uso de repetição não só é desejável, mas também é condição essencial para satisfazer o princípio da internacionalização, o qual demanda um texto simplificado que traga o mínimo possível de ambiguidade, bem como o princípio do reuso, o qual requer a eliminação, até onde for possível, de relações anafóricas que extrapolem uma frase individual. Por exemplo, nas seguintes páginas da web, sobre o mesmo produto (cf. partes em destaque abaixo):

Figura 6 - Página de produto, versão em inglês



CLARiiON Host Encryption

Host-based encryption for scalable, cost-effective data security

OVERVIEW **RESOURCES & BENEFITS** **RELATED PRODUCTS**

EMC CLARiiON Host Encryption delivers encryption to protect your information from unauthorized access or inadvertent loss. CLARiiON Host Encryption allows you to add a high level of protection to enable compliance with internal, private, and government standards.

As a host-based encryption product, CLARiiON Host Encryption lets you choose the logical unit numbers (LUNs) or volumes which contain sensitive data and need to be encrypted—no need to encrypt the entire environment or array, thereby minimizing management. **CLARiiON Host Encryption incorporates flexible, easy-to-use, centralized enterprise key management.**

Descriptive & Technical Info

 [Data sheet](#)

[View all Related Materials](#) 

Fonte: EMC (2008a).

Figura 7 - Página de produto, versão em português



CLARiiON Host Encryption

Criptografia baseada em host para segurança econômica e dimensionável dos dados

VISÃO GERAL **RECURSOS E BENEFÍCIOS** **PRODUTOS RELACIONADOS**

O EMC CLARiiON Host Encryption oferece criptografia para proteger suas informações contra o acesso não autorizado ou a perda acidental. O CLARiiON Host Encryption permite adicionar um alto nível de proteção para garantir conformidade com normas internas, privadas e governamentais.

Como é um produto de criptografia baseada em host, o CLARiiON Host Encryption permite que você escolha LUNs (Logical Unit Numbers, números de unidade lógica) ou volumes que contenham dados confidenciais e precisem ser criptografados — sem a necessidade de criptografar todo o ambiente ou array, consequentemente reduzindo o gerenciamento. **O CLARiiON Host Encryption incorpora o gerenciamento corporativo de chaves, um recurso centralizado, flexível e fácil de usar.**

Informações descritivas e técnicas

 [Data sheet](#)

[Exibir todo o material relacionado](#) 

Fonte: EMC (2008b).

Nesse caso, por exemplo, na última frase do primeiro parágrafo, seria possível usar um pronome ou outro tipo de anáfora (uma variante lexical, um hiperônimo, como *o/esse produto*) no lugar de *CLARiiON Host Encryption*, o que não ocorre. Contudo, a segunda sentença aparece repetida em outro texto, em contexto diferente (cf. partes em destaque abaixo):

Figura 8 - Uso de segmento de tradução em contexto diferente, em inglês

Table 2: CLARiiON Host Encryption key use cases

Use cases	Business/operational benefits
Security audit compliance and regulation	Deploy technologies to support enterprise security strategy Align with internal policies and external/governmental regulations
Shared services providers with consolidated storage	Protect data from unauthorized access and maintain confidentiality
Secure business continuity	Integrate encryption non-disruptively into an existing business continuity strategy Prevent the compromise of replicated media
Secure disk migration/rotation/upgrade	Prevent the compromise of transported data Protect data following its physical removal from a customer site Prevent the compromise of discarded media

CLARiiON Host Encryption incorporates flexible, easy-to-use, centralized enterprise key management, and enables consistent encryption methodology in heterogeneous environments.

Fonte: EMC (2010a).

Figura 9 - Uso de segmento de tradução em contexto diferente, em português**Tabela 2: Principais casos de uso do CLARiiON Host Encryption**

Casos de uso	Benefícios para os negócios/operações
Conformidade e normas de auditoria de segurança	Implementar tecnologias para dar suporte à estratégia de segurança corporativa Alinhar-se a políticas internas e normas externas/governamentais
Provedores de serviços compartilhados com armazenamento consolidado	Proteger os dados contra acesso não autorizado e manter a confidencialidade
Proteção da continuidade dos negócios	Integrar a criptografia sem causar interrupções em uma estratégia preexistente de continuidade dos negócios Impedir o comprometimento de mídias replicadas
Proteção de migração/rotação/upgrade de discos	Impedir o comprometimento de dados transportados Proteger os dados após sua remoção física do local de um cliente Impedir o comprometimento de mídias descartadas

O CLARiiON Host Encryption incorpora o gerenciamento centralizado e corporativo de chaves — fácil de usar e flexível — e viabiliza a metodologia de criptografia consistente em ambientes heterogêneos.

Fonte: EMC (2010b).

Aqui, não é possível adicionar uma referência anafórica, em função da posição que o segmento de texto ocupa no material. A questão chega a atingir outra dimensão quando se vê o seguinte³⁸ (cf. partes em destaque abaixo):

³⁸ Captura de tela retirada do *site* emc.com (EMC, 2013).

Figura 10 - Reaproveitamento de parte de segmento de tradução

Table 2. VNX Host Encryption Key Use Cases

Use cases	Business/Operational benefits
Security audit compliance and regulation	<p>Deploy technologies to support enterprise security strategy</p> <p>Align with internal policies and external government regulations</p>
Service providers with consolidated storage	Protect data from unauthorized access and maintain confidentiality
Secure business continuity	<p>Integrate encryption non-disruptively into an existing business continuity strategy</p> <p>Prevent replicated media from being compromised</p>
Secure disk migration/rotation/upgrade	<p>Prevent transported data from being compromised</p> <p>Protect data following its physical removal from a customer site</p> <p>Prevent discarded media from being compromised</p>

VNX Host Encryption incorporates flexible, easy-to-use, centralized enterprise key management, and enables consistent encryption methodology in VNX storage environments.

Fonte: EMC (2013).

Na figura acima, constata-se ainda que praticamente todo o segmento de texto se repete, com a exceção do nome, que muda de *CLARiiON Host Encryption* para *VNX Host Encryption*. Nesse sentido, não se poderia descartar que, em localização, um segmento de texto seja reaproveitado em casos que a natureza do agente mude. No exemplo apresentado, ambos se tratam de nomes de produto, mas poderia se tratar de um produto no primeiro caso e de uma solução no segundo, o que poderia levar a incongruências ou a problemas no aproveitamento desses segmentos. Sendo assim, como *a priori* um redator não consegue definir quando e onde determinada frase será usada no futuro, o princípio em geral é aplicado a todos os segmentos de texto possíveis.

3.3.1.3.2 Economia linguística

Aqui, o princípio está nos mecanismos de redução (p. 79); quando determinada denominação é muito repetitiva ao longo de um texto ou quando esta é muito extensa, empregam-se siglas ou uma versão reduzida da unidade, por exemplo, no caso de uma UT sintagmática. É o caso do seguinte par de tradução:

In the Facebook login **Web page**, log in to your Facebook profile.

Na **página** de login do Facebook, faça login no seu perfil do Facebook.

Aqui, o termo *página da Web* é reduzido a *página*. Outros elementos na frase permitem que o termo seja reduzido sem prejuízos à compreensão, como, por exemplo, o conhecimento pressuposto de que *Facebook* é um tipo de serviço acessado via internet. Igualmente determinante para isso é que o termo faz parte do sintagma nominal *página de login*, o que pode pressupor um acesso via web também.

3.3.1.3.3 Criatividade, ênfase e expressividade

Para este subtipo, a autora coloca a criação de denominações novas, a qual seria mais usual em textos não especializados, mas que também é usada em textos especializados, por exemplo, para dar ênfase ou originalidade.

É o caso de termos como *data deduplication*, *dedupe*, *commonality factoring* e *data coalescence*. Os quatro designam o processo de eliminar dados digitais redundantes (ou duplicados) para reduzir o espaço de armazenamento em disco para computadores (normalmente aplicável a servidores).

A justificativa para tantas designações diferentes está na concorrência de mercado, neste caso, do mercado de infraestrutura de armazenamento de dados digitais. Avamar (hoje parte da EMC), NetApp e Rocksoft (hoje parte da Quantum) eram algumas das empresas que competiam por essa fatia. No final, o termo *deduplication* (bem como a forma mais curta, *dedupe*) caiu nas graças comerciais, sendo adotado pelas demais empresas. Aqui, o uso de uma variante pode significar um desencontro de esforços em busca de inovação, visto que as

empresas são concorrentes e não divulgariam seus avanços antes de lançar algo no mercado, mas também uma tentativa de cada empresa de obter algum tipo de vantagem comercial por ser considerada a cunhadora do termo³⁹.

Seja como for, o exemplo ajuda a ilustrar que, na localização, esse tipo de causa pode ser mais frequente do que o suposto inicialmente.

3.3.1.4 Causas interlinguísticas

3.3.1.4.1 Coexistência entre termo local e empréstimo

Na interação entre idiomas diferentes (por exemplo, mediante tradução), os produtos resultantes podem apresentar variação entre um termo local e o termo original, ou empréstimo. Freixa (2002, p. 151) afirma que as causas para esse tipo de variação podem ser por proximidade geográfica e, com mais frequência, por proximidade cultural. Um especialista pode dar preferência à forma linguística de outro idioma por questão de prestígio ou para que a comunicação seja mais eficaz (por exemplo, caso determinada denominação já tenha sido veiculada internacionalmente). O idioma do termo mais fixado costuma corresponder ao da língua em que o conceito começou a ser veiculado.

Em áreas como o desenvolvimento de *software*, em que cada vez mais empresas globais optam pelo idioma inglês para a produção de seus produtos e respectiva documentação, a ocorrência de empréstimos é comum, por exemplo, em função de avanços tecnológicos (e, por extensão, da criação de novos conceitos). Por questões de estratégia comercial, não é raro que os conteúdos gerados por esse tipo de empresa sejam colocados em tradução antes que as próprias notícias sobre tal avanço atinjam um público especializado abrangente o suficiente para permitir a consolidação de uma forma equivalente em outras localidades.

Igualmente comum é a tentativa que os tradutores fazem de atribuir alguma forma equivalente no idioma de tradução, gerando assim uma concorrência entre as diferentes soluções adotadas por tradutores distintos e entre a forma no idioma original. Freixa (2002) coloca essa questão da seguinte maneira: a aparição de variantes e sinônimos resulta da necessidade

³⁹ Há na internet um fragmento de discussão sobre a denominação adotada para essas tecnologias de eliminação de dados eletrônicos duplicados. Observa-se especialmente no comentário principal uma tensão latente em ser o verdadeiro cunhador de um termo que caiu no gosto da clientela corporativa (NETAPP, 2009).

de denominar os conceitos novos de outros idiomas com rapidez e sob pontos de vista distintos, gerando dispersão denominativa na língua de chegada.

Na localização, é possível observar esse aspecto na terminologia de empresas diferentes. Por exemplo, para o termo *to download* em português brasileiro, a Microsoft lista o verbo *baixar*, ao passo que a Symantec usa *fazer o download*. Outro caso é *bandwidth throttling*, traduzido respectivamente por *limitação da largura de banda* e *regulagem da largura de banda*.

Estes são casos mais visíveis, visto que todos estão glossariados, mas esse tipo de variação é muito comum em termos gerais, até mesmo no contexto da terminologia de uma única empresa. É o que se vê nas traduções a seguir para o verbo *trigger*, todos retirados do *site* da Symantec:

These programs penetrate the Windows TCP/IP layer to send and receive data without **triggering** firewall alerts.

Esses programas penetram na camada TCP/IP do Windows para enviar e receber dados sem **acionar** alertas do firewall.

A flash file named today.swf was used to **trigger** the vulnerability in Internet Explorer.

Um arquivo Flash chamado today.swf foi usado para **disparar** a vulnerabilidade no Internet Explorer.

3.3.1.5 Causas cognitivas

3.3.1.5.1 Imprecisão conceitual

Essa causa relaciona a imprecisão conceitual diretamente à variação denominativa. Essa imprecisão pode estar relacionada, por exemplo, à inovação tecnológica, o que pode pressupor o surgimento de conceitos com fixação denominativa e conceitual baixa. A autora ainda afirma que especialistas de áreas diferentes podem supor que partilhem de alguns referentes,

mas os conceitos por trás de tais referentes não correspondem, por exemplo, por serem elaborados sobre pontos de vista diferentes.

3.3.1.5.2 Distância ideológica

A autora centra essa causa especialmente no desenvolvimento de conhecimento especializado mediante a formulação de teorias. Não é incomum que teóricos investiguem um mesmo fenômeno sob pontos de vista diferentes e cheguem a denominações diversas. Este trabalho pôde demonstrar isso, por exemplo, na revisão teórica sobre Fraseologia (p. 43-56). Por vezes, como a autora assinala, outros autores podem propor algum tipo de contestação ou contraproposta a determinada teoria vigente e, com isso, propor mudanças terminológicas e um vocabulário próprio para marcar as diferenças ou manter um alinhamento melhor com as novas concepções propostas (caso que pode ser observado na revisão teórica sobre Terminologia, p. 34-43).

Outro motivo apontado pela autora é evitar formas existentes que tenham conotação excessivamente negativa (um exemplo pode ser encontrado em discussões sobre erro em tradução, cf. p. 112).

Aproximando a questão para o âmbito comercial, também é possível encontrar variantes que concorrem por questões que poderiam ser identificadas como ideológicas (embora possam ser melhor definidas como mercadológicas ou dirigidas por *marketing*). É possível encontrar referências na internet de discussões e alegações acerca da autoria de determinado termo para definir tecnologias bem semelhantes. A tecnologia que vence o mercado normalmente passa a ter sua denominação adotada pela concorrência, como foi visto no caso de *deduplication* apresentado na página 69.

3.3.1.5.3 Diferenças na conceituação

No mesma perspectiva apresentada para a distância ideológica, Freixa aponta este subtipo como relacionado ao desenvolvimento de conhecimento especializado em nível superior. A causa principal está na perspectiva adotada por diferentes autores ao conceituar determinado referente. Conceituações diferentes, feitas sob disciplinas diferentes ou sob perspecti-

vas diferentes dentro de uma mesma disciplina, podem tratar do mesmo referente sob pontos de vista diversos. Isso resulta em sinonímia entre as denominações que decorrem de cada conceituação. Segundo a autora afirma, no entanto, essa relação não é consensual, visto que alguns autores a rechaçam, tratando-a como correferencial em vez de sinonímica.

3.3.2 Pressupostos para análise: autovariação e heterovariação

A probabilidade de cada uma dessas causas influir na variação passa por alguns pressupostos fundamentais. Para a autora, há dois tipos de variação denominativa: a *autovariação*, que ocorre quando um mesmo autor decide usar denominações diferentes para expressar uma mesma ideia ou conceito; e a *heterovariação*, que ocorre quando autores diferentes optam por expressar uma mesma ideia de maneiras diferentes. Nesse sentido, a motivação para usar uma variante é diferente quando se analisa o fenômeno em textos do mesmo autor ou de um autor diferente.

Posteriormente, Freixa (2005) apresenta outros pressupostos para a análise de *autovariação* versus *heterovariação*. No caso, a autora formula a hipótese de que a autovariação é diferente quando ocorre em um mesmo texto ou em textos distintos.

Esses pressupostos sugerem que há causas diferentes para a variação. A autora estabelece a seguinte relação em função das causas apontadas acima:

Tabela 5 - Relação entre causas e tipo de variação

Causa	Quando ocorre
Dialetal	Heterovariação
Funcional	Autovariação; somente entre textos diferentes
Interlinguística	Autovariação; no mesmo texto
Cognitiva	Autovariação/heterovariação (depende do subtipo analisado)

Fonte: Elaborado pelo autor com base em Freixa (2005).

Essa mudança facilita a escolha de critérios para análise. Contudo, observa-se que não há um caso que cubra a situação peculiar da localização; em função da metodologia de produ-

ção textual que é empregada, há uma diluição da autoria, visto que um texto pode ter vários autores, mas nenhum é apontado como tal (a empresa é autora e detentora dos direitos sobre o texto), isso somado a uma abordagem prescritivista à terminologia, que pretende evitar a variação.

Nesse cenário, embora seja possível fazer uma tentativa de enquadrar os textos produzidos, por exemplo, sob o pressuposto da autovariação (assim excluindo, por exemplo, as causas dialetais), é provável que esta não seja uma decisão metodológica muito acertada. Um motivo possível para isso é que, em um cenário corporativo de produção textual para localização, as atividades de redação podem ser realizadas nas mais diversas localidades geográficas, em continentes e realidades diferentes, muitas vezes por redatores que produzem textos em idioma estrangeiro, embora em geral existam instrumentos de controle estilístico (guias de redação) e terminológico (bases de dados terminológicas).

Observa-se nisso dois polos distintos: de um lado, uma diversidade geográfica, cultural e linguística que pode influir na produção do texto; por outro, a imposição de controles formais sobre essa produção. Na visão deste trabalho, essa oposição implica que não é possível seguir à risca a hipótese de Freixa sem incorrer em riscos.

3.3.3 Tipos de variação denominativa

Freixa (2001, p. 274-282) apresenta uma proposta de classificação tipológica para os casos de variação denominativa. A classificação toma por base parâmetros formais e semânticos, agrupando as mudanças em cinco tipos:

- 1 alteração gráfica;
- 2 alteração morfossintática;
- 3 redução;
- 4 alteração lexical;
- 5 diversas alterações complexas.

Segundo a abordagem metodológica aplicada a seu trabalho, a autora classifica as alterações em dois grupos: de um lado, as alterações simples (que correspondem aos níveis gráfico e morfossintático), e do outro, as alterações complexas (envolvendo o nível lexical e as reduções). A autora ainda promove uma segunda divisão para agrupar as variantes que envolvam mais de uma alteração complexa (tipo número 5). Além disso, seu trabalho tem um enfo-

que mais voltado a unidades nominais, embora tenha coletado alguns adjetivos e verbos (monolexicais) para análise. Isso reflete a tipologia apresentada pela autora, basicamente nominal.

A tabela a seguir detalha a tipologia apresentada pela autora, incluindo alguns exemplos listados por ela para cada caso:

Tabela 6 - Proposta de classificação formal

Ocorrência variacional	Tipo	Subtipo	Exemplo
I. Alteração gráfica	1. termo e forma artificial	a) termo e símbolo	coure / Cu
		b) termo e fórmula química	Amoníaco / NH ₃
		c) termo e outra fórmula artificial	llei de residus industrials de 1983 / llei 6/1983
	2. termo e abreviação	d) termo e sigla	clorofluorocarboni / CFC
		e) termo e abreviatura	acer inoxidable / acer inox.
	3. alteração ortográfica		esprai / spray
II. Alteração morfosintática	1. permanência da estrutura	a) alteração/presença de artigo	gestió de residus / gestió dels residus
		b) alteração de número	contaminació de l'aigua / contaminació de les aigües
		c) alteração de preposição	condicions del condensador / condicions en el condensador
	2. alteração na estrutura	d) alteração de gênero	màxima absoluta / màxim absolut
		e) [N+A] / [N+SP]	residus miners / residus de la mineria
		f) monolexical / polilexical	producte ecològic / ecoproducte
		g) outras alterações na estrutura	matèria primera / primera matèria
III. Redução	1. reduções na extensão	a) Tipo [N+X] = [N]	aigües residuals / aigües
		[N+A]= [N]	cicle de vida / cicle
		[N+SP] = [N]	semiconductor-electròlit / semiconductor
		[N ₁ +N ₂]= [N ₁]	
	b) Tipo [N+X+Y] = [N+X]	estalvi de recursos naturals / estalvi de recursos	

Ocorrência variacional	Tipo	Subtipo	Exemplo
		$[N+SP+A] = [N+SP]$	recollida selectiva dels residus / recollida selectiva
		$[N+A+SP] = [N+A]$	metall pesant tòxic / metall pesant
		$[N+A_1+A_2] = [N+A_1]$	temps de residència del gas / temps de residència
		$[N+SP_1+SP_2] = [N+SP_1]$	
		c) outras reduções na extensão	filtre de control de contaminació / filtre
	2. reduções na base	d) $[N+A] = [A]_N$	planta depuradora / depuradora
		e) $[N_1+N_2] = [N_2]$	gas CFC / CFC
		f) $[N_1+SP (de+N_2)] = [N_2]$	escuma de poliestirè / poliestirè
	3. outras reduções		emissió de compostos químics tòxics / emissió Tòxica
	1. unidades monolexicais		contaminació / pol·lució
		a) alteração na base	bé de consum / producte de consum
		$[N+SP (prep + art + N)]$	adob químico / fertilizant químico
		$[N+A]$	destruicó de la capa d'ozó / degradació de la capa d'ozó
	2. unidades polilexicais	outras alterações na base	
		b) alteração na extensão	dipòsit d'assentament / dipòsit de decantació
		$[N+SP (prep + art + N)]$	agricultura ecològica / agricultura biològica
		$[N+A]$	aigües residuals / aigües de la claveguera
		outras alterações na extensão	
IV. Alteração lexical			
V. Diversas alterações complexas	1. com semelhança formal	a) monolexical / polilexical	aliment / producte alimentari
		b) polilexical / polilexical	fang digerit / residu de digestió

Ocorrência variacional	Tipo	Subtipo	Exemplo
	2. sem semelhança formal	c) monolexical / polilexical	abocador / dipòsit de residus
		d) polilexical / polilexical	pesticida de síntesi / plaguicida químic

Fonte: Adaptado de Freixa (2001, p. 281-282).

Segue um resumo de cada um dos tipos apresentados pela autora.

3.3.3.1 I – Alteração gráfica

3.3.3.1.1 1 - Termo e forma artificial

Para a alteração entre termo e forma artificial, a autora agrupa a mudança entre uma unidade natural e uma unidade expressamente convencionada – no caso, uma UT e um símbolo, respectivamente. O paralelo entre as duas unidades deverá ocorrer sempre. Casos em que a unidade artificial não corresponde à UT são tratados pela autora em outra categoria (de alterações diversas). A autora cita o exemplo *C / matèria orgànica total e bé acidesa forta / SO₂*, em que, segundo ela, daria para se considerar como sinônimo contextual⁴⁰. Para esse grupo, a autora também considera as unidades que ocorrem dentro de unidades sintagmáticas, como no par de variantes *concentració de diòxid de sofre / concentració de SO₂*.

3.3.3.1.2 2 - Termo e abreviação

Este subtipo agrupa as mudanças entre uma unidade natural e a respectiva sigla ou abreviatura, como em *matèria orgànica total / MOT* e em *clorofluorocarboni / CFC* no primeiro caso, e em *acer inoxidable / acer inox* no segundo.

⁴⁰ Para exemplificar a tipologia, são apresentados os exemplos que a própria autora dá para cada caso.

3.3.3.1.3 3 - Alteração ortográfica

Em alteração ortográfica, a autora coloca as seguintes mudanças:

- a presença e ausência de hífen (*micro-organisme / microorganisme*);
- b presença e ausência de acento (*atmosfera / atmòsfera*);
- c “s” e “ss” (*dissolució / disolució*);
- d presença e ausência de espaço (*tetra brik / tetrabrik*);
- e “a” e “e” para representar vogal neutra (*claveguera / clavaguera*);
- f outros casos de variação ortográfica.

Algumas dessas mudanças não são necessariamente específicas de um idioma (a, b, d). Por exemplo, as mudanças a e d também são observadas na língua inglesa (*back-up / backup / back up*), e as mudanças a e b são encontradas em português também (*prerrequisito / pré-requisito*). As mudanças c, e são particulares a um idioma. Independentemente dessas particularidades, se a abordagem for mantida em um nível mais alto, será possível aplicá-la aos idiomas investigados neste trabalho.

3.3.3.2 II - Alteração morfossintática

3.3.3.2.1 1. Permanência da estrutura

- a presença e ausência de artigo (*gestió de residus / gestió dels residus; aigua de rentat / aigua del rentat; depuració de gasos / depuració dels gasos*);
- b alteração de número na expressão sintagmática (*contaminació de l'aigua / contaminació de les aigües; cobriment amb terra / cobriment amb terres; producció de gas / producció de gasos*);
- c alteração do gênero (*màxima absoluta / màxim absolut*);
- d alteração de preposição na estrutura [N+SP] (*condicions del condensador / condicions en el condensador*).

Das alterações apresentadas, observa-se que praticamente todos os casos são possíveis nos idiomas analisados neste trabalho, exceto pela mudança c no inglês, já que este idioma não apresenta flexões de gênero.

3.3.3.2.2 2. Alteração na estrutura

- a [N+A] e [N+SP] (*residus miners / residus de la mineria; aigües pluvials / aigües de pluja; contaminació hídrica / contaminació de les aigües*);
- b [N+A] e uma unidade monolexical em que o adjetivo da unidade sintagmática é representado por um prefixo ou forma prefixada na unidade monolexical (*tractament previ / pre-tractament; producte ecològic / ecoproducte*).

Os casos de alteração na estrutura podem ser encontrados nos dois idiomas analisados neste trabalho, embora a representação das estruturas seja diferente no caso do inglês ([A+N] ou [N+N] em vez de [N+A], como no par *security level / level of security*).

3.3.3.3 III - Redução

3.3.3.3.1 1. Reduções na extensão

- a redução de 2:1, em que a variante se constitui pela base da unidade (*abocador controlat / abocador; cicle de vida / cicle*);
- b redução de 3:2, em que a variante se constitui pela base mais a parte da extensão (sintagmática ou não) sintaticamente mais próxima deste (*estalvi de recursos naturals / estalvi de recursos*). Não há critério quanto à ligação da unidade omitida (pode estar relacionada à base ou ao segundo elemento);
- c reduções diversas, mas principalmente de mais de um elemento (*adob orgànic biològic / adob; planta compostatge de la brossa orgànica / planta de compostatge*).

A autora não traz muitos detalhes sobre as reduções de extensão. Afirma que são as mais frequentes e apresentam poucas particularidades. Classifica os subtipos de acordo com o número de constituintes pré e pós-redução. Em geral, tem caráter anafórico, exercendo uma clara função textual de retomada.

Assim como nos tipos morfossintáticos de alteração na estrutura referidos acima, os tipos apresentados aqui também encontram correspondência em português e inglês, embora essa afirmação também leve em conta as particularidades da língua inglesa para os casos

[N+A]= [N] e [N1+N2]= [N1], que seriam [A+N] ou [N+N]= [N] e [N1+N2]= [N2] (como em *risk level / level* ou *risk*).

3.3.3.3.2 2. Reduções na base

Normalmente, este tipo de redução ocorre quando a extensão da unidade tem um caráter mais distintivo que a base. Em geral, o motivo por trás da redução é lexical, e não anafórico (*planta depuradora / depuradora; gas CFC / CFC; escuma de poliestirè / poliestirè*).

3.3.3.3.3 3. Outras reduções

Aqui a autora agrupa casos variados, mas em especial quando a parte omitida é a extensão sintaticamente mais próxima da base (*captació d'aigua superficial / captació superficial; procés natural de degradació / procés de degradació*).

3.3.3.4 IV - Alteração lexical

Este é o tipo de variação mais comumente associado à sinonímia em teorias linguísticas. Os tipos propostos são apresentados a seguir:

3.3.3.4.1 1. Unidades monolexicais

Este subtipo agrupa as mudanças entre UT constituídas por apenas uma unidade e variantes igualmente unitárias (*contaminació / pol·lució*).

3.3.3.4.2 2. Unidades polilexicais

- a alteração na base, em que a extensão pode ser um adjetivo ou sintagma preposicional (*bé de consum / producte de consum*);
- b alteração na extensão, em que a extensão não necessariamente mantém a mesma estrutura sintagmática (*aigües residuals / aigües de la claveguera*).

3.3.3.5 V - Diversas alterações complexas

A autora criou este grupo para agrupar as variantes que apresentam mais de uma alteração complexa (lexical e/ou redução) em sua composição. Embora não deixe explícito, os subtipos apresentados aparentam criar uma separação entre as variantes que guardam alguma relação lexical entre seus componentes (o que a autora agrupa como *com parentesco formal*) e aquelas que apresentam mudança lexical total (*sem parentesco formal*).

3.3.3.5.1 1. Com parentesco formal

- a monolexical-polilexical (*agroecologia, agricultura ecológica; aliment, producte d'alimentació*);
- b polilexical-polilexical (*abocador de residus tòxics, abocador especial; aprofitament de residus sòlids urbans, valorització de residus*).

3.3.3.5.2 2. Sem parentesco formal

- a monolexical-polilexical (*abocador, dipòsit de residus; emissió, emanació gasosa*);
- b polilexical-polilexical (*acumulació de residus, amuntegament de deixalles; producció neta, reducció de la contaminació a l'origen*).

3.3.4 Síntese

Para a investigação proposta neste trabalho, será adotada a perspectiva de Freixa (2002) sobre o fenômeno da variação linguística, bem como sua classificação e tipologia, embora a denominação usada seja a de Suárez (*variante*).

Um dos motivos para a adesão está na abrangência que Freixa dá ao fenômeno da variação, com impacto direto em análises contrastivas entre traduções. Ou seja, tem influência direta na qualidade da parte computacional do QA terminológico, com o potencial de reduzir a ocorrência de falsos positivos. No texto de partida, ela aumenta o raio de identificação das USE, possibilitando a redução de silêncio. No texto de chegada, mesmo que trabalhos apontem para um percentual de apagamento da variação (SUÁREZ, 2004), mais de 70% das ocorrências indicam correspondência na tradução de variantes em algum grau; logo, pode resultar na redução do ruído.

Na contraparte, há aspectos que diferenciam os enfoques deste trabalho e os de Freixa, o que poderia justificar a escolha por outra base teórica – por exemplo, os idiomas tratados (catalão / português e inglês) e a abrangência das unidades lexicais (nominais⁴¹ / nominais, adjetivas e verbais). Mesmo assim, este trabalho considera que a tipologia apresentada seja aplicável, respeitando-se as particularidades de cada idioma (especialmente do inglês). Outro motivo que pode corroborar com essa adesão está na associação de outras metodologias, como a da Fraseologia, que podem facilitar a análise de outras unidades não cobertas, como as verbais.

Com relação às causas, segundo as particularidades gerais dos textos localizáveis, é possível assinalar de antemão que podem ser mais ou menos relevantes para este trabalho. A seguinte tabela apresenta as causas possivelmente aplicáveis ou não para este trabalho, com a probabilidade de ocorrência de cada uma, sugerida pelo autor em virtude das particularidades do *corpus* de estudo:

Tabela 7 - Causas para variação

Causa	Probabilidade de ocorrência
<i>Causas</i>	Nenhuma das causas dialetais apontadas pela autora se mostra aplicáveis

⁴¹ Embora ainda inclua algumas adjetivas e verbais monolexicais.

Causa	Probabilidade de ocorrência
<i>dialetais</i>	no âmbito da localização (mais detalhes a seguir, nas observações individuais para cada uma das subcausas).
Variação geográfica	Improvável. Em geral, empresas que trabalham com localização optam por uma variante de determinado idioma (em geral, o inglês dos EUA) para uniformizar seus textos. Como este trabalho analisa textos de uma única empresa, é pouco provável que essa causa apareça.
Variação cronológica	Improvável. Na localização, os textos são vistos como entidades que participam de um <i>continuum</i> . Uma vez que cumprem seu papel, ou são atualizados ou desaparecem. Nesse sentido, essa causa não seria possível. Contudo, não se descarta de todo essa noção, visto que é possível que mudanças arbitrárias na terminologia façam com que surjam variantes em textos localizados, mesmo que a distância temporal que separe um do outro seja medida em trimestres algumas vezes.
Variação social	Improvável. Embora a variação social seja possível no âmbito da localização, por exemplo, na interação entre os diferentes profissionais que atuam em um projeto desse tipo (como desenvolvedores de <i>software</i> , redatores, especialistas em alguma área), o tipo de conteúdo analisado neste trabalho situa-se em uma parte específica do processo, na qual a comunicação (escrita) tende a ser feita por profissionais do texto (redator, tradutor) e, portanto, tende a ser mais sistemática, uniformizada e a obedecer a quaisquer padrões de escrita determinados por uma empresa.
Causas funcionais	Para as causas funcionais, depende do caso específico. Dificilmente registro linguístico influirá, em função do nível de interlocução dos textos analisados. O grau de especialização, por sua vez, é uma causa possível. (mais detalhes a seguir, nas observações individuais para cada uma das subcausas).
Adequação ao registro linguístico	Improvável. Como o material analisado por este trabalho mantém um nível de interlocução consistente com a terminologia empregada, esta causa dificilmente aparecerá.

Causa	Probabilidade de ocorrência
Adequação ao grau de especialização	Provável. Pode ser relevante; em geral, textos de localização estão no nível semileigo-leigo; a suposição é a frequência de variantes seja grande (por exemplo, por meio de paráfrases explanatórias).
Causas discursivas	Com relação às causas discursivas, a relevância depende da subcausa, como se vê a seguir.
Evitar a repetição	Provável. Embora a repetição seja desejável em localização, é possível que isso ocorra; a hipótese é que vá aparecer mais nas traduções.
Economia linguística	Provável. Embora a economia linguística seja evitada em localização, visto que se busca criar um texto o mais explícito possível, é possível que isso ocorra; a hipótese é que vá aparecer mais nas traduções.
Criatividade, ênfase e expressividade	Improvável. É prática comum empresas diferentes buscarem denominações próprias para competir no mercado. No entanto, por se tratar de uma análise centrada no material de uma única empresa, é provável que essa causa não se manifeste.
Causas interlinguísticas	Como este trabalho analisa traduções de uma área bastante volátil em sua terminologia, é possível que essa causa esteja presente.
Coexistência entre termo local e empréstimo	Provável, considerando-se a área geral dos textos do <i>corpus</i> (computação; informática). Contudo, por se tratar de um <i>corpus</i> limitado a textos de uma só empresa, a probabilidade de ocorrer essa coexistência não é muito grande.
Causas cognitivas	Com relação às causas cognitivas, é possível que ocorram, dependendo da subcausa. Por exemplo, a imprecisão conceitual é uma causa possível, ao passo que as diferenças de conceituação não são (mais detalhes nas linhas abaixo nesta tabela).
Imprecisão conceitual	Provável. Neste caso específico, está diretamente relacionada com as causas interlinguísticas: termos novos são cunhados, mas o tradutor não dispõe de conhecimento suficiente para encontrar uma tradução – e dificilmente encontrará, dada a volatilidade da área e do rápido ciclo de inova-

Causa	Probabilidade de ocorrência
	ção.
Distanciação ideológica	Improvável. Pode ter relevância, mas é pouco provável. Em função do <i>corpus</i> analisado, a tendência é que não haja divergências.
Diferenças na conceituação	Improvável. Pouco relevante, visto que o <i>corpus</i> é composto exclusivamente de textos de uma empresa individual.

Fonte: Elaborado pelo autor.

Este trabalho adotará em parte a análise de causas para variação, com foco apenas naquelas que forem mais relevantes para a investigação proposta. Não se descarta, no entanto, que as causas ignoradas não venham a aparecer também. Por isso, elas não deixarão de ser consideradas durante a análise dos dados.

3.4 Tradução

O tema da tradução acompanha as relações interculturais desde a antiguidade, como se observa nas inscrições egípcias, em dois idiomas, que datam do Antigo Império do Terceiro Milênio a.C. (NEWMARK, 1981, p. 3), nas traduções de Platão por Cícero no séc. I a.C. e nas traduções bíblicas de São Jerônimo no séc. IV (BASSNETT, 2002). Seu estudo no ocidente, de que se tem notícia, data de I a.C. com as discussões de Horácio e Cícero acerca da prática da tradução de clássicos gregos para o latim. Desde então, os estudos sobre o tema têm tratado, mesmo que implicitamente, da função que a tradução tem como produto e sua fidelidade para com o original (ou qualquer noção que venha a servir para explicar o vínculo entre o original e sua tradução).

Em seu estudo sobre a história das teorias da tradução, Bassnett (2002) mostra que até o séc. XIX as visões teóricas ocidentais acompanhavam a prática da tradução de perto, operando como modelos a serem seguidos pelos tradutores. Trata-se, portanto, de um campo teórico que nasce da prática – a exemplo de outros, como o da Terminologia. Bassnett também mostra que o tradutor tinha um papel social ativo especialmente em questões culturais e políticas; a atividade do tradutor demonstrava uma preocupação com a fidelidade ao original, mas também com seu público.

Com o passar do tempo, noções teóricas como a de fidelidade ou a oposição entre tradução literal e livre não escaparam de um escrutínio acadêmico, dada a falta de objetividade de suas definições. Afinal, uma tradução deve ser fiel a quê? Ou a quem? Ao texto original e à problemática de defini-lo? Ao público consumidor da tradução? É possível estender a pergunta à oposição entre tradução literal e livre. A noção de tradução literal toma por base o caráter formal das línguas, ou seja, na repetição de formas semelhantes entre dois idiomas. Essa noção sobreviveu com certa facilidade nos estudos europeus, em função de certa similaridade constatável entre os idiomas do continente, mas não sobrevive a uma breve incursão rumo ao oriente, por exemplo.

Assim, com os estudos contemporâneos a partir da segunda metade do séc. XX, alguns temas recorrentes começaram a ser questionados. A fidelidade deu lugar a outras noções, como a da equivalência, que passariam a explicar de maneira mais satisfatória a relação que há entre um original e sua tradução. A oposição entre tradução livre e literal também enfraqueceu – embora ainda seja observada até hoje na prática, inclusive por pessoas envolvidas tangencialmente ou até mesmo diretamente com o ofício de tradutor.

Esses avanços levaram ao surgimento de teorias como as comunicativas e funcionalistas, que são eficazes para descrever e analisar o fenômeno em todas as etapas do processo tradutório, permitindo abordá-lo de maneira consistente com base em diversas noções, como o tipo de tradução (técnica, literária), a modalidade (tradução de *software*, dublagem, legendagem), o enfoque (por exemplo, microtextual sob a noção de unidade de tradução), entre outras – todas sob a premissa básica da equivalência e especialmente sob a tutela da funcionalidade, em que a primeira estabelece um vínculo entre o original e sua tradução e a segunda dá ao tradutor crédito pelo produto de seu trabalho e atribui suas escolhas tradutórias principalmente ao papel que a tradução cumpre na cultura de chegada.

A tradução como atividade também supõe avanços com o passar do tempo. A tradução profissional, ou pelo menos uma grande parte desta, progrediu mais ou menos no mesmo ritmo desses avanços teóricos – só que na direção oposta. Propulsionada pelo desenvolvimento industrial e tecnológico, surge a partir dos anos 80 a indústria da localização⁴², a qual faz uso maciço de tecnologia computacional para eliminar custos e segmentar o processo tradutório. O texto deixa de ser uma unidade coerente e passa a ser um conjunto de segmentos textuais reaproveitáveis. Contexto e cotexto fogem das mãos do tradutor. O aspecto funcional continua em foco, mas foi deslocado das etapas que envolvem tradução para outras, como as de con-

⁴² Cf. *Localização* na página 22 para uma introdução geral no tópico e a página 101 neste capítulo para obter mais informações sobre o paradigma relacionado.

cepção de um produto ou em revisões posteriores do material traduzido por profissionais de *marketing* ou departamento jurídico. Em função disso, o controle e o crédito que começavam a ganhar nitidez na teoria nunca estiveram tão perto de sumir na prática.

É em meio deste conflito entre os avanços teóricos e os (potenciais) retrocessos práticos que este trabalho inscreve seu ponto de análise, entre a metodologia comunicativa, funcionalista e incluyente de Hurtado Albir e o paradigma da localização excludente e antagonista de Pym. A premissa é que a metodologia e a concepção de tradução de Hurtado Albir permitam uma análise satisfatória dos produtos da localização, observando-se as particularidades apontadas por Pym em seu paradigma.

Esta revisão teórica tratará do tema em função dos estudos contemporâneos da tradução (ou seja, da segunda metade do séc. XX em diante), apresentando uma síntese sobre as concepções de tradução, percorrendo junto com Pym (2010, 2011) os grandes paradigmas, da equivalência à localização, e introduzindo o modelo comunicativo, descritivo e funcional proposto por Hurtado Albir (2001) com enfoque nas noções mais aplicáveis a este trabalho, tais como equivalência, técnicas e erro de tradução.

3.4.1 Concepções de tradução

Dos quatro grandes temas tratados nesta revisão, a tradução é provavelmente aquele que vem sendo estudado há mais tempo. Nessa condição, supõe-se que também apresente diferentes designações, baseadas em variadas perspectivas. Terminologicamente, *tradução* pode apresentar significados diferentes, designando processo, produto e conceito – de maneira análoga aos outros temas tratados neste trabalho, terminologia, fraseologia e variação. Bell (1991 apud HURTADO ALBIR, 2001, p. 40) faz essa distinção, indicando que o termo *tradução* pode designar:

- 1 o ato de traduzir: o processo ou a atividade;
- 2 uma tradução: o produto que resulta do processo de traduzir ou o texto traduzido;
- 3 a tradução: o conceito abstrato que abarca tanto o processo de traduzir quanto seu produto.

Segundo Hurtado Albir (2001, p. 37), existem muitas definições para *tradução*, com variados enfoques e concepções distintas: atividade interlinguística, aspecto textual, caráter

comunicativo, processo. Segue um apanhado das diferentes designações catalogadas pela autora:

Tabela 8 - Definições para tradução

Autor	Definição de tradução	Enfoque
Vinay e Darbelnet (1995)	Passar de uma língua A para uma língua B expressando a mesma realidade.	Atividade interlinguística
Seleskovitch e Lederer (1984)	Transmitir o sentido de um texto – em oposição a fazer a conversão do idioma em que o texto está formulado.	Aspecto textual
Catford (1965)	Substituir o material textual em uma língua por material textual equivalente em outra.	Aspecto textual
House (1977)	Substituir um texto original por outro semântica e pragmaticamente equivalente na língua de chegada.	Aspecto textual
Nida e Taber (1969)	Reproduzir, mediante uma equivalência natural e exata, a mensagem da língua original na língua receptora.	Ato de comunicação
Hatim e Mason (1990)	Processo comunicativo que tem lugar em um contexto social	Ato de comunicação
Hermans (1991)	Prática comunicativa e, portanto, um tipo de comportamento social; tem lugar em uma situação comunicativa, e os problemas de comunicação podem ser definidos como problemas de “coordenação” interpessoais, os quais fazem parte da ampla família de problemas de interação social.	Ato de comunicação
Toury (1980)	Ato intrassistêmico de comunicação	Ato de comunicação
Snell-Hornby (1988)	Ato transcultural	Ato de comunicação

Autor	Definição de tradução	Enfoque
Hewson e Martin (1991)	Equação cultural	Ato de comunicação
Reiss e Vermeer (1984)	O princípio dominante de toda tradução é sua finalidade.	Ato de comunicação
Nord (1988)	Ato comunicativo cujo critério fundamental é a funcionalidade.	Ato de comunicação
Vázquez Ayora (1977)	O procedimento tradutório consiste em analisar a expressão do texto original em termos de orações pré-nucleares, traduzir essas orações em orações pré-nucleares na língua de chegada e, finalmente, transformar essas estruturas da língua de chegada em expressões estilisticamente apropriadas.	Processo
Seleskovitch e Lederer (1984)	Processo de tradução está mais relacionado com operações de compreensão e reexpressão do que de comparação entre línguas.	Processo
Steiner (1975)	Processo de transformação, interpretativo, hermenêutico; em seu modelo esquemático, a tradução consiste em uma mensagem proveniente de uma língua fonte que passa para uma língua receptora, sofrendo um processo de transformação.	Processo
Delisle (1980, 2009)	A atividade tradutória se define como determinar a significação dos signos linguísticos encontrados em uma mensagem em função <i>daquilo que se quis dizer</i> e restituir essa mensagem na íntegra mediante os signos de outra língua.	Processo
Bell (1991)	A tradução pode ser definida como processo, produto e conceito.	Processo

Embora a autora crie uma separação entre os pontos de vista de cada autor, essa divisão não é clara. Por exemplo, Vinay e Darbelnet (1995) definem a tradução como uma atividade interlinguística, enquanto Catford (1965) a define como atividade textual, mas em uma definição que permanece ancorada no plano da língua. Outro exemplo é House (1977), que a autora situa nas definições textuais, mas que já traz aspectos pragmáticos também. Mais importante que a separação entre perspectivas é a conclusão de Hurtado Albir (2001, p. 40) para essa diversidade, de que isso não faz mais do que demonstrar toda a complexidade do tema e permitir sua caracterização como texto, ato de comunicação e atividade cognitiva.

Para Hurtado Albir (2001, p. 40-42), a tradução se caracteriza por três aspectos básicos: é um ato de comunicação, uma operação entre textos e um processo mental.

Como ato de comunicação, Hurtado Albir afirma que as traduções são feitas com uma finalidade comunicativa, para que um destinatário que desconhece a língua possa compreender o texto original. Essa finalidade varia segundo o tipo de trabalho ou o público alvo, e para cada caso o tradutor adota métodos diferentes e chega a soluções distintas.

Ao definir tradução como operação entre textos, a autora trata de afastá-la do plano da língua e situá-la no da fala – em uma referência à oposição saussuriana entre língua (*langue*) e fala (*parole*). Nesse sentido, não se traduzem unidades isoladas e fora de contexto, mas sim se traduzem textos. Ao analisar a tradução, é preciso considerar os mecanismos de funcionamento textual (elementos de coerência e coesão, tipos e gêneros textuais) e lembrar que esses mecanismos variam para cada idioma⁴³.

Como atividade cognitiva, Hurtado Albir descreve a tradução como uma atividade realizada pelo tradutor que exige uma competência específica (a competência tradutória). Ao desempenhar essa atividade, o tradutor realiza um processo mental que consiste em compreender o sentido transmitido por um texto e reformulá-lo com os meios de outro idioma, tendo em conta as necessidades do destinatário e a finalidade da tradução. A ordem é importante nessa definição: primeiro o tradutor interpreta o texto, o contexto e a finalidade da tradução para então comunicar.

A junção desses três aspectos básicos culmina na seguinte definição: a tradução é um processo interpretativo e comunicativo que consiste na reformulação de um texto com os mei-

⁴³ Essa definição está em desalinhamento com a noção de equivalência do paradigma da localização, sobre o qual este trabalho trata em *Localização* (p. 96).

os de outra língua que se desenvolve em um contexto social e com uma finalidade determinada (HURTADO ALBIR, 2001, p. 41)⁴⁴.

A definição proposta por Hurtado Albir se reflete de maneira consistente no arcabouço metodológico que a autora desenvolve para seu trabalho, bem como em suas análises. A autora propõe que se integrem os níveis de análise, observando não só as relações internas desses mecanismos (de texto a texto), mas também as externas (fator espaço-temporal, importância do receptor, tipo de trabalho, finalidade da tradução) e as competências e os processos mentais implicados.

Essa definição será adotada neste trabalho, visto que está alinhada com as outras bases teóricas empregadas.

3.4.2 Paradigmas da tradução

Na história dos estudos contemporâneos da tradução no ocidente, séries de designações e conceituações têm sido elaboradas sob diversas bandeiras teóricas. Ao fazer sua análise e estudo, Pym (2010, 2011) afirma que é possível agrupá-las em conjuntos de princípios básicos que são partilhados por teorizações distintas. A premissa está na existência de ideias gerais que apresentam coerência interna e um ponto de vista compartilhado entre diferentes perspectivas teóricas.

Sob essa premissa, o autor traça um histórico sobre os estudos ocidentais com base em cinco paradigmas principais – equivalência, funcionalismo, descritivismo, indeterminismo e localização⁴⁵ –, ocupando-se das respectivas implicações, com destaque especial para a repercussão do paradigma da localização para o tradutor, o qual seria incompatível com teorias funcionalistas e descritivas da tradução⁴⁶.

As subseções a seguir apresentam uma breve descrição dos paradigmas apresentados pelo autor.

⁴⁴ Tradução do autor deste trabalho. No original: *proceso interpretativo y comunicativo consistente en la reformulación de un texto con los medios de otra lengua que se desarrolla en un contexto social y con una finalidad determinada* (HURTADO ALBIR, 2001, p. 41).

⁴⁵ Pym apresenta seis paradigmas, mas apenas os cinco apresentados neste trabalho estão voltados para textos (produtos) e se aplicam a este trabalho. O sexto paradigma, da tradução cultural, está voltado para o movimento de pessoas (indivíduos).

⁴⁶ Para mais detalhes sobre essa questão de incompatibilidade, consulte *Localização* (p. 96).

3.4.2.1 Equivalência

Esse paradigma agrupa as teorias que constituem seus conceitos em torno da noção de que algo dito em determinado idioma pode ter o mesmo valor, importância ou função quando traduzido para outro idioma. Nesse sentido, a relação entre o original e a tradução é de valor igual, não importa se a relação está no nível da forma ou função, entre outras. As teorias sob esse paradigma apresentam dois subparadigmas ou tendências.

O primeiro é a equivalência “natural”, que pressupõe um vínculo prévio entre o original e sua tradução, situado no plano da língua. Ou seja, trata-se de uma equivalência prévia ao ato da tradução. Autores como Vinay e Darbelnet (1995) e Vázquez-Ayora (1977) fazem parte desse paradigma. Para Vinay e Darbelnet, em uma das primeiras teorias, a equivalência se referia ao que é possível fazer nos casos em que não há simetria entre as formas de dois sistemas linguísticos diferentes. O exemplo usado pelos autores é o vínculo tradutório entre *la soupe* dos militares franceses e *tea* dos britânicos (VINAY e DARBELNET, 1995, p. 41). Nesse caso, denominações e designações mudam de um idioma para o outro a fim de manter uma função equivalente. Disso, Pym conclui que a noção de equivalência foi funcionalista desde a sua origem.

Depois, com Eugene Nida (1964), outra noção vem à tona: no caso de *la soupe = tea*, há duas maneiras de ver a equivalência. De um lado, há a equivalência formal (*la soupe = the soup*) e do outro, a equivalência *dinâmica* (*la soupe = tea*), dinâmica no sentido de cumprir a mesma função cultural na outra língua. Assim, o termo se tornou mais abrangente, visto que se referia à noção de valor em qualquer nível linguístico, da forma à função. Os dois autores cujos trabalhos fundamentam este (HURTADO ALBIR, 2002; PYM, 2010) seguem essa perspectiva.

Um dos problemas básicos da equivalência natural está na premissa de que existe uma correlação sistêmica entre os idiomas. Esse vínculo prévio pressupõe, por exemplo, que um texto, traduzido de volta para o idioma do original, resultará exatamente no mesmo texto original. A partir dessa conclusão, surgem as propostas situadas no segundo subparadigma – o da equivalência *direcional*. Com a direcionalidade, ao traduzir de uma língua para outra, e depois fazer o caminho de volta, a tradução recíproca não será idêntica ao original. Nesse tipo de equivalência, a direcionalidade cumpre um papel fundamental por mostrar que as traduções são resultado de decisões ativas tomadas pelo tradutor. Como o tradutor é quem decide, os resultados nunca serão os mesmos. Essa relação de direcionalidade introduz uma série de teo-

rias sobre como os tradutores tomam decisões tradutórias, já que precisam constantemente fazer uma escolha entre várias alternativas possíveis.

Enquanto as teorias da equivalência natural tendem a apresentar vários procedimentos ou técnicas de tradução, na equivalência direcional a tendência é existir apenas duas noções antagônicas, como tradução livre *versus* tradução literal, o que remete à visão pré-contemporânea mais comum nos estudos que tratam do tópico, em que a tradução pode ser no nível de *palavra por palavra* ou *sentido por sentido*. Conforme afirma Pym, é possível observar essa característica nas oposições entre os tipos de equivalência apresentados por alguns teóricos:

Tabela 9 - Oposições para equivalência segundo alguns teóricos da tradução

Autor	Oposições de equivalência
Nida (1964)	<ul style="list-style-type: none"> ○ <i>formal</i>: seguir de perto as palavras e os padrões textuais do original; ○ <i>dinâmica</i>: recriar a função que as palavras teriam na situação original.
Levy (1969)	<ul style="list-style-type: none"> ○ <i>ilusória</i>: produzir um texto que esteja tão bem adaptado à cultura de destino que não seja percebido como tradução; ○ <i>anti-ilusória</i>: gerar um texto que mantenha alguns traços do original, permitindo que o leitor o identifique como uma tradução.
House (1997)	<ul style="list-style-type: none"> ○ <i>overt</i>: indica que o leitor sabe que o texto é uma tradução; ○ <i>covert</i>: indica que o leitor desconhece que o texto é uma tradução.

Fonte: Elaborado pelo autor com base em Pym (2010, p. 30).

Percebe-se que cada uma dessas oposições se refere a aspectos diversos. Assim, Pym afirma que equivalência era usada de pelo menos três maneiras diferentes:

- para conceituar a adaptação cultural (equivalência dinâmica);
- para se referir à reprodução de diferentes níveis e funções “naturais” do texto de partida (quando o termo de fato recobra o discurso milenar da fidelidade); e

- para discutir escolhas diferentes com as quais o tradutor se depara.

Uma crítica sobre as teorias nesse paradigma está relacionada ao fato de que elas sugerem uma prevalência do original sobre a tradução. Ao se perguntar a que uma tradução equivale, normalmente a resposta envolve algo encontrado no original. Nesse sentido, o original determinaria a relação de equivalência, prevalecendo sobre a tradução. Além disso, em geral, os autores que seguem esse paradigma desconsideravam o cliente na relação entre original, tradução e tradutor.

O resultado é um paradigma complexo, mas que com alguma frequência é reduzido a algumas de suas formulações mais ingênuas (como se verá, por exemplo, no paradigma da localização). Por trás de todas essas conceituações está a ideia comum de que o jeito de traduzir depende da natureza do texto de partida, já que a tradução equivale a este. É a partir dessa perspectiva que autores como Pym (2010) e Hurtado Albir (2001) apresentam uma definição de equivalência produtiva para os estudos da área (cf. *Equivalência tradutória* na página 107, seção *Noções centrais de análise tradutória*).

3.4.2.2 Funcionalismo

Este paradigma afirma que o fator dominante sobre cada tradução é sua função, suplantando assim o texto de partida e qualquer fidelidade a ele. Nessa perspectiva, a tradução é vista como algo que serve a funções bem diferentes das do original, já que se dirige a um público diferente, em uma situação cultural distinta. O tradutor deve decidir, em conjunto com o cliente, qual é a finalidade e, então, agir de acordo com ela. Essa teoria não abole a equivalência, apenas faz dela um caso especial, a ser buscado quando é necessário manter a função do original na tradução, ao que Reiss e Vermeer (1984) denominam *constância da função*⁴⁷. Ambos são autores de referência do funcionalismo e propõem a Teoria do Escopo (*ibid* 1996). Nessa teoria, o tradutor passa a ter participação ativa em suas traduções. A própria maneira de encarar a tradução muda. Mais do que a atividade de passar um texto de um idioma para outro ou como um objeto estático (“a tradução”), ela passa a ser tratada como um projeto. Nesse sentido, para esse paradigma as instruções do cliente podem ser muito mais importantes que o texto original, e o tradutor não é mais uma figura solitária que confronta um texto estrangeiro, mas sim um parceiro ativo em um ato complexo de comunicação.

⁴⁷ No original, *Funktionskonstanz*.

Uma das críticas sobre o paradigma é que, dado o rigor com que são formuladas suas teorias, sua aplicação não resulta em uma boa relação custo-benefício. A questão em pauta nesse caso é se os tradutores deveriam realmente realizar um extenso trabalho teórico antes de começar a traduzir. Pym faz alusão a 76 perguntas sobre o texto original que Nord (1988) afirma serem necessárias como parte do processo tradutório.

Um dos grandes méritos das teorias sob esse paradigma é ter colocado o tradutor como alguém que deve negociar com todas as partes envolvidas, como clientes, leitores usuários finais e autores, aproximando a teoria da prática profissional corrente.

3.4.2.3 Descritivismo

Esse paradigma sustenta que é possível revelar regras ocultas por meio da análise científica de produtos culturais. Para as teorias que seguem esse paradigma, as traduções são fragmentos de linguagem em contexto, em oposição ao que inúmeras teorias opinaram a respeito do ideal de tradução. Para o pesquisador israelense Gideon Toury (1980), a abordagem descritiva deve aceitar como axiomático que todas as traduções são equivalentes a seus originais. Somente assim a pesquisa poderá descrever os modos de tal equivalência.

Embora definido como descritivo, os trabalhos realizados segundo esse paradigma foram além do trabalho de descrever. Essa abordagem foi capaz de teorizar o papel das traduções no mundo. Para a escola de tradução israelense, a cultura de chegada pode ser vista como um sistema (ou um “polissistema”). Neste, as traduções podem ser centrais, desempenhando um papel inovador e ajudando a moldar a cultura, ou periférico, obedecendo a padrões estabelecidos e desempenhando um papel complementar ou reiterativo. Isso mina totalmente as prescrições sobre como fazer tradução: se uma cultura se vê inferior em relação a outra em alguma esfera, penderá para traduções literalistas; caso contrário, não.

Uma crítica sobre o paradigma, inclusive entre teóricos descritivistas, está no fato de que o lado que recebe a tradução não consegue explicar todas as relações envolvidas no processo, como proferido por Toury. Sem assegurar a importância do texto original, não é possível explicar, por exemplo, as relações entre culturas.

3.4.2.4 Indeterminismo

Esse paradigma agrupa as teorias que consideram que um tradutor não pode estar absolutamente certo a respeito dos significados que está traduzindo. Há duas correntes básicas. A primeira acredita que o texto original transcende suas interpretações individuais, o que faz com que as traduções sejam transitórias e tenham qualidade inferior. No lugar da equivalência (seja ela qual for), o máximo que se consegue obter é “similaridade”. A segunda crê que todo ato de comunicação apresente algum grau de indefinição e, assim, o indeterminismo também se aplica ao texto original, tornando-o um entre vários textos possíveis. Elementos desse segundo tipo de indeterminismo podem ser encontrados na teórica Rosemary Arrojo (1993).

Para o tradutor, as principais consequências disso é que eles não precisam mais ter respeito pelo sentido exato do texto original: eles estão livres para “trair” a forma de expressão e para dar vazão a seu papel criativo. A indefinição passa a ser um problema com o qual o tradutor deve lidar. Contudo, esse paradigma parece não dizer como o tradutor deve fazer isso. Uma crítica que esse paradigma recebe é a respeito de que essas teorias não trariam nenhum benefício prático para o tradutor e não ajudariam a explicar a complexidade envolvida no ato de traduzir. Apesar disso, o paradigma tem sua importância revelada no fato de que, independentemente da teoria aplicada, sempre é necessário lidar com algum grau de incerteza.

3.4.2.5 Localização

Ao contrário dos outros paradigmas, o paradigma da localização nasce fora do eixo acadêmico, saindo das observações de Pym a respeito de uma prática industrial que cresce cada vez mais em escala global. Conforme aponta o autor, enquanto a teoria do escopo reduzia a equivalência a um caso especial, os estudos descritivos a viam em toda parte e o indeterminismo questionava sua existência, a indústria da tradução enfrentava problemas mais práticos no âmbito da globalização econômica (PYM, 2009, p. 7-8)⁴⁸.

Um desses problemas está na necessidade de adaptar produtos computacionais segundo convenções (por exemplo, sistema imperial no inglês americano e sistema métrico no por-

⁴⁸ Tradução do autor deste trabalho. No original: *While Skopos theory reduced equivalence to a special case, descriptive studies were finding it everywhere, and indeterminism continued to question its existence, the translation industry was confronting the more practical problems of economic globalization* (PYM, 2009, p. 7-8).

tuguês brasileiro) ou particularidades de cada localidade (por exemplo, a cobrança de determinados impostos sobre transações bancárias em determinado país e sobre outros em uma localidade diferente). Essa necessidade não é estranha aos paradigmas mais recentes, como o do escopo e descritivismo; a questão está no ponto exato em que o tradutor se insere nesse novo paradigma. Como Pym aponta (2009, p. 8), o principal paradoxo da localização para a tradução é que, embora pareça promover a adaptação cultural, seus processos tradutórios, na verdade, se baseiam em modelos simplistas de equivalência. Todos os aspectos que exigem adaptação são pensados nas etapas pré e pós-tradutórias, o que relega ao tradutor o trabalho de converter cadeias textuais de um idioma para outro.

Outro problema é a velocidade que as empresas imprimem ao lançamento de seus produtos, em escala cada vez mais global, para suprir demandas e permanecer comercialmente relevantes. Esta pressa se reflete em todas as etapas da cadeia de produção de determinado bem, inclusive em sua tradução. Esse ritmo acelerado se reflete, por exemplo, no emprego de sistemas de gestão de conteúdo ou memórias de tradução, bem como na produção de textos em idioma simplificado, os quais facilitam e aceleram a tradução de grandes volumes de texto.

Pym conclui que a teoria do escopo e o descritivismo, cada qual a sua maneira, fizeram um apelo à prática da tradução, em geral na esperança de que pudessem permitir que a prática surgisse da teoria, trazendo princípios sólidos sobre como fazer tradução (para a teoria do escopo) e sobre como ela é geralmente feita (para os descritivistas) (2009, p. 8). Agora, o que se vê é que a prática em si – ou pelo menos uma ramificação bastante expressiva desta – não precisa de teorias e aparentemente as contradiz inclusive no fundamento básico da equivalência.

Para Pym, enquanto *localização* denomina a indústria de idiomas mais bem-sucedida de nossa era, em particular nas áreas de *software*, documentação de produtos e comércio eletrônico, *tradução* tem sido associada a uma indústria artesanal que não sofreu mudanças por séculos. Em geral, a tradução é vista como uma pequena parte da localização, e a localização é vista como uma modalidade de tradução. Os dois termos, contudo, denominam maneiras possivelmente antagônicas de tratar a comunicação intercultural. Na maioria dos pontos de vista encontrados na localização, a tradução é tratada como uma espécie de atividade rotineira, um mero problema de linguagem. Para alguns teóricos da tradução, os discursos sobre localização introduzem modismos, mas nenhuma novidade (2004, p. 16).

Pym coloca a questão desse paradigma sob uma perspectiva mais abrangente do que a de localização e tradução – ou mesmo de globalização. Para ele, trata-se de um problema de *distribuição*. Essa questão será tratada a seguir.

3.4.2.5.1 Conceitos fundamentais do paradigma

Pym (2004) faz uma reflexão a respeito dos termos-chave da localização. As seções a seguir apresentam um resumo dos principais pontos apresentados pelo autor.

3.4.2.5.1.1 Distribuição

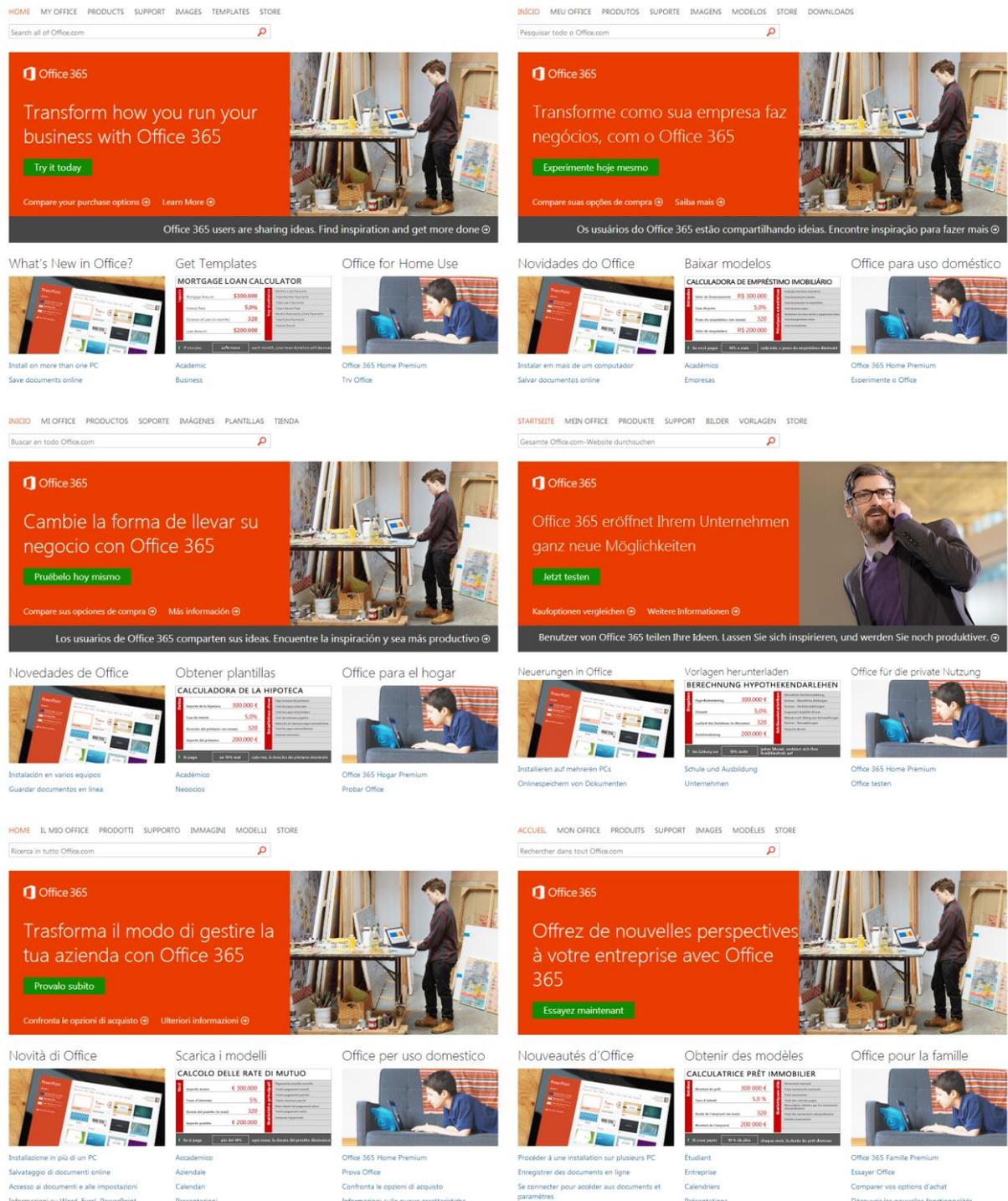
Distribuição, em uma concepção bem material do termo, é determinante para o processo de localização como um todo. Esse conceito diz respeito à distribuição de objetos em um espaço e tempo determinados – o que atingiu um novo patamar em função da Internet. Não faz muito tempo, um fã ficava conhecendo somente parte limitada do acervo de seu artista internacional favorito. Hoje está a um clique de distância de praticamente todos os álbuns, lados B e o que mais se puder imaginar.

Essa facilidade de acesso, essa facilidade de distribuição impactou profundamente a maneira com que as pessoas trabalham com um computador e, para as empresas do ramo, isso repercute com igual profundidade nos ciclos de produção e lançamento de produtos eletrônicos. Mais ou menos na mesma época, a pessoa que quisesse instalar o Windows 95 em seu computador teria que fazer a troca dos 13 disquetes que armazenavam seus arquivos de instalação. Hoje, uma instalação do Windows consiste em um arquivo .iso com gigabyte de tamanho acessível *on-line*.

Esse encurtamento na distância para se obter algo na Internet – no caso, um *software* e, por extensão, os textos que o acompanham – muda drasticamente nossa relação com esse conteúdo. No meio digital, todo mundo está cercado por localizações. Desde a versão do Microsoft Word em português, passando por Google, YouTube, Google Drive, e até notícias

lidas em portais como Terra, CNN e Associate Press. Por exemplo, os seguintes *links* levam para páginas do *software* Microsoft Office na Internet em vários idiomas⁴⁹:

Figura 11 - Página do Microsoft Office na Internet em diversos idiomas



Fonte: Microsoft (2013a, 2013b, 2013c, 2013d, 2013e e 2013f).

⁴⁹ Capturas de tela tiradas do *site* do Microsoft Office nos seguintes idiomas, respectivamente: inglês britânico, português brasileiro, espanhol europeu, alemão, italiano e francês.

Ao observar os *sites* acima, na visão de Pym (2004), é possível conceber o fenômeno de duas maneiras. Uma das concepções consiste na visão mais comum encontrada nas teorias da tradução, em que o original e a tradução são textos que ocupam determinado espaço em duas realidades distintas – no sistema cultural da língua de partida e no sistema da língua de chegada. Já a segunda, proposta por Pym, por estar fundamentada na noção de distribuição, concebe o original e a tradução como parte de um mesmo sistema.

Para o autor, a distribuição é essencialmente material e dinâmica; trata-se de movimentos reais no espaço e no tempo (PYM, 2004, p. 24). O autor opõe essa visão à concepção tradicional das teorias da tradução. Por exemplo, na visão tradicional, quando um *software* em inglês é traduzido para o português, ocorre um contato entre dois sistemas culturais (inglês e português), mas o *software* original não sai do lugar. Ele permanece onde está, operando no sistema cultural do inglês. Em vez de distribuição ativa, ocorre reprodução ou adaptação passiva. Nessa concepção tradicional, o original age como um representante estático da tradução.

Embora essa visão seja válida e represente o senso comum, o autor afirma que ela se baseia em uma noção idealista de sistemas culturais (como línguas e literatura). Nessa visão, os sistemas origem e meta estão estruturados em termos de presença e ausência. Quando um texto entra no sistema, ele fica lá por um longo período e sua presença afeta todos os outros textos no sistema (PYM, 2004, p. 25). Isso funciona para os modelos mais representativos nas teorias da tradução, mas não para a localização, especialmente por causa da velocidade dos ciclos de produção industrial.

Na localização, um texto localizado não serve para representar um texto anterior (o original), mas sim faz parte do mesmo processo de distribuição material constante, que começa em uma cultura e pode continuar em muitas outras. Em vez de pular entre sistemas estáveis de diferença mútua, a distribuição pode ser representada como um festival de fogos de artifício, na metáfora usada pelo autor, em que os textos atingem um ápice de visibilidade e depois vão esmaecendo.

Logo, o *software* não está disponível em âmbito geral para os falantes do idioma inglês, mas sim são usados em vários nichos especializados dentro desse âmbito e são rapidamente suplantados em função dos ciclos de produção mencionados acima. A empresa que não entra nesse jogo está fora, perde a relevância. Basta considerar a quantidade de versões e atualizações que são lançadas a cada ano para cada produto. Usando de exemplo um dos materiais coletados no *corpus* analisado neste trabalho, o *software* Norton AntiVirus já teve 14 versões de 2000 a 2013. Cada versão reaproveita boa parte do material produzido nas anteriores, e

apenas as partes novas ou modificadas são traduzidas, embora às vezes o material reaproveitado possa passar por edição também.

Quanto mais arraigado à Internet for o produto, mais intenso esse movimento fica. O Google Docs foi lançado em 2007. De lá até hoje, sofreu várias atualizações, novos recursos foram adicionados. Mas quem saberia dizer em que versão o produto está?⁵⁰ Sua existência está totalmente baseada no agora; uma vez que se torne irrelevante, deixa de existir. Por isso, é necessário um trabalho constante para estabelecer e manter a distribuição em todos os mercados. Esse trabalho constante acaba gerando várias ações de distribuição em pequena escala, como divulgação publicitária, correção e atualização do *software* e adaptação para outras localidades.

3.4.2.5.1.2 Localização

A localização, vista como a adaptação de um texto para determinado público, cobre tudo aquilo que ao longo de décadas os teóricos da tradução tentam legitimar, por exemplo, no paradigma da finalidade associado ao da equivalência. No entanto, ao contrário das teorias da tradução, a localização tira o controle das mãos do tradutor (PYM, 2001, p. 5). Logo abaixo, as conclusões (p. 102) desenvolvem um pouco mais o tema.

3.4.2.5.1.3 Internacionalização e globalização

A internacionalização ou globalização representaria uma aproximação ao ideal de se chegar a uma linguagem técnica universal, visto que busca eliminar marcas culturais características de determinada localidade para gerar um texto não marcado e descontextualizado. Ao contrário de previsões que afirmam que as diferenças culturais nunca seriam subjugadas pelo imperialismo comercial, o universalismo parece possível no âmbito da localização (PYM, 2001, p. 4).

⁵⁰ A pergunta não é válida para funcionários do Google.

3.4.2.5.2 Repercussão do paradigma para o tradutor

Embora a perspectiva da localização encontre eco nas teorias da tradução, o problema está no espaço que ela reserva para essa atividade e para o tradutor. Para a localização, a tradução é uma pequena parte de todo o processo de adaptação e fica relegada a uma mera conversão de códigos linguísticos, quase equiparável às noções de equivalência encontradas nas primeiras teorias modernas da tradução.

Por exemplo, Esselink (2000, p. 4) expõe a questão, definindo tradução como o processo de converter palavras em língua oral ou escrita em outro idioma, em que o significado do original deve ser reproduzido com precisão e integridade no idioma de destino. O autor ainda complementa com a definição apresentada pela extinta LISA (referência recuperável, hoje, apenas tangencialmente em outros trabalhos, como em Pym ou no próprio Esselink), de que a tradução é só uma das atividades encontradas na localização. Além dessa, é possível citar gestão de projetos, engenharia de *software*, testes e editoração eletrônica.

Um dos problemas, segundo Pym, está na aparente contradição entre os principais termos ou processos atrelados à localização. Por um lado, a localização prevê a adaptação de textos segundo o idioma e a cultura de determinado mercado. Por outro, a internacionalização vislumbra um texto simplificado, fragmentado e descontextualizado. A contradição aparente está na perspectiva para cada caso. Na primeira, observa-se o processo como um todo. Na segunda, a fragmentação e descontextualização estão na perspectiva do tradutor. Assim, a localização é ao mesmo tempo compatível e incompatível com as teorias textuais e funcionalistas da tradução, visto que os projetos de localização preveem adaptação e consideram a finalidade, mas relegam ao tradutor o trabalho menor de converter códigos linguísticos.

3.4.2.5.3 Síntese

Como se vê a respeito desse paradigma, há diversos questionamentos quanto ao papel reduzido que o tradutor tem em projetos de localização. Em geral, o tradutor se vê relegado a um mero conversor de cadeias sintagmáticas. Pym qualifica o papel do tradutor nessa área como algo muito pequeno, que o desvaloriza. Em última análise, o ofício é visto normalmente como só uma etapa na carreira profissional de alguém, e um tradutor deveria rapidamente

tentar migrar para áreas mais interessantes, como a gestão de projetos ou a redação técnica. Esse ponto de vista é partilhado por muitas empresas que adotam processos de localização ou que prestam serviços de tradução ao setor de localização.

No entanto, pode-se ganhar um bom dinheiro trabalhando como tradutor na área. Talvez o trabalho não seja o mais interessante na área. Mas também pode ser que exista um aparente jogo de forças no meio acadêmico, que vai das comparações entre tradução literária (mais interessante) e tradução técnica (mais entediante). Agora, esse jogo é levado mais além com o surgimento de um novo paradigma, em uma escala que vai do literário ou mais compensador, passando pela tradução de línguas de especialidade, até a tradução de projetos de localização, que atualmente recebe o título de menos interessante para o tradutor. Mas será que a atividade do tradutor técnico em si realmente muda com a introdução desse novo paradigma? Será que o tédio descrito por Pym para quem traduz um pedaço de um *software* é maior que aquele de quem traduzia o manual de uma máquina ou de um telefone nos anos 80?

Essa tradução – ora desprovida de contexto, ora desprovida de cotexto – dá oportunidades para quem quer fazer dela seu ofício diário. Será que não há desafios associados a trabalhar essa tradução? Será que a tentativa de sanar essa falta de contexto e cotexto não constitui por si um desafio – por exemplo, acessando o próprio *software* na versão original que está sendo atualizada (ou acessando uma versão anterior), recorrendo à potencialmente vasta documentação sobre este, seja na Internet, seja instalada no computador, ou ainda vasculhando amplas memórias de tradução devidamente catalogadas?

Seja como for, não importa a valoração que se queira dar, o fato é que há trabalho neste setor. Considerando a situação do mercado de tradução no Brasil, seria o caso de ver com bons olhos o surgimento e o aumento de oportunidades profissionais no âmbito da tradução, algo que se deve em boa parte à expansão de projetos de localização no país. Hoje, embora não seja possível falar de tradução como quem fala de, por exemplo, indústria alimentícia – no sentido de que não há uma indústria institucionalizada a ponto de dar perspectivas concretas de trabalho com carteira, que seria a situação de quem cursa alguma engenharia –, ao menos é possível ter uma noção um pouco mais concreta de oportunidades de trabalho, embora estas ainda requeiram uma certa dose de empreendedorismo e estejam muito direcionadas a quem domina a língua inglesa e quer trabalhar com esse idioma.

Essas afirmações, embora sejam um pouco generalistas, trazem algo de palpável. Esta é uma época que passa a valorizar cada vez mais a profissão do tradutor. Talvez nunca se tenha visto algo desse tipo voltado a esse ofício tão negativado. Isso tudo tem a finalidade de dizer que este trabalho é relevante para tradutores que atuem na área de localização, mas es-

pecialmente para qualquer indivíduo que se interesse por localização. Tem relevância por observar um fenômeno que ocorre em uma área que, embora seja pouco teorizada (seja na linguística ou em qualquer outra área teórica), cresce cada vez mais, atrai cada vez mais profissionais – entre eles, tradutores – e pode contribuir para uma tarefa que é talvez uma das mais onerosas que há no processo inteiro, no sentido de que o tempo dedicado não justifica os benefícios colhidos.

O fato de Pym lançar a localização como um novo paradigma aparentemente antagônico com muitos avanços conquistados nos estudos da tradução não quer dizer que não seja possível usar os instrumentos teóricos e metodológicos que foram desenvolvidos para a tradução em conteúdo produzido nessa área. Hurtado Albir (2001), em sua proposta abarcadora, aborda as traduções produzidas para localização como uma modalidade de tradução. É igualmente possível usar as técnicas de tradução descritas por alguns autores como instrumento metodológico desse tipo de texto. Talvez seja prudente considerar, contudo, as observações feitas por Pym para não apresentar resultados imprecisos ou equivocados em função de um enfoque igualmente equivocado. Ao usar textos e produtos localizados como objeto de análise, é preciso ter bem claro o papel limitado do tradutor na área da localização. É preciso ter bem claro que a equivalência no nível da tradução como processo de localização não é necessariamente a mesma que se vê no produto final (após as etapas de pós-tradução). É preciso ter em mente o frágil poder de decisão que o tradutor tem sobre o texto. É preciso lembrar-se da invisibilidade do tradutor, que dentro da localização talvez atinja um aspecto nunca antes visto nas Teorias da Tradução. Em função disso, serão apresentadas abaixo algumas noções centrais para a análise tradutória.

3.4.3 Noções centrais de análise tradutória

As noções apresentadas nesta seção para a análise tradutória tomam por base o trabalho de Hurtado Albir (2001), no qual a autora faz uma ampla revisão teórica sobre a teoria da tradução e propõe um modelo comunicativo, descritivo e funcional para tal análise. Nesse trabalho, a autora apresenta as seguintes noções como os temas centrais dessa análise: equivalência, unidade de tradução, invariável, método, técnica, estratégia, problema e erro. O quadro a seguir traz um resumo das noções, conforme apresentado pela autora, bem como a pertinên-

cia de cada uma para este trabalho. As noções relevantes serão descritas em detalhes em subseções posteriores.

Tabela 10 - Noções de análise tradutória e sua probabilidade de ocorrência neste trabalho

Noção de análise	Descrição	Probabilidade de ocorrência
Equivalência tradutória	Noção relacional que define a existência de um vínculo entre a tradução e o texto original. É estabelecida em função da situação comunicativa e o contexto socio-histórico da atividade de tradução. Tem caráter relativo, funcional e dinâmico.	Provável em função da problemática que envolve o tema na localização, conforme discutido em <i>Localização</i> (p. 101) e também porque estabelece o vínculo primordial entre original e tradução.
Unidade de tradução	Unidade comunicativa com a qual o tradutor trabalha. Situa-se no nível do texto e sua estruturação é variável, podendo consistir desde macrounidades, como um capítulo inteiro de um texto, até microunidades, como um sintagma nominal ou um prefixo.	Provável dada sua concepção como ponto de enfoque momentâneo do tradutor e a sua estruturação dinâmica. Como este trabalho trata como as UT e UFE são traduzidas segundo um glossário bilíngue, é possível situá-las dentro da noção de unidade de tradução.
Invariável tradutória	Noção relacional que define a natureza do vínculo entre a tradução e o texto original. Tem um caráter verbal, contextual, funcional e dinâmico.	No âmbito geral da localização, improvável, especialmente em função do conceito de <i>distribuição</i> (cf. <i>Distribuição</i> na página 98) introduzido no paradigma da localização. A partir do momento em que os textos em vários idiomas passam a fazer parte do mesmo movimento de distribuição, sem que se pressuponha um original, o problema de vincular o original com a tradução deixa de ser um problema. Contudo, é relevante às etapas específicas de TEP QA, visto que são as únicas em que a relação de vínculo é relevante. Para os tradutores e editores, os quais discutem a qualidade da tradução, essa noção relacional está presente.
Método de tradução	Execução de um processo de tradução que é regulado por determinados princípios em função do objetivo do tradu-	Provável, pois ajuda a situar o tipo de opção ao qual o tradutor recorre para trabalhar com textos localizáveis, que em geral estão desprovidos de

**Noção de
análise**
Descrição
Probabilidade de ocorrência

	<p>tor, respondendo a uma opção global que recorre o texto inteiro. Os métodos mudam segundo o contexto e a finalidade da tradução.</p>	<p>contexto e cotexto.</p>
Técnica de tradução	<p>Procedimento que o tradutor emprega para chegar à equivalência tradutória no nível de microunidades textuais. Esse procedimento é determinado mediante a comparação entre o original e a tradução. Seu uso é regido por alguns parâmetros, como funcionalidade, gênero textual, modalidade de tradução, finalidade da tradução e método.</p>	<p>Provável, na condição de instrumento metodológico para explicar a presença de ruído em uma lista de validação terminológica.</p>
Estratégia tradutória	<p>Procedimentos usados pelo tradutor para resolver os problemas encontrados durante o processo tradutório e para melhorar sua eficácia em função de suas necessidades específicas. Podem ser conscientes e inconscientes, verbais e não verbais, e internos e externos.</p>	<p>Improvável. A estratégia procura explicar, por exemplo, como o tradutor resolve determinados problemas de tradução ou para compreender o texto original. Situa-se mais em nível processual, e a análise feita neste trabalho concentra-se no resultado.</p>
Problema de tradução	<p>Dificuldades de caráter objetivo que podem surgir para o tradutor durante seu trabalho.</p>	<p>Improvável. Assim como a estratégia, o problema de tradução tem mais relação com o processo tradutório do que com o resultado. Para a análise feita neste trabalho, os problemas tangenciam os resultados, mas não estão estreitamente ligados.</p>
Erro de tradução	<p>Uma equivalência tradutória inadequada. Os erros são determinados por critérios textuais, contextuais e funcionais.</p>	<p>Provável, visto que o objetivo principal de uma lista de validação é, de fato, identificar possíveis erros. Metodologicamente, o erro de tradução serve como instrumento classificatório das ocorrências. Apesar disso, como o objetivo do trabalho não é catalogar ou estudar como a questão do erro é tratada em localização, essa categorização não precisa ser muito aprofundada, já que ela não interfere diretamente no processamento computacional de um QA.</p>

Fonte: Elaborado pelo autor com base em Hurtado Albir (2001, p. 308).

3.4.3.1 Equivalência tradutória

Hurtado Albir (2001, p. 223) define a equivalência tradutória como uma noção de natureza contextual que é funcional, relativa, dinâmica e flexível. Por defini-la de tal maneira, a autora cria uma oposição expressa à visão prescritiva de equivalência, que colocava a equivalência no plano estritamente linguístico. Nesse sentido, a equivalência é oposta à maneira com que o paradigma da localização concebe o fenômeno (cf. *Repercussão do paradigma para o tradutor* na página 102). Na proposta da autora, a equivalência é vista como um vínculo mútuo, que abarca a diferença, a manipulação, a visibilidade, a literalidade, etc. dependendo das escolhas do tradutor. Com isso, a equivalência passa simplesmente a determinar o vínculo que há entre o original e a tradução, seja ele qual for.

Com relação à concepção de equivalência no paradigma da localização, o que se observa na prática, ao contrário do que Pym afirma, não é necessariamente uma conversão de códigos linguísticos, embora seu dinamismo esteja muito mais limitado que o observado na tradução de texto literário – pelo menos em uma análise que se restringe a unidades terminológicas glossariadas.

3.4.3.2 Unidade de tradução

A unidade de tradução, segundo a concepção da autora, consiste na manifestação palpável da equivalência tradutória no nível do texto, embora não deixe de ser metalinguagem, uma abstração. Ela a caracteriza como uma unidade comunicativa que se situa no nível textual e que tem estruturação variável.

Por situar-se em nível textual, essa unidade não está presa à determinada forma linguística, mas sim vinculada à significação. Sendo assim, o tradutor não usa palavras individuais como unidade de tradução, mas sim unidades maiores. Expandindo um pouco mais a questão e considerando o processo tradutório, o tradutor trabalha no nível bitextual. Nesse sentido, essa unidade não é necessariamente síncrona. Ou seja, a unidade com que o tradutor trabalha ao ler e compreender o original não necessariamente equivale diretamente àquela produzida na tradução.

Com relação a sua dimensão e estruturação, a autora afirma que essa unidade varia para cada caso. Pode ser:

- uma macrounidade: o texto como um todo;
- uma unidade intermediária: no caso de um texto escrito, estaria no nível do parágrafo;
- uma microunidade: um sintagma nominal, um verbo ou um morfema.

Dentro do processo de localização, geralmente essa unidade está no nível micro, embora também possa ser situada no nível médio, ao se considerar os segmentos de tradução armazenados em uma memória computacional. Por considerar a unidade de tradução como uma unidade comunicativa, a autora atribui a essa noção as características dinâmicas que são necessárias para se fazer uma análise em traduções mais diversas possíveis, como legendagem e localização.

3.4.3.3 Método de tradução

O método de tradução diz respeito à maneira com que o tradutor encara o texto original e elabora a tradução segundo determinados princípios. Para a autora, o método tem um caráter supraindividual e consciente (embora possa ser inconsciente de vez em quando) e responde a uma opção global que recorre o texto todo.

A autora faz uma proposta de classificação dos métodos que toma por base a finalidade da tradução. Para a autora, a pertinência do uso de um método tem relação com o contexto e a finalidade da tradução. Isso significa que ele varia segundo o caso. Ela apresenta os seguintes quatro métodos básicos:

- método interpretativo-comunicativo: voltado para a compreensão e reexpressão do sentido do texto original, observando na tradução a mesma finalidade que o original e produzindo o mesmo efeito no destinatário. Mantém a tradução e o gênero textual;
- método literal: voltado à conversão dos elementos linguísticos do texto original, traduzindo palavra por palavra, sintagma por sintagma ou frase por frase, a morfologia, a sintaxe e/ou a significação do texto original. O objetivo desse método não é fazer com que a tradução cumpra a mesma finalidade do original, mas sim reproduzir o sistema linguístico de partida ou a forma do texto original;
- método livre: procura transmitir o mesmo sentido que o original, mantendo funções similares e a mesma informação;

- método filológico: consiste em inserir na tradução notas com comentários filológicos, históricos, etc. O original se converte em objeto de estudo, e a tradução se dirige a um público erudito ou de estudantes.

Destes métodos, o mais provável a ser usado na localização é o primeiro, o interpretativo-comunicativo. No entanto, esse método seria observável somente ao se considerar o processo de localização como um todo. A tomar por base o paradigma da localização, no ponto de vista do tradutor, o método literal seria a escolha mais provável. Isso tem uma implicação importante no objeto de análise deste trabalho, visto que diminui a probabilidade de que ocorram, por exemplo, reexpressões do sentido do original. Ou seja, pressupõe uma probabilidade maior de que as entradas do glossário sejam efetivamente usadas pelo tradutor nos segmentos de tradução.

3.4.3.4 Técnicas de tradução

A técnica de tradução é um procedimento concreto, geralmente verbal, que é visível no resultado da tradução. Segundo a autora, é usado para obter-se a equivalência tradutória afetando as partes menores do texto. Seu caráter é funcional e dinâmico, e seu uso é condicionado por alguns fatores:

- gênero textual (por exemplo, um contrato, panfleto turístico);
- tipo de tradução (por exemplo, técnica, literária);
- modalidade (por exemplo, escrita, consecutiva);
- método (por exemplo, comunicativo, livre).

As técnicas apresentam as seguintes características básicas:

- afetam o resultado da tradução;
- são catalogadas em comparação com o original;
- estão no nível das microunidades textuais;
- têm caráter discursivo e contextual;
- são funcionais.

A proposta de classificação de Hurtado Albir (2001) toma por base Molina (2001) e Molina e Hurtado Albir (2002), incluindo novas propostas para dar conta de mecanismos que ainda não haviam sido descritos. As técnicas apresentadas pela autora são as seguintes, incluindo um indicativo da probabilidade de ocorrência neste trabalho:

Tabela 11 - Técnicas de tradução e sua probabilidade de ocorrência neste trabalho

Técnica de tradução	Descrição	Exemplo	Probabilidade de ocorrência
Adaptação	Substituir um elemento cultural por outro próprio da cultura receptora.	Traduzir <i>baseball</i> por <i>futebol</i>	Improvável
Ampliação linguística	Incorporar elementos linguísticos. Recurso que costuma ser usado especialmente em interpretação consecutiva ou dublagem.	Traduzir <i>no way</i> por <i>de maneira alguma</i> , resultando em mais palavras na tradução; oposto de compactação linguística.	Improvável
Amplificação	Introduz “precisões” não formuladas no original: informações adicionais, paráfrases explanatórias, notas do tradutor; oposto de elisão.	Traduzir <i>Ramadã</i> por <i>mês de jejum dos muçulmanos</i> .	Provável
Calco	Traduzir literalmente uma palavra ou sintagma estrangeiro; lexical e estrutural.	Traduzir <i>public school</i> (Reino Unido) por <i>escola pública</i>	Improvável
Compensação	Introduz-se em outro lugar da tradução um elemento de informação ou efeito estilístico que não poderia ser refletido no mesmo lugar em que aparece no original.	A tradução de pronomes com níveis de formalidade diferentes no francês, como <i>tu/Vous</i> para o inglês, que inclui apenas <i>you</i> . Nesse caso, seria necessário compensar de alguma maneira na tradução, usando alguma outra formulação que indique formalidade no inglês.	Improvável; a prática é condenável nas etapas de tradução da localização, visto que pode gerar erros ao se reaproveitar os segmentos de uma memória de tradução descontextualizada.
Compactação linguística	Sintetizar elementos linguísticos. Recurso usado especialmente em interpretação simultânea e legendagem.	Traduzir <i>Yes, so what?</i> por <i>E?</i> em vez de <i>Sim, e daí?</i>	Improvável
Criação dis-	Estabelecer uma equivalên-	Traduzir nomes de filmes, como	Improvável

Técnica de tradução	Descrição	Exemplo	Probabilidade de ocorrência
cursiva	cia efêmera, totalmente imprevisível fora do contexto.	<i>Tommy Boy</i> por <i>Mong e Lóide</i>	
Descrição	Substituir um termo ou expressão pela descrição de sua forma ou função.	Traduzir <i>panetone</i> por <i>el bizcocho tradicional que se toma en Noche Vieja en Italia</i> .	Provável
Elisão	Não formular elementos de informação presentes no texto original.	Omitir <i>nos Estados Unidos</i> no original <i>em Nova York, nos Estados Unidos</i> em uma tradução para o inglês americano.	Provável
Equivalente consagrado	Usar um termo ou expressão reconhecido (por dicionário, uso linguístico) como equivalente na tradução.	Traduzir a expressão <i>kick the bucket</i> por <i>bater as botas</i> .	Improvável
Generalização	Usar um termo mais geral ou neutro.	Traduzir <i>ventilation flap</i> por <i>janela</i> (em vez de <i>basculante</i>).	Provável
Modulação	Mudar o ponto de vista, o enfoque ou a categoria de pensamento em relação com a formulação do original; pode ser lexical e estrutural.	Escolher entre <i>Golfo arábico</i> ou <i>Golfo pérsico</i> dependendo da ideologia.	Improvável
Empréstimo	Integrar uma palavra ou expressão de outra língua no estado em que se encontra.	Pode trazer mudanças ortográficas, como em <i>futebol</i> , ou não, como em <i>lobby</i> .	Provável
Particularização	Usar um termo mais preciso ou concreto.	Traduzir <i>window</i> por <i>basculante</i> .	Provável
Substituição	Mudar elementos linguísticos por paralinguísticos (entonação, gestos) e vice-versa.	Traduzir o gesto árabe de levar a mão ao coração por <i>obrigado</i> ; uso especialmente em interpretação.	Improvável
Tradução literal	Traduzir palavra por palavra um sintagma ou expressão.	Traduzir <i>kick the bucket</i> por <i>chutar o balde</i> (e não <i>bater as botas</i>)	Pode ser, mas para explicar algum erro – não parece provável

Técnica de tradução	Descrição	Exemplo	Probabilidade de ocorrência
Transposição	Alterar a categoria gramatical.	Traduzir <i>Some formulas are taking a while to calculate</i> por <i>O cálculo de algumas fórmulas está levando algum tempo</i> , em que o verbo <i>to calculate</i> é traduzido pelo substantivo <i>o cálculo</i>).	Talvez seja o mais provável de todos; também explicaria ocorrências segundo o modelo transformacional (mais informações na página 47)
Variação	Trocar elementos linguísticos ou paralinguísticos (entonação, gestos) que afetam aspectos da variação linguísticas: mudança de registro textual, estilo, dialeto social, dialeto geográfico.		Pouco provável, dadas algumas particularidades da localização, como a linguagem simplificada, internacionalizada. Essa técnica não deve ser confundida com outro aspecto analisado neste trabalho, a variação terminológica (cf. <i>Variação</i> na página 57).

Fonte: Elaborado pelo autor com base em Molina (2001) e Molina e Hurtado Albir (2002).

3.4.3.5 Erro de tradução

A autora apresenta uma concepção funcional e cognitiva para tratar os erros em tradução. Há três questões que são fundamentais para a análise do erro. A primeira é o aspecto textual, contextual e funcional da análise sobre a ocorrência de erros. Em segundo está a necessidade de estabelecer uma tipologia dos erros, a fim de criar uma metalinguagem. Por fim, deve-se considerar o caráter cognitivo do erro. Destes três, os dois primeiros são mais relevantes para este trabalho e serão discutidos em detalhes a seguir; o último está mais relacionado com o processo de tradução e tem um fundo didático sobre a formação do tradutor.

Com relação à ocorrência de erros, um elemento importante consiste em analisar sua gravidade. Segundo a autora, alguns autores criam categorizações dependendo do nível textual (macro ou microestrutura), de sua relação com o texto original (sentido incorreto, falta de sentido), entre outros. No setor da localização – como seria de se esperar de uma atividade industrial –, foram criadas métricas para a avaliação da qualidade de uma tradução. A métrica

consiste basicamente em uma avaliação de erros. É o que se observa na métrica de QA da extinta LISA⁵¹. No setor, em geral os trabalhos são avaliados em categorias distintas, e o impacto que um erro tem na avaliação depende do nível de gravidade atribuído (leve, médio, severo), se a avaliação adotada seguir a LISA, ou tanto da gravidade quanto do tipo de erro (por exemplo, sentido incorreto teria um peso maior que um erro de ortografia).

Para Hurtado Albir (2001, p. 303), a gravidade do erro não deveria depender de sua natureza (como sentido incorreto ou ortografia). Seria necessário considerar o todo, sob uma perspectiva textual, contextual e funcional que considere o elemento em questão em relação:

- ao texto como um todo;
- ao contexto de ocorrência;
- à finalidade da tradução;
- ao método escolhido;
- ao tipo e modalidade.

Embora a discussão sobre a gravidade de um erro seja válida e produtiva para a área da localização, este trabalho não pretende entrar em mais detalhes. Como instrumento metodológico, a análise de ocorrências de erro na lista de validação não receberá um tratamento aprofundado e servirá apenas para fins de catalogação. Dado seu enfoque na avaliação de ruído, um erro de tradução representaria o único tipo de ocorrência que idealmente o tradutor deveria encontrar em uma lista de validação.

Com relação à tipologia de erros, a autora apresenta uma proposta que cria uma separação entre:

- os erros relacionados ao texto original (como sentido incorreto, falta de sentido, omissão, entre outros) e aqueles que dizem respeito à produção do texto de chegada (ortografia, sintaxe, etc.);
- os erros funcionais e os absolutos, em que o primeiro está relacionado à falha do tradutor em fazer determinada tradução funcionar como o original e em que o segundo pode não ser um erro dependendo do contexto (como o uso de uma preposição mal empregada para assinalar um personagem estrangeiro em uma história);
- os erros no resultado da tradução e aqueles cometidos durante o processo tradutório.

Dessa separação, o primeiro e o segundo tipo são relevantes para este trabalho. Por fim, a autora apresenta a seguinte tipificação, a qual será usada para catalogar os erros encontrados na análise deste trabalho:

⁵¹ Recuperável em *sites* de outras empresas atuantes no setor, como a SDL (SDL, 2013).

- inadequações que afetam a compreensão do original:
 - omissão;
 - sentido incorreto;
 - falta de sentido;
 - outro sentido;
 - adição;
 - supressão;
 - referência extralinguística malsolucionada;
 - inadequação de variação linguística (tom, dialetos, idioleto, etc.).
- inadequações que afetam a expressão na língua de chegada:
 - ortografia e pontuação;
 - gramática;
 - léxico;
 - aspectos textuais (coerência, progressão temática, elementos de referência, conectores);
 - redação (formulação problemática ou pouco clara, falta de riqueza expressiva, pleonasmos);
- inadequações pragmáticas, por não estar de acordo com a finalidade da tradução (em relação ao tipo de trabalho, o gênero textual e suas convenções, etc.).

Com a adoção dessas noções, acredita-se que este trabalho poderá prover uma análise mais coerente com o que acontece nas traduções na área da localização, além de permitir o uso de uma linguagem comum que descreva e designe alguns pontos específicos à análise dos dados como um todo.

Os próximos capítulos dessa revisão abordam a descrição dos procedimentos metodológicos abordados neste trabalho, com base na revisão teórica aqui apresentada sobre os tópicos de terminologia, com fraseologia e variação, e de tradução.

4 PROCEDIMENTOS METODOLÓGICOS

Este capítulo descreve os materiais que serão analisados, as ferramentas usadas para tal fim e a metodologia de análise. Sobre os materiais e as ferramentas, o capítulo detalha o *corpus* de estudo e os critérios para sua seleção, descreve as informações contidas no *corpus* coletado e apresenta as ferramentas que foram usadas para sua coleta, consolidação e análise. Com relação à metodologia, apresenta o QA terminológico que foi executado, os critérios adotados para analisar e classificar os dados, como a análise foi feita e como foram organizados os dados resultantes desse processo.

4.1 Materiais e ferramentas

4.1.1 *Corpus*

O *corpus* consiste em material produzido em processos de localização e envolve três tipos de texto muito comuns na área (glossário terminológico, *software* e ajuda). As subseções a seguir apresentam os critérios adotados de seleção e os detalhes do *corpus* que será analisado neste trabalho.

4.1.1.1 Critérios de seleção

Para reunir o *corpus* a ser usado neste trabalho, foram adotados alguns critérios. Embora alguns possam parecer evidentes, todos estão descritos aqui para prover transparência nos resultados da análise. A seguinte tabela apresenta os critérios utilizados, fundamentados na proposta de Berber Sardinha (2004) e na experiência do autor em projetos de localização:

Tabela 12 - Critérios para a seleção do *corpus*

O <i>corpus</i> deve...	Comentários
Conter materiais que sejam produto da localização	Trata-se de algo evidente; como a análise se propõe a mostrar um problema real em uma atividade profissional específica – a localização –, não poderia ser de outra forma.
Ser bilíngue, no par de idiomas EN-US / PT-BR	Em função da maneira com que o QA terminológico é executado pelo <i>software</i> , o formato do <i>corpus</i> precisa corresponder àquilo que é usado na localização. Já o par de idiomas foi escolhido por dois motivos: primeiro, é o par com o qual o autor deste trabalho tem experiência na área; segundo, o idioma inglês americano é em geral o escolhido e usado por empresas que empregam metodologias de localização e o português brasileiro é representativo do âmbito institucional ao qual este trabalho se insere.
Representar três dos gêneros textuais mais comuns encontrados em localização: <i>software</i> , respectiva ajuda (instrucional) e glossário	Espera-se que a presença de conteúdo de gêneros diferentes possa trazer à tona particularidades específicas a cada um. Além disso, a tradução do par <i>software</i> e ajuda consiste em um dos formatos de projeto de localização mais comuns. Mais detalhes sobre esses gêneros em <i>Interface de usuário</i> (p. 121) e <i>Ajuda</i> (p. 123).
Ter uma dimensão representativa de uma situação real de QA terminológico	Não houve preocupação em reunir grandes volumes de texto. Um dos objetivos é mostrar o tamanho de uma lista de QA terminológico em situação real de trabalho. O volume definido para os <i>subcorpora</i> seria processável por quatro pessoas (dois tradutores, um editor, um revisor textual) em um pouco mais de um mês de trabalho (com dedicação exclusiva a tal projeto). Para mais informações sobre o volume, consulte <i>Síntese do corpus</i> (p. 125).
Respeitar a propriedade material ou autoral	Em função de confidencialidade e propriedade intelectual, o material recolhido deve ser publicamente acessível a qualquer pessoa com um computador, em acesso disponibilizado pelo detentor do projeto. Essa tarefa trouxe limitações, especialmente pela dificuldade de encontrar terminologia formalizada e estruturada disponível ao público. Atualmente, no levantamento feito neste trabalho, apenas três empresas disponibilizam publicamente suas bases de dados terminológicos. São elas: Microsoft, Symantec e SAP ⁵² .
Envolver produtos voltados para um público abrangente	A ideia é que, com isso, a terminologia técnica seja relativamente diluída, possibilitando traduções menos “presas” ao original.

Fonte: Elaborado pelo autor.

⁵² Microsoft (2013g) foi a pioneira nesse quesito, seguida pela Symantec (2013b); a última o fez recentemente, no ano de 2013 (SAP, 2013).

4.1.1.2 Descrição do *corpus*

Antes de falar das informações contidas no *corpus* em si, é necessário apresentar brevemente sua origem. O objetivo é demonstrar a relação direta com a localização, bem como contextualizar a natureza das informações nele contidas.

O *corpus* contém material textual que pertence à empresa Symantec, uma corporação global que é especializada em *software* para a segurança de dados digitais, como antivírus, firewall, pacotes de segurança na Internet e backup. Além de produzir o *software* usando metodologia baseada em localização, a empresa disponibiliza sua base terminológica para consulta *on-line* em mais de vinte idiomas. Além disso, desenvolve produtos tanto para consumidor final quanto para corporações. Trata-se de um dos carros-chefes dessa empresa, com enfoque justamente no usuário final.

Nessa breve introdução, observam-se os principais motivos para a escolha de conteúdos produzidos por essa empresa. Primeiro, o material é produto de localização, o que permite a reprodução do processo de QA aplicado na área. Em segundo lugar, está a disponibilidade de uma base terminológica multilíngue, visto que não são muitas as empresas que deixam esse tipo de conteúdo disponível⁵³.

Em termos concretos, o *corpus* consiste em três *subcorpora* bilíngues: um glossário, o material textual de uma interface de usuário (UI) e ajuda. Todos os *subcorpora* pertencem à mesma área temática, ou seja, segurança de dados digitais, e a UI e a ajuda pertencem ao mesmo *software* na mesma versão: Norton AntiVirus versão 2011. Já o glossário abrange também outros produtos da empresa, como Norton 360, Norton Firewall, Norton Online Backup e Norton Safe Web (todos listados como entradas do glossário). Embora cada *subcorpus* contenha material de um gênero textual diferente, todos estão interligados de alguma maneira: a ajuda se refere ao *software*, e o glossário contém entradas terminológicas que valem para ambos.

Em síntese, o *corpus*:

- pertence a uma empresa que aplica localização a seus produtos;

⁵³ De fato, até onde esta pesquisa pôde observar, três empresas mantêm suas bases terminológicas abertas ao público. São elas Microsoft, Symantec e SAP. A ideia inicial deste trabalho era contrastar amostras de duas empresas que aplicassem metodologia da área. A SAP abriu o acesso a seu acervo terminológico quando este trabalho já estava nos estágios finais de desenvolvimento. A Microsoft seria incluída; contudo, dado o volume de dados processável no respectivo *corpus*, optou-se por deixar essa análise para outros trabalhos. Além disso, o glossário Microsoft tem um escopo muito mais abrangente que o da Symantec. O primeiro tende a agrupar produtos de ordem diversa; o último está mais centrado em segurança de dados digitais.

- consiste em materiais que são produto da localização;
- é bilíngue no par de idiomas EN-US / PT-BR;
- é composto por três *subcorpora*:
 - glossário: contém entradas terminológicas
 - UI: contém *strings*⁵⁴ de *software*
 - ajuda do *software*: contém conteúdo em html.

As subseções a seguir descrevem um pouco mais cada um dos *subcorpora*, apresentando suas particularidades e possíveis pontos de limitação ou impacto na análise. Em seguida, é apresentado um quadro que resume as características do *corpus*.

4.1.1.2.1 Glossário

A Symantec abriu um espaço colaborativo em seu *site*, o *Norton Together*, no qual usuários em geral podem se cadastrar e, nos termos da empresa, permitir que estes se envolvam mais de perto com seus produtos favoritos da linha Norton – mediante a participação em atividades como tradução e revisão terminológica. Em suma, é uma maneira de as empresas fazerem *crowdsourcing*⁵⁵ sem aparentar muito, abrindo espaço para usuários colaborarem ou trabalharem com uma roupagem colaborativa. Muitas empresas adotam modelos parecidos, entre elas Microsoft, Google e Facebook. Em geral, o tipo de conteúdo disponibilizado para tradução tem um nível de confidencialidade baixo para a empresa, uma vez que os produtos já estão lançados no mercado ou o conteúdo já está publicado *on-line*.

Normalmente há uma contrapartida, contudo. Por exemplo, a empresa pode oferecer acesso gratuito a serviços, embora em geral essa contrapartida esteja disponível tanto para colaboradores quanto para não colaboradores. No caso do *Norton Together*, o retorno está na possibilidade de baixar gratuitamente pacotes de idiomas para qualquer *software* que faça parte dessa iniciativa. No final, todos os usuários se beneficiariam com seus produtos em seu idioma.

⁵⁴ Fragmento curto de texto ou sequência de caracteres que é usado para compor o conteúdo textual de um *software* (MACMILLAN, 2013a).

⁵⁵ Desenvolvimento de *software* que é realizado por um grande grupo de indivíduos, geralmente não vinculados a uma empresa ou entidade específica – ou seja, membros do público geral – via Internet. (MACMILLAN, 2013b).

A base terminológica entra nessa forma de interação como parte desse processo de tradução. Como a maior parte – senão a totalidade – das empresas que fazem localização, a Symantec também usa sua terminologia como instrumento prescritivo. Nesse sentido, o glossário está disponível para que os colaboradores sejam consistentes com a terminologia usada nos produtos Norton.

O glossário apresenta a terminologia adotada pela Symantec para seus produtos da linha Norton. Trata-se de entradas terminológicas no par de idiomas EN-US / PT-BR, sendo em sua maioria unidades nominais e sintagmáticas. Uma pequena parte das unidades nominais consiste em nomes próprios, em geral, nomes de produtos e tecnologias Symantec (como Norton AntiVirus, Symantec ServiceDesk e One Button Disconnect), mas também nomes de produtos de terceiros (como Internet Explorer e Authenticode, produtos da Microsoft). Em menor número, são listadas unidades verbais (*display / exibir* e *push / enviar*), seguidas por adjetivais (*quarantined / em quarentena* e *sensitive / confidencial*) e algumas adverbiais (*by default / por padrão* e *successfully / com êxito*). Entre as unidades nominais e verbais, encontram-se algumas UFE, como *backup / fazer backup* – e *quarantine / colocar em quarentena*. Também registra alguns elementos de UI, por exemplo, *Poor Trust* e *Bad Trust*. A figura a seguir mostra uma consulta à terminologia da Symantec no *Norton Together*:

Figura 12 - Consulta à terminologia da Symantec no *site Norton Together*

The screenshot shows the Symantec Language Portal interface. On the left is a navigation menu for 'Norton Together' with options like Home, Profile, Progress Overview, Support, Getting Started, FAQ, Guidelines, Community, Forum, and Language Portal. The main content area is titled 'Symantec Language Portal' and features a search bar with 'a' entered. Below the search bar, there are dropdown menus for 'Source Language' (English) and 'Target Language' (Portuguese, Brazil), and a 'Find Terminology' button. The search results are displayed in a table with columns for Source, Target, Context, and Comment.

Source	Target	Context	Comment
Discover scan	verificação de descoberta		
Fibre Transport	Fibre Transport	unable to determine if %s is a Fibre Transport server\n %s is not a Fibre Transport server\n %s is a Fibre Transport server\n	
Fibre Channel arbitrated loop	Fibre Channel Arbitrated Loop	Direct connect Fibre Channel arbitrated loop (FC-AL) configuration	
handshake	aperto de mão	Downstream TLS handshake successful TLS handshake with downstream MTA [0] failed. TLS handshake with downstream MTA [0] was successfully completed.	
Data Processing Service Connector	Data Processing Service Connector		
Media Manager	Gerenciador de mídia	Media Manager Also if both Media Manager and disk storage units are specified the duration of disk write	

Fonte: Elaborado pelo autor.

O glossário não consolida unidades relacionáveis sob a mesma entrada, nem traz qualquer associação ou vínculo entre elas. Todas são tratadas como termos separados. Por exemplo, agrupa as entradas segundo o modelo transformacional (p. 47), como se vê a seguir:

Figura 13 - Consulta a support no *site* Norton Together

Source	Target	Context	Comment
customer support	suporte ao cliente		
Global Support	Suporte global	Select your country or language under Global Support.'	
Help & Support	Ajuda e suporte		
help desk support	suporte de helpdesk		
MySupport	MySupport	MySupport is a secure interactive and personalized website that allows you to directly communicate with Enterprise Support via the web to initiate track update and close service requests as well as review your closed cases and their resolutions.'	
NetBackup Support matrix	Matriz de suporte do NetBackup (*)		
pre-sales support	suporte a pré-vendas		
Onsite Support Assistance	Ajuda de Suporte Local		
One Click Support	Suporte em um só clique		
Platinum support	suporte Platinum		
online support	suporte online		
Norton Support	Suporte Norton		
Premium Support Services	Serviços Premium Platinum		
technical support	suporte técnico		
supported	suportado		
support agreement	contrato de suporte		
support	suporte		
support	suportar		

Fonte: Elaborado pelo autor.

Como se pode observar, as formas superficiais de uma mesma unidade (*suporte*, *suportar* e *suportado*) são todas registradas separadamente. Ortograficamente, o glossário registra unidades comuns com letras minúsculas, e grafa em caixa-alta as iniciais de nomes próprios. Algumas unidades apresentam variação ortográfica. Embora fique latente que se trate da mesma unidade, o glossário as separa em duas entradas. Por exemplo:

- *autofill* e *auto-fill*.

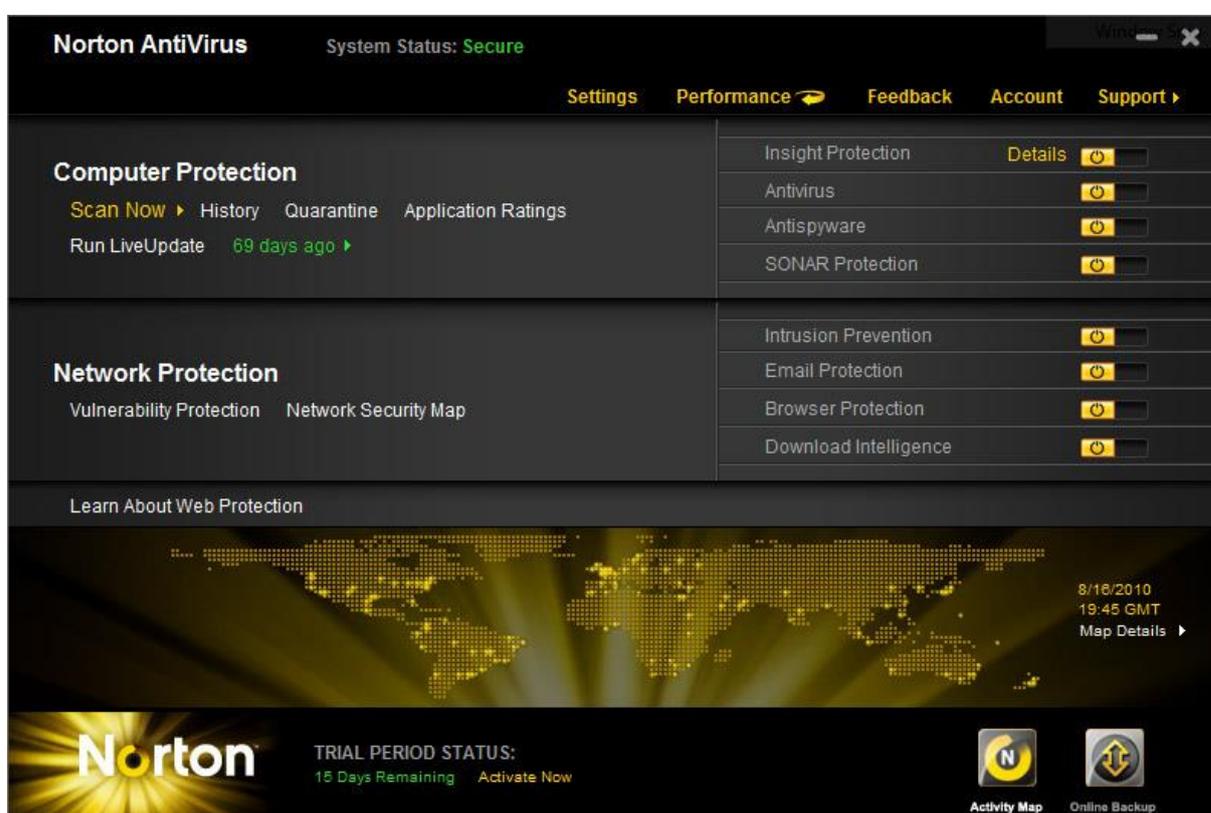
No entanto, essas informações não constam no glossário, sendo todas produto de análise deste trabalho. A única informação extra apresentada são frases (apenas em inglês) para

contextualizar a entrada. Não há informações gramaticais (classe, por exemplo). O usuário identificaria essa informação pela leitura de cada entrada nos dois idiomas.

4.1.1.2.2 Interface de usuário

Este *corpus* consiste em *strings* textuais do *software* Norton AntiVirus. Este *software* foi escolhido para análise dado seu enfoque em produtos voltados para consumidor final. A seguinte figura mostra uma tela desse *software*:

Figura 14 - Tela do Norton AntiVirus 2011



Fonte: Elaborado pelo autor.

Em geral, traduzir uma interface de usuário consiste em extrair suas *strings* e processá-las algum outro *software*, como uma ferramenta CAT. Por se tratar de texto fora de seu ambiente original, em certo sentido, trata-se de conteúdo traduzido fora de contexto. Embora existam ferramentas CAT no mercado que permitam a tradução de interfaces dentro de seu contexto, dependendo do caso, traduzir uma lista de *strings* fora de contexto pode ser a reali-

dade do tradutor trabalhando em um projeto desse tipo (cf. a *Figura 16* na página 128 para ver como as *strings* do *software* são apresentadas ao tradutor).

Em geral, o objetivo de uma UI é ajudar o usuário a navegar pelas funções do *software*. Em função disso, os elementos de interação mais comuns de encontrar são itens de menu⁵⁶, dicas de ferramenta, botões, opções e mensagens. A seguinte tabela traz uma breve síntese desses elementos, mas não pretende ser uma lista definitiva nem extensiva dos elementos que podem ser encontrados em uma interface:

Tabela 13 - Elementos textuais típicos em interfaces gráficas

Elemento de interface	Tipo de conteúdo textual
Item de menu	<p>Geralmente sintagmas nominais ou verbais no infinitivo. Exemplos:</p> <ul style="list-style-type: none"> • Scan / Verificar • Setup / Configuração
Dica de ferramenta	<p>Sintagmas curtos, em geral verbais. Exemplos:</p> <p>Copies all contents of this screen to the Clipboard.</p> <p>Copia todo o conteúdo desta tela para a Área de transferência.</p>
Opções	<p>Predominantemente sintagmas nominais, mas também apresentam alguns sintagmas verbais. Exemplos:</p> <p>Don't ask me again until I restart my computer.</p> <p>Não perguntar novamente até que o computador seja reiniciado.</p>
Botões	<p>Predominantemente sintagmas verbais, mas também apresentam alguns sintagmas nominais. Exemplos:</p> <ul style="list-style-type: none"> • Add to Quarantine / Adicionar à quarentena • Scan Now / Verificar agora

⁵⁶ Inclui também ícones. Mas estes também apresentam algum tipo de texto, por exemplo, em dica de ferramenta.

Elemento de interface

Tipo de conteúdo textual

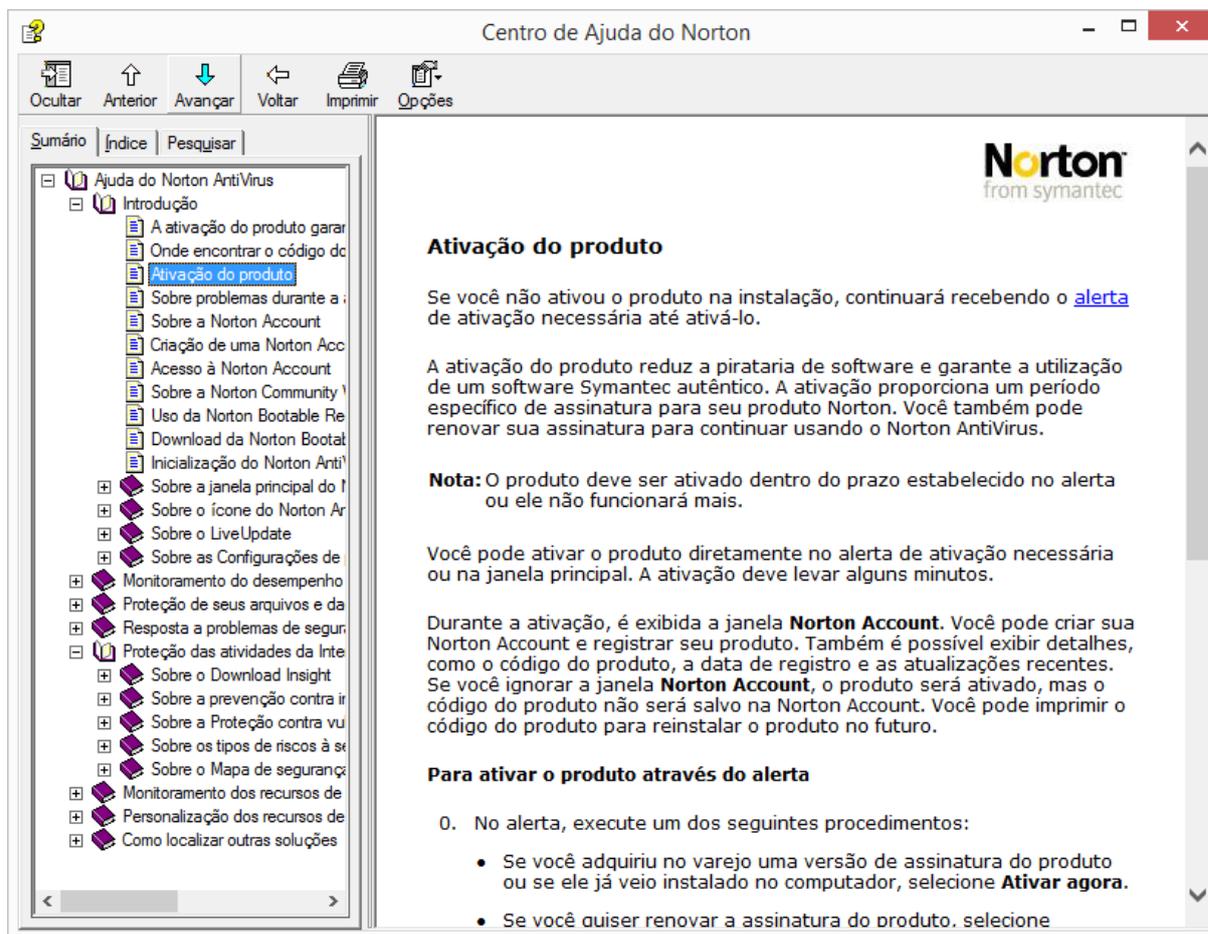
Mensagens	<p>Frases simples que, normalmente, informam o usuário sobre o resultado decorrente de uma ação, sobre um erro ou sobre os riscos envolvidos em executar determinada ação. As mensagens compõem a parte mais extensa do conteúdo textual de um <i>software</i>, e, em geral, sua extensão varia desde uma frase pequena, até um pequeno parágrafo. Exemplos:</p> <p style="padding-left: 40px;">Security activity has occurred during Silent Mode.</p> <p style="padding-left: 40px;">Ocorreu uma atividade de segurança durante o modo silencioso.</p> <hr style="border-top: 1px dotted #000;"/> <p style="padding-left: 40px;">Are you sure you wish to clear all AntiSpam logs?</p> <p style="padding-left: 40px;">Tem certeza de que deseja limpar todos os logs do antispam?</p> <hr style="border-top: 1px dotted #000;"/> <p style="padding-left: 40px;">After performing a manual removal, run Quick Scan to confirm a successful removal.</p> <p style="padding-left: 40px;">Após realizar a remoção manual, execute a Verificação rápida para confirmar o êxito da remoção.</p>
Outros elementos	<p>Também existem outros elementos em uma UI. Por exemplo, indicadores sobre o status de determinada tarefa. Esses indicadores podem consistir em adjetivos. Exemplos:</p> <ul style="list-style-type: none"> • Detected / Detectado • Cancelled / Cancelado • Duplicated / Duplicado • Moderated / Moderado

Fonte: Elaborado pelo autor.

4.1.1.2.3 Ajuda

Este *subcorpus* consiste na documentação de ajuda para o *software* Norton AntiVirus. Sua função é ensinar a usá-lo, trazendo os procedimentos para desempenhar determinadas atividades com o *software*, explicação para mensagens de erro, etc. A ajuda tem o objetivo de educar o usuário e ajudá-lo a configurar ou a usar o produto. A seguinte figura ilustra como a ajuda é apresentada ao usuário:

Figura 15 - Ajuda do Norton AntiVirus 2011



Fonte: Elaborado pelo autor.

Esse *subcorpus* tem uma apresentação mais tradicional, ou seja, está em um formato mais de acordo com o que se costuma designar um *texto* como produto. Apresenta diversos capítulos, seções e subseções, títulos, parágrafos e incorpora elementos como listas numeradas ou com marcadores. Em geral, trata-se de material instrutivo, isto é, espera-se uma boa ocorrência de sentenças no imperativo, títulos que descrevem algum processo do *software* (foco em processos) e listas de instruções.

4.1.1.2.4 Restrições quanto ao uso dos *subcorpora* na análise

Talvez o fator que mais tenha o potencial de interferir na análise deste trabalho é a data de publicação dos *subcorpora*. Existe uma disparidade de datas. O glossário Symantec foi coletado em 2013 e traz a terminologia mais atualizada. O *software* e a respectiva ajuda, por

sua vez, datam de 2011. Em função disso, é possível que erros sejam introduzidos na lista do QA terminológico. O motivo seria a dinâmica de atualização de termos que ocorre no setor da localização ou, mais especificamente, das traduções dos termos originais. Em geral, os termos traduzidos são mais voláteis, e podem ter ocorrido atualizações ao longo dos últimos dois anos que não se refletem no *software* antigo. De fato, apareceram alguns casos. Para um exemplo disso, confira a descrição para as ocorrências de *online* e *on-line* (p. 194) nos resultados para a análise tradutória, em *Erros de tradução* → seção *Ortografia e pontuação* (p. 194).

4.1.1.3 Síntese do *corpus*

A tabela a seguir apresenta uma síntese do *corpus* coletado para análise, incluindo alguns detalhes a respeito das fontes.

Tabela 14 - Sumário com informações sobre cada *subcorpus*

Subcorpus	Detalhes
	<i>Software Norton AntiVirus 2011</i> (versão 18.1)
	<i>Modo de obtenção</i> Processamento com a ferramenta CAT SDL Passolo 2011
Interface de usuário	<i>Informações gerais sobre o produto</i> (versão 2013) ⁵⁷ http://br.norton.com/antivirus/
	<i>Licença de uso</i> ⁵⁸ https://www.symantec.com/content/en/us/about/media/NAV-NIS-2011-Subscription-EULA-USE.pdf
	Documentação de ajuda offline que acompanha o <i>software</i> acima
Ajuda	<i>Modo de obtenção</i> Alinhamento via SDL Trados WinAlign 2007
	Terminologia oficial Symantec para a linha de produtos Norton (AntiVirus, Online Backup, Internet Security, entre outros)
Glossário	Inclui um campo que registra o contexto em inglês, mas nenhum outro indicativo de produto, definição, classe gramatical, etc., para os termos ⁵⁹
	<i>Modo de obtenção</i> Download mediante extração de dados do portal de idiomas Symantec no seguinte link (requer cadastramento gratuito) ⁶⁰ : http://together.norton.com/symtd/search

Fonte: Elaborado pelo autor.

Já a tabela abaixo contém informações sobre o volume de dados encontrado em cada *subcorpus*.

⁵⁷ Fonte: Symantec (2013a).

⁵⁸ Fonte: Symantec (2011a).

⁵⁹ Vale dizer que, normalmente, as bases de dados terminológicas de empresas desse porte trazem mais informações; ou seja, as empresas fazem a escolha de partilhar apenas parte da informação com o público geral.

⁶⁰ Fonte: Symantec (2013b).

Tabela 15 - Volume de palavras por *subcorpus*

<i>Subcorpus</i>	Volume		
	<i>Segmentos</i>	<i>Palavras em EN-US</i>	<i>Palavras em PT-BR</i>
Interface de usuário	3677	27 444	30 655
Ajuda	5837	74 600	80 695
Glossário	4578	10 578	13 635

Fonte: Elaborado pelo autor.

Enfim, para o glossário, ainda há alguns cruzamentos específicos que é possível fazer, como a relação de entradas compostas no original que são traduzidas para entradas simples e vice-versa. A seguinte tabela apresenta os detalhes:

Tabela 16 - Relação de entradas simples e compostas no glossário

	Glossário	
	EN-US	PT-BR
Entradas simples	797	691
Entradas compostas	3 794	3 900
Relação composta (EN-US):simples (PT-BR)	45	
Relação simples (EN-US):composta (PT-BR)	151	

Fonte: Elaborado pelo autor.

4.1.2 Ferramentas

Algumas ferramentas computacionais foram necessárias para realizar algumas das tarefas envolvidas na coleta e análise dos dados. As subseções a seguir apresentam os detalhes sobre cada uma.

4.1.2.1 Ferramentas para coleta de dados do *corpus*

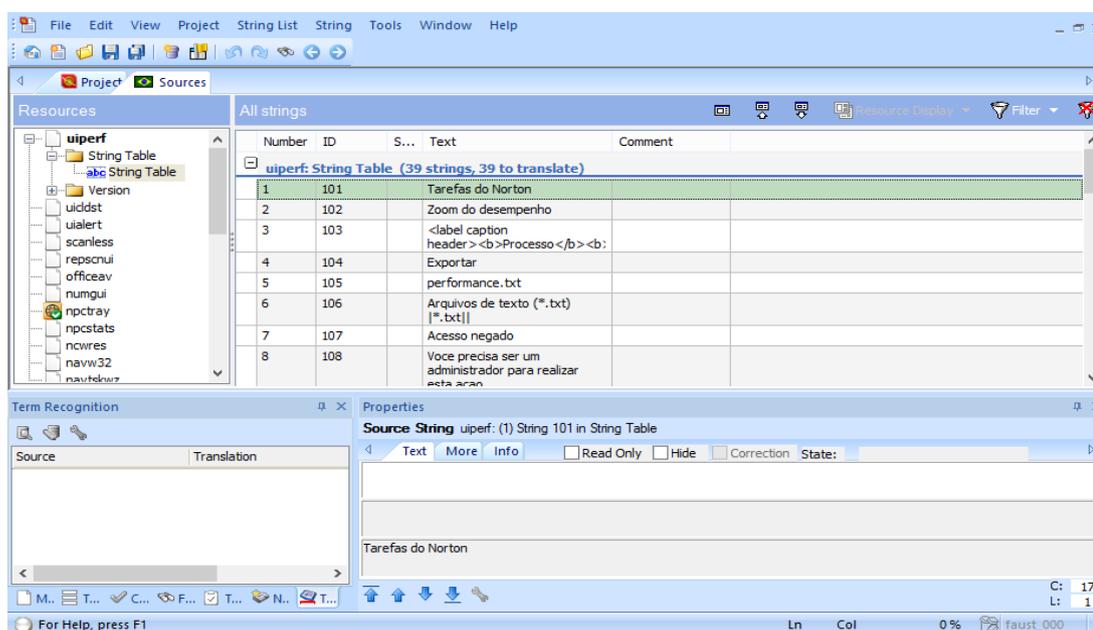
Em virtude dos formatos de arquivo da interface de usuário e da ajuda do *software* selecionado para análise, foi necessário empregar algumas ferramentas que permitissem extrair e alinhar o conteúdo textual desses arquivos. Para coletar as *strings* do *software*, foi usada a

ferramenta SDL Passolo 2011; em seguida, o Microsoft Excel foi usado para alinhar os segmentos de texto extraídos. Já para coletar e alinhar o conteúdo textual da ajuda, foi usado o SDL Trados WinAlign 2007.

4.1.2.1.1 SDL Passolo 2011

O SDL Passolo versão 2011⁶¹ é uma ferramenta CAT voltada à tradução de interfaces de usuário. Esse *software* gera uma prévia de caixas de diálogo, menus, etc., colocando as *strings* em seu contexto de uso. A ferramenta permite exportar as *strings* em variados formatos, como xliif⁶², bem como copiá-las para a área de transferência do Windows. A seguinte imagem mostra a tela de tradução de *strings* da ferramenta, com segmentos de texto pertencentes ao Norton AntiVirus 2011:

Figura 16 - SDL Passolo 2011 exibindo strings do Norton AntiVirus 2011



Fonte: Elaborada pelo autor.

Como foi afirmado acima, esta ferramenta foi usada para copiar as *strings* do *software* nas versões EN-US e PT-BR. Posteriormente, com a ajuda do Microsoft Excel (cf. abaixo), essas strings foram consolidadas em um único arquivo bilíngue.

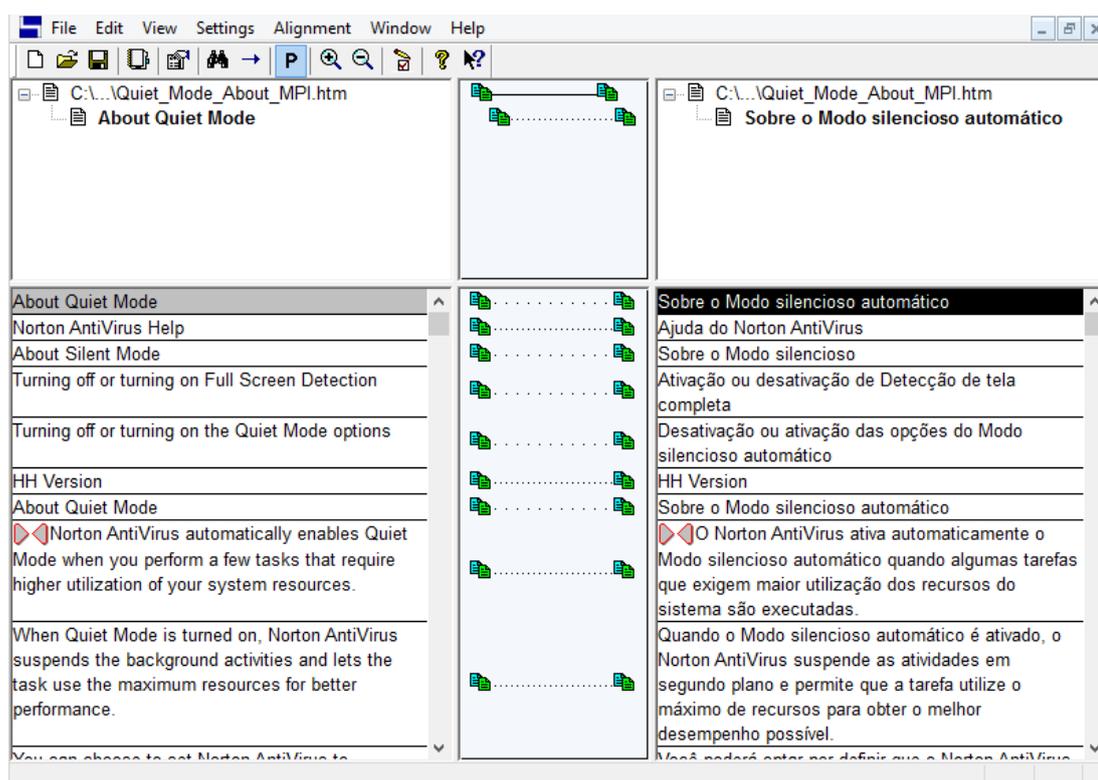
⁶¹ Mais informações sobre a ferramenta em <http://www.sdl.com/products/sdl-passolo/>

⁶² Formato padrão do setor da localização para arquivos bilíngues (OPENTAG.COM, 2013c).

4.1.2.1.2 SDL Trados WinAlign 2007

O SDL Trados WinAlign versão 2007 é uma ferramenta de alinhamento de texto que faz parte do pacote SDL Trados Studio 2011⁶³. Esta ferramenta permite gerar um arquivo bilingue com base em pares de arquivos em idiomas distintos, executando o alinhamento de diversos formatos (entre eles, .html) com base em parâmetros formais (como marcas de parágrafo e pontuação). Além disso, permite exportar o conteúdo alinhado em algum formato bilingue (como o .tmx⁶⁴). A imagem abaixo apresenta a tela de alinhamento do *software*, com segmentos da ajuda do Norton AntiVirus 2011:

Figura 17 - SDL Trados WinAlign 2007 alinhando arquivos da ajuda do Symantec Norton AntiVirus 2011



Fonte: Elaborado pelo autor.

Neste trabalho, conforme exposto acima, a ferramenta foi usada para gerar um *corpus* bilingue com base nas versões da ajuda do *software* em EN-US e PT-BR.

⁶³ Ferramenta CAT da líder de mercado no segmento, a SDL. Mais informações sobre a ferramenta em <http://www.translationzone.com/products/sdl-trados-studio/>

⁶⁴ TMX é um formato padrão (derivado de XML) criado no setor da localização para uniformizar a troca de memórias de tradução entre ferramentas computadorizadas (OPENTAG.COM, 2013b).

4.1.2.2 Ferramentas para análise e manipulação dos dados

Algumas ferramentas foram necessárias para gerar a lista de QA objeto de estudo deste trabalho e para consolidar os dados para análise. O *software* ApSIC Xbench foi usado para simular um QA terminológico tal qual é realizado na área da localização. Com relação à consolidação dos resultados desse QA, foi usado o Microsoft Excel 2007.

4.1.2.2.1 ApSIC Xbench 2.9

O ApSIC Xbench versão 2.9 é um *software* de referência que incorpora à sua interface um concordanciador com recurso de expressões regulares e uma função de QA.

O motivo para a escolha desse *software* foi, em primeiro lugar, por ter uso gratuito e, em segundo, pela velocidade e praticidade. Trata-se de uma ferramenta que processa rapidamente grandes volumes de informação e retorna uma lista que pode ser exportada em outros formatos, para manipulação em outro *software*, como um editor de planilhas ou um banco de dados. Além disso, o Xbench aceita uma grande variedade de formatos de arquivo, inclusive texto simples separado por tabulações. Como alguns dos arquivos do *corpus* seriam gerados nesse formato, isso permitiu o uso direto dos arquivos sem que fosse necessário processar ou converter os arquivos.

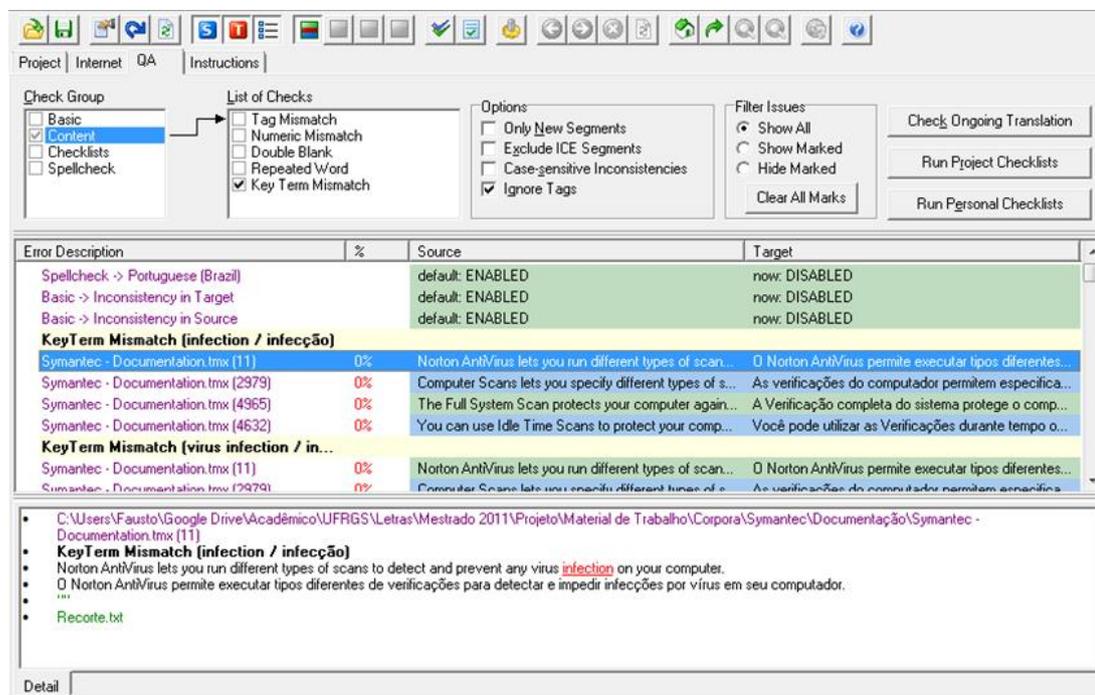
Outra opção seria usar uma ferramenta como o SDL Trados Studio 2011, o qual tem funções de QA terminológico mais sofisticadas, que usam *fuzzy matching*⁶⁵ ou rotinas de processamento (por exemplo, dar prioridade à verificação de unidades sintagmáticas maiores para não retornar ocorrências duplicadas). No entanto, essa ferramenta não gera uma lista exportada que inclua todas as informações necessárias para a análise realizada neste trabalho. Por exemplo, a lista não inclui os segmentos de texto em que ocorre determinado item.

O *software* fornece uma função de QA que inclui uma série de verificadores, como espaçamento duplo, inconsistência entre a tradução de segmentos idênticos, entre outros. Entre

⁶⁵ *Fuzzy matching* pode ser entendido como a capacidade de fazer cruzamentos de padrões que não sejam 100% precisos. Por exemplo, uma pesquisa em *fuzzy matching* por “abrir” pode retornar “abriam”, “abrir” ou ainda “fabril” como possíveis correspondentes, dependendo do algoritmo usado. Teoricamente, o uso de *fuzzy matching* em verificadores terminológicos tende a trazer mais resultados, sejam eles ruído ou não.

estes está a função *Key Term Mismatch*, a qual executa o QA terminológico. A figura a seguir

Figura 18 - Lista de validação terminológica gerada pelo *software* ApSIC Xbench



Fonte: Elaborado pelo autor.

ilustra a lista de validação que o *software* gera para terminologia:

Posteriormente, essa lista pode ser exportada em outro formato, como o .xls, para ser trabalhada em outros produtos de *software*. Neste trabalho, o Xbench foi usado para gerar a lista de QA e exportá-la no formato .xls.

4.1.2.2.2 Microsoft Excel 2007

O Microsoft Excel versão 2013 foi amplamente usado neste trabalho, especialmente na consolidação dos dados obtidos pela execução do QA. Além disso, também foi empregado na coleta de dados, para consolidar as *strings* do *software* em um arquivo bilíngue. O recurso de tabelas dinâmicas do *software* foi fundamental para quantificar e filtrar os resultados segundo os critérios de análise.

4.1.2.3 Síntese das ferramentas

A tabela a seguir traz a relação das ferramentas usadas e o respectivo uso.

Tabela 17 - Ferramentas eletrônicas usadas na coleta e análise de dados

Ferramenta	Descrição	Uso
SDL Passolo 2011	Ferramenta CAT usada especificamente para traduzir strings de <i>software</i> .	Aplicada em arquivos que compõem a interface de usuário do <i>software</i> para coletar parte do <i>corpus</i> bilíngue.
SDL Trados WinAlign 2007	Ferramenta de alinhamento de textos em idiomas diferentes.	Aplicada em conteúdo no formato .html referente à ajuda do <i>software</i> , para gerar parte do <i>corpus</i> bilíngue.
ApSIC Xbench 2.9	Ferramenta de tem função de concordanciador e de QA.	Aplicada na execução do QA terminológico e exportação dos resultados em fomato do Microsoft Excel.
Microsoft Excel 2007	Ferramenta de planilhas eletrônicas.	Usada para catalogar e quantificar os resultados do QA terminológico. Também foi usada para alinhar os segmentos textuais extraídos do <i>software</i> .

Fonte: Elaborado pelo autor.

4.2 Metodologia

Nesta seção, são apresentados os processos empregados para coletar os dados do *corpus* e gerar a lista de validação terminológica e analisar os dados resultantes. Para a geração da lista, a seção descreve como seus dados foram organizados para a análise. Para a etapa de análise, são introduzidos os critérios e parâmetros que foram adotados para a classificação dos dados.

4.2.1 Coleta de dados para o *corpus*

Para a coleta de dados, foram usados três processos distintos, dadas as particularidades de apresentação e armazenamento de cada *subcorpus*. As subseções a seguir detalham cada um dos processos de coleta.

4.2.1.1 Glossário

O glossário⁶⁶ foi baixado mediante a cópia direta de dados do portal de idiomas Symantec. Os dados eram recuperáveis mediante consultas que retornavam as entradas terminológicas em tabelas html. As entradas terminológicas copiadas desse portal foram consolidadas em formato tabular para uso na ferramenta de QA terminológico⁶⁷.

Os seguintes passos foram executados para gerar o glossário bilíngue:

1. acessar a pesquisa terminológica no portal de idiomas Symantec;
2. fazer consultas para cada letra do alfabeto em inglês⁶⁸ até retornar todas as entradas possíveis;
 - a. a cada consulta, copiar a tabela html retornada pelo portal e colá-la no Microsoft Excel;
3. no Microsoft Excel, fazer uma busca por duplicatas;
4. eliminar todas as duplicatas encontradas, mantendo apenas entradas exclusivas⁶⁹;
5. salvar o arquivo em formato .txt separado por tabulações.

⁶⁶ Consulte *Glossário* (p. 118) para mais detalhes sobre o glossário.

⁶⁷ Para informações gerais sobre o processo de QA terminológico, cf. *TEP QA em detalhes* (p. 31); para detalhes sobre a ferramenta de QA terminológico usada neste trabalho, cf. *ApSIC Xbench 2.9* (p. 130).

⁶⁸ Por exemplo, ao consultar “a”, a página retornava todas as entradas que continham a letra “a” em qualquer parte de sua forma – como *handshake*, *Fiber Transport* e *technical support*.

⁶⁹ Como as consultas ao portal eram muito abrangentes, foi necessário buscar as duplicatas a fim de eliminá-las, resultando em um glossário mais próximo da base *on-line* mantida pela Symantec.

4.2.1.2 Interface de usuário

As *strings* de *software* foram copiadas do próprio Norton AntiVirus em suas versões em EN-US e PT-BR através do *software* SDL Passolo 2011⁷⁰. Como havia duas versões diferentes para o Norton AntiVirus, as *strings* copiadas de cada uma foram alinhadas com a ajuda do Microsoft Excel. O texto copiado estava na mesma ordem nos dois idiomas, o que facilitou a tarefa. O conjunto bilíngue de *strings* foi salvo em formato .txt separado por tabulações para processamento na ferramenta de QA terminológico.

Os seguintes passos foram executados para gerar um arquivo bilíngue com as *strings* do *software*:

1. no SDL Passolo, abrir os arquivos do *software* que continham as *strings* em EN-US e PT-BR;
2. para cada arquivo, selecionar todo o conteúdo da coluna que armazena as *strings* e copiá-lo;
3. no Microsoft Excel, colar o conteúdo em EN-US em uma coluna e o conteúdo em PT-BR na coluna ao lado;
4. salvar o arquivo em formato .txt separado por tabulações.

4.2.1.3 Ajuda

A documentação de ajuda acompanhava o *software* em um arquivo .chm, que agrupa vários arquivos .html internamente. Havia um arquivo .chm para cada idioma. Os respectivos arquivos .html foram copiados e alinhados com o uso do *software* SDL Trados WinAlign⁷¹. Como esse *software* usa algoritmos para fazer o alinhamento, havia maior possibilidade de que ocorressem problemas de segmentação nesse *corpus* (cf. *Alinhamento* na página 158 para mais detalhes sobre os problemas de alinhamento encontrados). O conjunto alinhado da ajuda foi salvo em formato .tmx para processamento na ferramenta de QA terminológico.

⁷⁰ Consulte *Interface de usuário* (p. 121) e *SDL Passolo 2011* (p. 128) para informações sobre a interface de usuário e o *software* usado em sua extração.

⁷¹ Consulte *Ajuda* (p. 123) e *SDL Trados WinAlign 2007* (p. 129) para informações sobre a ajuda e o *software* usado em sua extração.

Os seguintes passos foram executados para gerar um arquivo bilíngue com o conteúdo textual da ajuda:

1. no SDL Trados WinAlign, criar um projeto de alinhamento com estas configurações:
 - a. *Source Language: English (United States)*;
 - b. *Target Language: Portuguese (Brazil)*;
 - c. *File Type: HTML Files*;
2. na guia *Files*:
 - a. em *English (United States)*, inserir os arquivos .html referentes da ajuda em inglês;
 - b. em *Portuguese (Brazil)*, inserir os arquivos .html referentes da ajuda em português;
 - c. clicar em *Align File Names* para que o *software* alinhe os arquivos automaticamente com base no nome⁷²;
3. na guia *Export*, selecionar *Translation Memory Exchange Format (TMX)* como a opção de exportação do arquivo alinhado resultante;
4. executar o alinhamento dos arquivos;
5. exportar o resultado em um arquivo no formato .tmx.

4.2.2 Execução do QA terminológico

Nesta etapa, uma verificação automatizada de terminologia foi executada para gerar duas listas de validação:

- a lista de validação original gerada pelo *software*, ou seja, contendo os itens que o tradutor deveria conferir;
- uma lista de validação ajustada especialmente para conter também os itens que o *software* deveria ignorar..

⁷² Como cada par de arquivo .html em EN-US e PT-BR tem o mesmo nome para cada idioma, a ferramenta consegue identificar facilmente qual arquivo em inglês deve ser comparado com qual arquivo em português, assim facilitando o alinhamento dos segmentos neles contidos.

4.2.2.1 Geração da lista de validação terminológica original do *software* de QA

A lista de validação foi gerada com o *software* ApSIC Xbench, configurado para executar a função *Key Term Mismatch* com a configuração *Ignore tags* ligada. Essa configuração garante que a formatação do material em *tags* não interfira no processo de detecção. Os seguintes passos foram executados para gerar a lista:

1. criar um projeto no ApSIC Xbench;
2. incluir o glossário e defini-lo como *Key Terms*; essa configuração faz com que as entradas desse *subcorpus* sejam usadas como padrão a ser casado entre o original e a tradução;
3. incluir a UI e a ajuda e defini-las como *Ongoing Translation*; essa configuração faz com que os segmentos desses dois *subcorpora* sejam usados como o material que deve ter os padrões contidos no glossário;
4. na guia *QA*, selecionar apenas a função *Content -> Key Term Mismatch* e a opção *Ignore Tags*;
5. executar a verificação por meio do botão *Check Ongoing Translation*;
6. exportar a lista em formato Microsoft Excel (.xls) via opção *Export QA Results*.

A seguinte figura mostra o arquivo .xls gerado pelo *software*:

Figura 19 - Exerto da lista de validação de QA gerada e exportada via ApSIC Xbench em formato .xls

Exported QA Report		Generated by	
		ApSIC Xbench 2.9 Build 474	
		http://www.apsic.com/xbench	
QA Options in non-default state			
<i>Basic -> Untranslated Segments (0)</i>		default: ENABLED	now: DISABLED
<i>Content -> Tag Mismatch (0)</i>		default: ENABLED	now: DISABLED
<i>Content -> Numeric Mismatch (0)</i>		default: ENABLED	now: DISABLED
<i>Content -> Double Blank (0)</i>		default: ENABLED	now: DISABLED
<i>Content -> Repeated Word (0)</i>		default: ENABLED	now: DISABLED
<i>Spellcheck -> Portuguese (Brazil) (0)</i>		default: ENABLED	now: DISABLED
<i>Basic -> Inconsistency in Source (0)</i>		default: ENABLED	now: DISABLED
<i>Basic -> Inconsistency in Target (0)</i>		default: ENABLED	now: DISABLED
KeyTerm Mismatch (infection / infecção)			
<i>NAV_Ajuda.tmx (11)</i>	0%	Norton AntiVirus lets you run different types of scans to detect and prevent any virus infection on your computer.	O Norton AntiVirus permite executar tipos diferentes de verificações para detectar e impedir infecções por vírus em seu computador. <i>Symantec_Glossário.txt</i>
<i>NAV_Ajuda.tmx (2979)</i>	0%	Computer Scans lets you specify different types of scans to detect and prevent any virus infection on your computer.	As verificações do computador permitem especificar tipos diferentes de verificações para detectar e impedir infecções por vírus em seu computador. <i>Symantec_Glossário.txt</i>
<i>NAV_Ajuda.tmx (4965)</i>	0%	The Full System Scan protects your computer against infection without compromising the performance of your computer.	A Verificação completa do sistema protege o computador contra infecções sem comprometer o desempenho. <i>Symantec_Glossário.txt</i>
<i>NAV_Ajuda.tmx (4632)</i>	0%	You can use Idle Time Scans to protect your computer against infection without compromising the performance of your computer.	Você pode utilizar as Verificações durante tempo ocioso para proteger o computador contra infecções sem comprometer o desempenho. <i>Symantec_Glossário.txt</i>
KeyTerm Mismatch (virus infection / infecção por vírus)			
<i>NAV_Ajuda.tmx (11)</i>	0%	Norton AntiVirus lets you run different types of scans to detect and prevent any virus infection on your computer.	O Norton AntiVirus permite executar tipos diferentes de verificações para detectar e impedir infecções por vírus em seu computador. <i>Symantec_Glossário.txt</i>
<i>NAV_Ajuda.tmx (2979)</i>	0%	Computer Scans lets you specify different types of scans to detect and prevent any virus infection on your computer.	As verificações do computador permitem especificar tipos diferentes de verificações para detectar e impedir infecções por vírus em seu computador. <i>Symantec_Glossário.txt</i>

Fonte: Elaborado pelo autor.

4.2.2.2 Geração de uma lista de validação contendo os itens que o *software* de QA ignora

Para possibilitar a análise das ocorrências consideradas acertos pelo *software* de QA, ou seja, os itens excluídos da lista de validação, foram agregados resultados extras à lista original do *software*. Para tal fim, os seguintes passos foram feitos:

1. criar uma cópia exata do glossário;
2. substituir a contraparte em português brasileiro por **[abc000]**, um conteúdo que seguramente não existe na UI nem na ajuda;
3. repetir os passos 1-6 de *Geração da lista de validação terminológica* acima, agora usando essa cópia do glossário no lugar do original;
4. A seguinte figura mostra o arquivo .xls gerado pelo *software*:

Figura 20 - Exerto da lista de validação adaptada para incluir todos os itens processados pelo *software* de QA

Exported QA Report		Generated by	
		ApSIC XBench 2.9 Build 474	
		http://www.apsic.com/xbench	
QA Options in non-default state			
<i>Basic -> Untranslated Segments (0)</i>		default: ENABLED	now: DISABLED
<i>Basic -> Inconsistency in Source (0)</i>		default: ENABLED	now: DISABLED
<i>Basic -> Inconsistency in Target (0)</i>		default: ENABLED	now: DISABLED
<i>Content -> Tag Mismatch (0)</i>		default: ENABLED	now: DISABLED
<i>Content -> Numeric Mismatch (0)</i>		default: ENABLED	now: DISABLED
<i>Content -> Double Blank (0)</i>		default: ENABLED	now: DISABLED
<i>Content -> Repeated Word (0)</i>		default: ENABLED	now: DISABLED
KeyTerm Mismatch (virus infection / [abc000])			
<i>NAV_Ajuda.tmx (11)</i>	0%	Norton AntiVirus lets you run different types of scans to detect and prevent any virus infection on your computer.	O Norton AntiVirus permite executar tipos diferentes de verificações para detectar e impedir infecções por vírus em seu computador. <i>Symantec_Glossário_Alterado.txt</i>
<i>NAV_Ajuda.tmx (2979)</i>	0%	Computer Scans lets you specify different types of scans to detect and prevent any virus infection on your computer.	As verificações do computador permitem especificar tipos diferentes de verificações para detectar e impedir infecções por vírus em seu computador. <i>Symantec_Glossário_Alterado.txt</i>
<i>NAV_Ajuda.tmx (3618)</i>	0%	Norton Online Backup provides a secure online backup solution that safeguards your important data against system crash, accidental deleting, virus infection , and other disasters.	O Norton Online Backup proporciona uma solução de backup on-line segura que protege seus dados importantes contra travamento do sistema, exclusão acidental, infecção por vírus ou outros desastres. <i>Symantec_Glossário_Alterado.txt</i>
KeyTerm Mismatch (security level / [abc000])			
<i>NAV_Ajuda.tmx (43)</i>	0%	It provides enhanced security level from the time your start your computer.	Essa opção fornece um nível de segurança aprimorado quando você inicia o computador. <i>Symantec_Glossário_Alterado.txt</i>
<i>NAV_Ajuda.tmx (2328)</i>	0%	Here, you can view the security level of the Web site from where the file was downloaded.	Nessa página, é possível visualizar o nível de segurança do site de onde o arquivo foi obtido por download. <i>Symantec_Glossário_Alterado.txt</i>
<i>NAV_Ajuda.tmx (1129)</i>	0%	The Enable Boot Time Protection option provides enhanced security level from the time you start your computer.	Além disso, a opção Ativar proteção do tempo de inicialização oferece nível de segurança aprimorado desde o momento em que você inicia o computador. <i>Symantec_Glossário_Alterado.txt</i>

Fonte: Elaborado pelo autor.

Como se pode observar em comparação com a figura anterior, agora o equivalente em PT-BR das entradas do glossário apresenta a sequência [abc000] no lugar do termo original.

Com isso, dois arquivos foram gerados no formato .xls, um com a lista original do *software* de QA e outra que inclui todos os segmentos que passaram pela análise deste.

5.

4.2.3 Preparação e organização dos dados da lista de QA

Para permitir o armazenamento correto das informações que seriam coletadas durante a análise da lista de QA, foi necessário adaptar a lista gerada pelo *software*. Para isso, os arquivos.xls gerados na execução do QA terminológico acima foram ajustados e consolidados em um único arquivo no formato .xlsx. Os passos a seguir descrevem em detalhes como isso foi feito:

1. no Microsoft Excel, abrir os dois arquivos .xls gerados na etapa anterior;
2. copiar o conteúdo de ambos, excluindo as linhas de cabeçalho, em uma nova planilha .xlsx;
3. buscar e remover as duplicatas⁷³;
4. no novo arquivo .xlsx, criar cabeçalhos para cada uma das colunas para as informações existentes, ou seja, aquelas já retornadas pelo *software* de QA, e também para aquelas a serem preenchidas durante a análise:
 - a. colunas existentes para as seguintes informações: número do segmento de tradução da ocorrência, *subcorpus*, entrada do glossário envolvida, contexto da ocorrência em EN-US, contexto da ocorrência em português (PT-BR);
 - b. colunas para preencher durante a análise: classes gramaticais das unidades do glossário, classes gramaticas da ocorrência segundo o contexto, resposta do *software* de QA para a ocorrência, tipo, categoria e subcategoria da ocorrência, e indicativo se a unidade da ocorrência já era glossariada;
5. para cada linha da planilha, preencher a coluna *Subcorpus* segundo o arquivo listado na coluna *Segmento*⁷⁴;

⁷³ Isso foi necessário para que a lista contivesse apenas os itens que o *software* de fato processou, sem duplicações, assegurando assim a integridade dos resultados quantificados e filtrados mais tarde.

⁷⁴ Como o nome do arquivo se refere especificamente a um dado *subcorpus*, ele foi usado como base para denominar o *subcorpus* de cada ocorrência e, posteriormente durante a análise, filtrar mais facilmente os resultados.

6. copiar, mediante processo automatizado, os termos listados nos cabeçalhos *Key Term Mismatch* dos arquivos .xls originais e colá-los nas respectivas linhas referentes ao termo em questão⁷⁵;
7. classificar as ocorrências da lista original do *software* de QA como *Inclui na lista*;
8. classificar as demais ocorrências, ou seja, aquelas que o *software* ignora, como *Exclui da lista*;

A seguinte figura ilustra como ficou a apresentação final do conteúdo, antes da análise dos dados:

⁷⁵ Como se pôde observar nas linhas das figuras apresentadas na seção anterior, o termo era exibido apenas na primeira linha de uma série de ocorrências. Este passo foi necessário para facilitar a filtragem dos resultados após a análise dos dados.

Figura 21 - Tabela de registro dos dados para análise

Segmento	Subcorpus	Entrada do glossário	Inglês (EN-US)	Português (PT-BR)	GR Glossário (EN-US)	GR Glossário (PT-BR)	GR Segmento (EN-US)	GR Segmento (PT-BR)	Resposta do software de QA	Tipo de ocorrência	Categoria	Sub-categoria	Glossariada?	Comentários
Symantec - Ajuda.tmx (11)	Ajuda	infection / infecção	Norton A...	O Norton...										
Symantec - Ajuda.tmx (2979)	Ajuda	infection / infecção	Computer...	As verific...										
Symantec - Ajuda.tmx (4965)	Ajuda	infection / infecção	The Full S...	A Verifica...										
Symantec - Ajuda.tmx (4965)	Ajuda	infection / infecção	The Full S...	A Verifica...										
Symantec - Ajuda.tmx (4632)	Ajuda	infection / infecção	You can u...	Você pode...										
Symantec - Ajuda.tmx (4632)	Ajuda	infection / infecção	You can u...	Você pode...										
Symantec - Ajuda.tmx (11)	Ajuda	virus infection / infecção por vírus	Norton An...	O Norton...										
Symantec - Ajuda.tmx (2979)	Ajuda	virus infection / infecção por vírus	Computer...	As verifica...										

Fonte: Elaborado pelo autor.

A tabela apresenta as seguintes informações:

- **Segmento:** identifica o segmento de tradução no qual o erro potencial apontado pelo *software* ocorre;
- **Subcorpus:** indica o *subcorpus* no qual a ocorrência aparece. Dois valores são possíveis:
 - **Ajuda** para a documentação de ajuda;
 - **UI** para a interface de usuário;
- **Entrada do glossário:** lista a entrada glossariada que originou a ocorrência;
-
- **Inglês (EN-US):** lista a contraparte em inglês do segmento de tradução no qual a ocorrência aparece;
- **Português (PT-BR):** lista a parte em português do segmento de tradução no qual a ocorrência aparece;
- **GR Glossário (EN-US):** indica a classe gramatical da unidade glossariada em inglês americano;
- **GR Glossário (PT-BR):** indica a classe gramatical da contraparte em português brasileiro;
- **GR Segmento (EN-US):** indica a classe gramatical da unidade em relação ao contexto encontrado na coluna *Inglês (EN-US)*;
- **GR Segmento (PT-BR):** indica a classe gramatical da unidade em relação ao contexto encontrado na coluna *Português (PT-BR)*;
- **Resposta do *software* de QA:** indica como a ocorrência foi tratada pelo *software*. Dois tipos possíveis:
 - **Inclui na lista de QA** indica que o *software* adicionou o item à lista como um possível erro de tradução para que o tradutor possa averiguar;
 - **Exclui da lista de QA** indica que o *software* considerou o item como correto e, portanto, não o inseriu na lista.
- **Tipo de ocorrência:** indica como a ocorrência foi classificada durante a análise dos dados. Dois tipos possíveis:
 - **Positivo** assinala que o item foi processado corretamente pelo *software*;
 - **Falso positivo** significa que o item foi processado incorretamente pelo *software*.

- **Categoria:** indica a classificação específica para determinada ocorrência. Para detalhes sobre as categorias específicas, cf. *Crítérios de análise* (p. 145);
- **Subcategoria:** apresenta a classificação específica a cada uma das categorias definidas no item acima. Para detalhes sobre as categorias específicas, cf. *Crítérios de análise* (p. 145);
- **Glossariada?:** indica se a unidade identificada na ocorrência consta em glossário ou não;
- **Comentários:** campo reservado para quaisquer observações não estruturadas que sejam pertinentes durante a análise.

O conteúdo em si dessa tabela não foi publicado na íntegra neste trabalho, mas foi usado de base para a apresentação dos resultados em *Análise dos dados e resultados* (p. 152).

Uma das aplicações dessa tabela foi a geração de tabelas dinâmicas no Microsoft Excel, para quantificar os resultados. Detalhes sobre o tipo de tabela que foi gerado estão em *Quantificação e filtragem dos resultados* na página 151. Essas tabelas dinâmicas estão publicadas no capítulo de análise e resultados citado acima.

4.2.4 Análise da lista de validação terminológica

O processo de análise consiste classificar cada uma das ocorrências da lista de validação terminológica, segundo os critérios apresentados na revisão teórica deste trabalho (cf. capítulo *Revisão da Literatura*, pp. 34-114), indicando se é positivo ou falso positivo. A análise considera tanto as ocorrências que o *software* inclui na lista de QA (ou seja, consideradas possíveis erros de tradução) quanto aquelas que são excluídas (ou seja, consideradas corretas do ponto de vista tradutório). O processo deverá retornar em números o volume de positivos e falsos positivos na lista, além de indicar alguns pontos que norteiem o desenvolvimento de verificadores mais eficazes, por exemplo, indicando o tipo de fenômeno linguístico que responde pela maior parte do ruído. As subseções a seguir descrevem os critérios e a execução da análise.

4.2.4.1 Critérios de análise

Conforme tratado na revisão teórica, foram estabelecidos critérios terminológicos e tradutórios para fundamentar a análise da lista de QA feita neste trabalho. Com base nesses critérios, foram criadas as seguintes classificações para as ocorrências da lista de QA:

- Classificação fraseológica, baseada na proposta de Bevilacqua (1999, 2004);
- Classificação variacional, baseada na tipologia de Freixa (2002);
- Classificação tradutória, baseada em Hurtado Albir (2001), especificamente nas noções de técnicas de tradução e erro de tradução.

Conforme visto na revisão teórica, cada aspecto específico coberto pelas propostas foi apurado, para verificar sua viabilidade como possível responsável por causar ruído na lista de QA. Com isso, a seguinte subclassificação foi elaborada, com base na classificação acima:

- Fraseológica:
 - Estabilidade sintática (p. 51);
 - Modelo transformacional (MT) (p. 47);
- Variacional:
 - Toda a tipologia de Freixa (cf. páginas 74-81, com uma tabela que resume os tipos e subtipos na página 75);
- Tradutória:
 - Técnicas de tradução (p. 109);
 - Erros de tradução (cf. 112);

Além disso, outros critérios foram arrolados para cobrir aspectos linguísticos mais fundamentais, como elipses, flexão verbal, concordância nominal, etc. Para fins de classificação, estes foram agrupados pelo nome *Linguística*. As subclassificações dessa categoria foram sendo relacionadas durante o andamento da análise de dados, mas alguns itens foram antevistos, como, por exemplo:

- Número;
- Gênero;
- Flexão verbal;
- Outra unidade, mesma forma;
- Referência à interface de usuário.

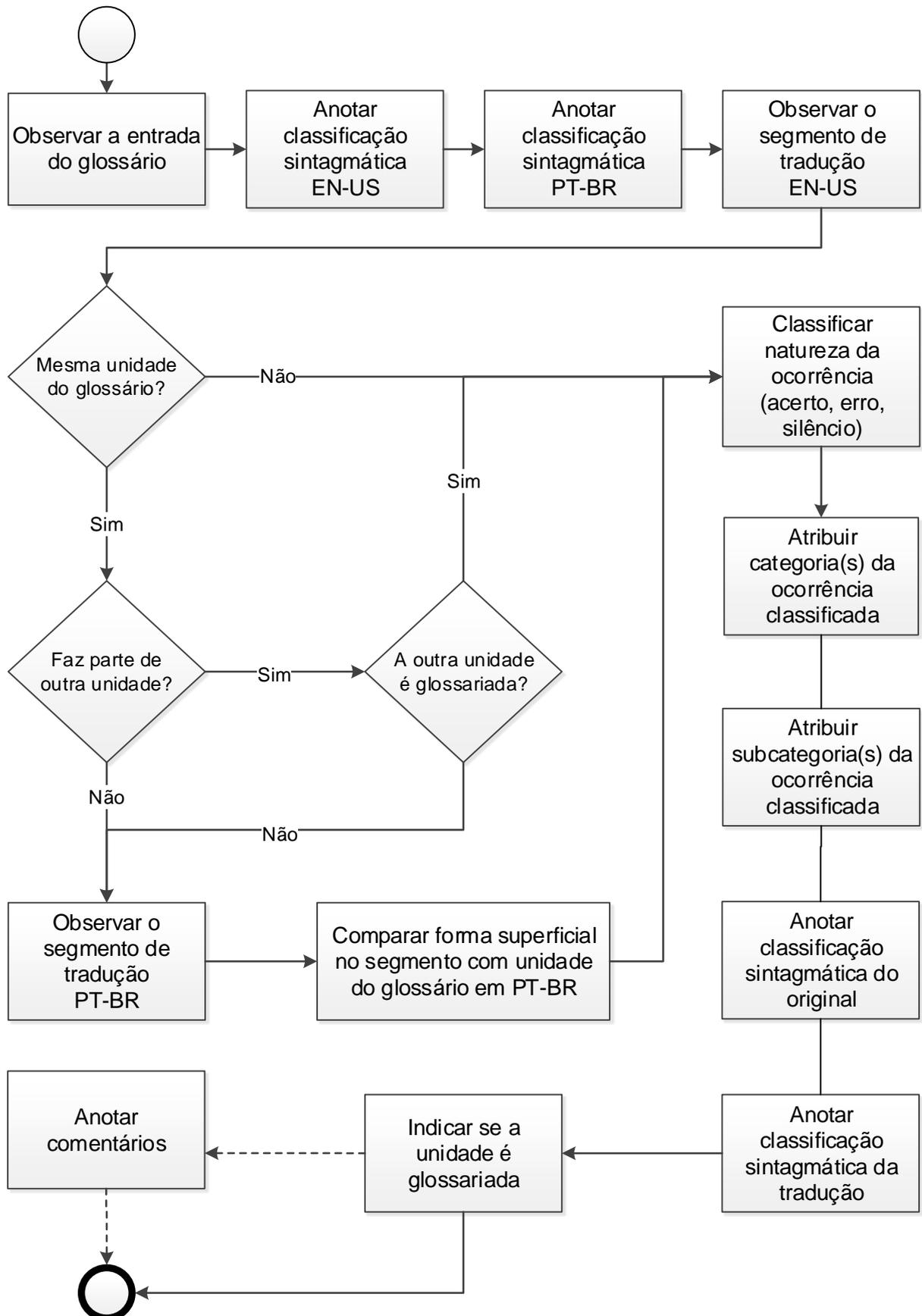
A lista completa de classificações verificadas pode ser encontrada em *Ocorrências linguísticas* na página 217. Outras duas categorias não previstas na análise foram encontradas, trata-se de ocorrências resultantes de erro de alinhamento do *subcorpus* de ajuda e de erros encontrados no glossário. Para mais informações, cf. *Alinhamento* (p. 158) e *Glossário* (p. 159).

4.2.4.2 Execução da análise

Em linhas gerais, o processo consiste em analisar individualmente cada item na lista de QA e: 1) comparar original e tradução do glossário com original e tradução do texto; 2) classificar de acordo com os parâmetros de análise.

A seguinte figura apresenta o processo diagramado em detalhes, seguido da descrição de cada passo. O diagrama serve para facilitar a compreensão dos processos envolvidos, especialmente em função da complexidade das decisões que devem ser tomadas na etapa 3.

Figura 22 - Diagrama processual da análise feita neste trabalho



Os passos no diagrama acima podem ser descritos da seguinte maneira:

- 1 observar a entrada do glossário;
- 2 anotar no campo *GR Glossário EN-US* a classe gramatical ou a classificação sintagmática da unidade em inglês;
- 3 anotar no campo *GR Glossário PT-BR* a classe gramatical ou a classificação sintagmática da unidade em português;
- 4 observar o segmento de tradução em EN-US;
- 5 validar que se trata de fato da mesma unidade presente no glossário em EN-US;
 - a em caso afirmativo:
 - i verificar se a unidade faz parte de outra, sintagmática;
 - 1 em caso afirmativo:
 - a verificar se essa unidade sintagmática consta no glossário;
 - i em caso afirmativo, ir para 6;
 - ii em caso negativo, ir para 5 .a .ii;
 - 2 em caso negativo, ir para 5 .a .ii;
 - ii observar o segmento de tradução em PT-BR;
 - iii verificar a forma linguística superficial no segmento e compará-la com a unidade presente no glossário em PT-BR;
 - b em caso negativo, ir para 6;
- 6 no campo *Tipo de ocorrência*⁷⁶, classificar sua avaliação (positivo ou falso positivo) segundo a comparação em 6;
- 7 no campo *Categoria da ocorrência*, atribuir a categoria ou as categorias que expliquem a ocorrência classificada em 6;
- 8 no campo *Subcategoria da ocorrência*, atribuir a subcategoria ou as subcategorias que expliquem a ocorrência classificada em 6 e 7;
- 9 no campo *GR Segmento (EN-US)*, anotar a classificação sintagmática da unidade que ocorre no original;
- 10 no campo *GR Segmento (PT-BR)*, anotar a classificação sintagmática da unidade que ocorre na tradução;
- 11 no campo *Glossariada?*, indicar se a unidade encontrada na tradução consta no glossário;

⁷⁶ Para obter mais informações sobre cada um dos campos envolvidos nas etapas 4-9 do processo, consulte *Preparação e organização dos dados* (p. 140).

12 (opcional) no campo *Comentários*, anotar quaisquer comentários que sejam pertinentes.

A fim de limitar o escopo da análise e viabilizar este trabalho em tempo e volume hábeis, foi estabelecido um recorte para a lista de validação. Este recorte consiste na análise das cerca de 2 500 primeiras ocorrências na lista.

Ao término deste processo, a tabela se apresentava da seguinte maneira:

Figura 23 - Tabela de registro dos dados após a análise

Segmento	Subcorpus	Entrada do glossário	Inglês (EN-US)	Português (PT-BR)	GR Glossário (EN-US)	GR Glossário (PT-BR)	GR Segmento (EN-US)	GR Segmento (PT-BR)	Resposta do software de QA	Tipo de ocorrência	Categoria	Sub-categoria	Gloss-ariada?	Comen-tários
Symantec - Ajuda.tmx (11)	Ajuda	infection / infecção	Norton A...	O Norton...	N	N	SN	SN	Inclui na lista de QA	Falso positivo	Linguística	Parte de outra unidade	S	
Symantec - Ajuda.tmx (2979)	Ajuda	infection / infecção	Computer...	As verific...	N	N	SN	SN	Inclui na lista de QA	Falso positivo	Linguística	Parte de outra unidade	S	
Symantec - Ajuda.tmx (4965)	Ajuda	infection / infecção	The Full S...	A Verifica...	N	N	N	N	Inclui na lista de QA	Falso positivo	Tradutória	Técnica -> Generalização		
Symantec - Ajuda.tmx (4965)	Ajuda	infection / infecção	The Full S...	A Verifica...	N	N	N	N	Inclui na lista de QA	Falso positivo	Linguística	Número		
Symantec - Ajuda.tmx (4632)	Ajuda	infection / infecção	You can u...	Você pode...	N	N	N	N	Inclui na lista de QA	Falso positivo	Tradutória	Técnica -> Generalização		
Symantec - Ajuda.tmx (4632)	Ajuda	infection / infecção	You can u...	Você pode...	N	N	N	N	Inclui na lista de QA	Falso positivo	Linguística	Número		
Symantec - Ajuda.tmx (11)	Ajuda	virus infection / infecção por vírus	Norton An...	O Norton...	SN	SN	SN	SN	Inclui na lista de QA	Falso positivo	Linguística	Número		
Symantec - Ajuda.tmx (2979)	Ajuda	virus infection / infecção por vírus	Computer...	As verifica...	SN	SN	SN	SN	Inclui na lista de QA	Falso positivo	Linguística	Número		

Fonte: Elaborado pelo autor.

4.2.5 Quantificação e filtragem dos resultados

Após a análise da lista de QA terminológico, foi feita uma etapa de pré-processamento para quantificar e filtrar os resultados. Para isso, foi empregado o recurso de tabelas dinâmicas presente no Microsoft Excel. Foram criadas tabelas dinâmicas para elucidar as seguintes situações:

- total de positivos e falsos positivos para os itens incluídos e excluídos na lista de QA pelo *software*;
 - influência dos *subcorpora* nesses totais;
- índice de falsos positivos para cada classificação;
- índices individuais para cada classificação, incluindo as ocorrências por sub-classificação.

Os resultados gerados por essas tabelas dinâmicas são apresentados no próximo capítulo.

5 ANÁLISE DOS DADOS E RESULTADOS

A análise dos dados apresenta a análise e interpretação de alguns números, como o número de falsos positivos por categoria, a relação entre positivos e falsos positivos, entre outros cruzamentos de dados que ajudam a nortear ou priorizar partes do desenvolvimento de um *software* de QA. É necessário, contudo, fazer a ressalva de que este trabalho não propõe nada de absoluto fora do âmbito do *corpus* estudado, por se tratar de uma análise preliminar e centrada em um só produto de apenas uma empresa. Ainda assim, os resultados podem trazer pistas ou indícios sobre algum tipo de padrão, por exemplo, no contraste entre *software* e ajuda, ou para alguma das classificações analisadas.

São apresentados, inicialmente, alguns números e análises preliminares em âmbito geral, como o número de falsos positivos por *subcorpus* e por categoria de classificação. Esta parte inicial compreende tanto a análise dos itens incluídos na lista de QA quanto daqueles que são excluídos pelo *software*. A seguir, são feitas descrições mais aprofundadas sobre cada uma das categorias registradas para as ocorrências da lista de QA. Esta parte centra-se exclusivamente nos itens incluídos pelo *software* na lista de QA.

5.1 Análise global dos resultados

Esta seção apresenta uma síntese das ocorrências encontradas, considerando as quatro categorias da análise mencionadas na revisão teórica (fraseológica, variacional, tradutória e linguística), bem como as duas categorias que foram identificadas durante a análise (alinhamento e glossário). Esta síntese toma a forma de um quadro comparativo de todas as ocorrências e é feita em linhas gerais. Além disso, traz um comparativo entre todas as categorias analisadas. Os detalhes específicos de cada categoria, contudo, são apresentados nas seções seguintes.

Os dados analisados nesta seção compreendem principalmente os itens que o *software* inclui na lista de QA, ou seja, aqueles que devem ser revisados pelo tradutor. Para introduzir outra dimensão de análise, também são apresentados as ocorrências que o *software* exclui da lista, ou seja, aquelas que estariam corretas. Contrastar estes itens pode sugerir novas direções

para estudos futuros. Contudo, como o foco deste trabalho é analisar a ocorrência de falsos positivos na lista, este contraste não será aprofundado.

5.1.1 Acertos e erros do *software*

As tabelas a seguir apresentam os números gerais de positivos e falsos positivos encontrados para as ocorrências que o *software* inclui e exclui da lista de QA:

Tabela 18 - Positivos e falsos positivos para ocorrências que o *software* inclui na lista de QA

Tipo de ocorrência	Inclui na lista	
	#	%
Falso positivo	2086	97,61%
Positivo	51	2,39%
Total Geral	2137	100,00%

Fonte: Elaborado pelo autor.

Tabela 19 - Positivos e falsos positivos para ocorrências que o *software* exclui da lista de QA

Tipo de ocorrência	Exclui da lista	
	#	%
Falso positivo	274	9,91%
Positivo	2491	90,09%
Total Geral	2765	100,00%

Fonte: Elaborado pelo autor.

Estas tabelas apresentam as seguintes informações:

- **Exclui da lista:** ocorrências que o *software* considera corretas no QA terminológico, ou seja, aquelas que ele não inclui na lista de validação; ocorre para a forma tanto do termo original quanto da tradução, registrada no glossário, aparecem nos segmentos de tradução;
- **Inclui na lista:** ocorrências que o *software* considera erradas no QA terminológico, ou seja, aquelas que ele inclui na lista de validação; ocorre quando a forma do termo original registrada no glossário aparece nos segmentos de tradução, mas a forma da respectiva tradução não é encontrada;

- **Tipo de ocorrência:** indica se o *software* acertou ou não em ter incluído ou excluído a ocorrência da lista de QA; resulta da análise dos dados do *corpus* em relação às categorias de classificação reunidas na revisão teórica e consistem em:
 - **Positivo:** ocorrências que o *software* acertou, seja por não incluir acertadamente na lista de QA, seja por ter incluído (cf. *Erros de tradução* na página 188 da análise tradutória para ver exemplos de acertos do *software*, os quais se traduzem em erros efetivos de tradução);
 - **Falso positivo:** ocorrências que o *software* errou, tanto para os itens que incluiu na lista de QA quanto para aqueles que foram incluídos.

Como as tabelas acima demonstram, há uma relativa proporção entre as ocorrências que o *software* inclui e exclui da lista (2137 e 2765 ocorrências, respectivamente). Quanto às ocorrências excluídas pelo *software*, uma pequena parte corresponde a falsos positivos. Trata-se de unidades que poderiam ser relacionadas umas com as outras, como *download / fazer o download* e *download / download* (mais sobre o assunto nas seções que tratam do MT nas páginas 169 e 224).

Na direção oposta, a grande maioria das ocorrências incluídas pelo *software* resulta em falsos positivos (97,61%). Estes casos envolvem as categorias apresentadas neste trabalho para estudar o volume de falsos positivos na lista de QA, segundo a metodologia adotada (cf. *Critérios de análise* na página 145). A considerar pelo grande volume, fica evidente o impacto negativo que as ocorrências têm na geração dessa lista. É possível afirmar com segurança que, para o *corpus* analisado, a ocorrência de falsos positivos é predominante na lista de QA.

5.1.1.1 Influência do *subcorpus*

Uma das questões que surgem, dada as particularidades dos dois tipos de texto coletados para análise (uma UI e a respectiva ajuda), é se suas diferenças teriam alguma influência em números nas incidências de falsos positivos ou na detecção do *software*. Como se pode observar pelas tabelas a seguir, a grande maioria das ocorrências se concentra na ajuda.

Tabela 20 - Ocorrências que o *software* inclui na lista de QA por *subcorpus*

<i>Subcorpus</i>	Inclui na lista	
	#	%
Ajuda	1965	91,95%
UI	172	8,05%
Total Geral	2137	100,00%

Fonte: Elaborado pelo autor

Tabela 21 - Ocorrências que o *software* exclui da lista de QA por *subcorpus*

<i>Subcorpus</i>	Exclui da lista	
	#	%
Ajuda	2467	89,22%
UI	298	10,78%
Total Geral	2765	100,00%

Fonte: Elaborado pelo autor

Como se vê, a relação do volume de ocorrências incluídas e excluídas entre a ajuda e a UI mostra que a última corresponde a mais ou menos 10% do total em todos os casos. Conforme apresentado na metodologia (cf. *Interface de usuário* na página 121 e *Síntese do corpus* na página 125), a UI compõe quase 27% do volume de texto total do *corpus*. Será que essa diferença estaria dentro de algum desvio esperado nessa análise estatística? Será resultante de alguma particularidade da UI? Como ideia para futuros trabalhos, talvez análises mais aprofundadas desses *subcorpora*, como, por exemplo, sobre suas densidades terminológicas, possam trazer mais respostas nesse sentido.

Entrando um pouco mais nos detalhes, as seguintes tabelas são uma réplica das tabelas apresentadas acima, agora seccionadas por *subcorpus*:

Tabela 22 - Ocorrências que o *software* inclui na lista de QA para o *subcorpus* de ajuda

Tipo de ocorrência	Inclui na lista	
	#	%
Falso positivo	1723	98,68%
Positivo	23	1,32%
Total Geral	1746	100,00%

Fonte: Elaborado pelo autor.

Tabela 23 - Ocorrências que o *software* exclui da lista de QA para o *subcorpus* de ajuda

Tipo de ocorrência	Exclui da lista	
	#	%
Falso positivo	243	9,85%
Positivo	2224	90,15%
Total Geral	2467	100,00%

Fonte: Elaborado pelo autor.

Tabela 24 - Ocorrências que o *software* inclui na lista de QA para o *subcorpus* de UI

Tipo de ocorrência	Inclui na lista	
	#	%
Falso positivo	149	98,68%
Positivo	2	1,32%
Total Geral	151	100,00%

Fonte: Elaborado pelo autor.

Tabela 25 - Ocorrências que o *software* exclui da lista de QA para o *subcorpus* de UI

Tipo de ocorrência	Exclui da lista	
	#	%
Falso positivo	31	10,40%
Positivo	267	89,60%
Total Geral	298	100,00%

Fonte: Elaborado pelo autor.

O que se vê, mais uma vez, são estatísticas semelhantes, e a UI está a uma proporção de 10% do volume da ajuda. Uma possibilidade é que as ocorrências recebam influência de classe gramatical. Esse tipo de análise, contudo, foge do escopo desta pesquisa e, conforme dito acima, sugere a necessidade de estudos mais aprofundados no futuro.

5.1.1.2 Ocorrências por classificação

Esta seção apresenta as ocorrências que o *software* marcou como erro, mas que, na verdade, consistem em falsos positivos e seriam simplesmente ignoradas pelo tradutor ao revisar a lista de QA. Um dos objetivos de tal análise é prover informações que ajudem na tomada de decisões para o desenvolvimento de melhorias no *software* de QA. Nesse sentido, as ocorrências mais frequentes, se solucionadas, trariam mais benefícios para o tradutor. A seguinte tabela apresenta o volume e a porcentagem de falsos positivos para cada tipo de classificação usada no trabalho:

Tabela 26 - Ocorrências de falsos positivos por classificação para os itens que o *software* inclui na lista de QA

Categoria	Falso positivo	
	#	%
Linguística	1494	71,62%
Variacional	214	10,26%
Glossário	193	9,25%
Tradutória	73	3,50%
Alinhamento	59	2,83%
Fraseológica	53	2,54%
Total Geral	2086	100,00%

Fonte: Elaborado pelo autor.

Esta tabela apresenta as seguintes informações:

- **Categoria:** indica a categoria de análise à qual o falso positivo está associado;
- **Falso positivo:** apresenta as ocorrências que o tradutor descarta ao revisar a lista de QA.

Como se pode observar, a classificação linguística é a que apresenta a maioria dos casos, seguida de variacional, glossário, tradutória, alinhamento e fraseológica. Como foi afirmado acima, a análise revelou ainda duas outras categorias de classificação, *Alinhamento* e *Glossário*. Para estas, observa-se que problemas no glossário representam boa parte das ocorrências (9,25%) enquanto que os problemas na segmentação do *corpus* representam 2,83% dos casos. A análise para cada uma das categorias é apresentada a seguir.

5.2 Alinhamento

As ocorrências relacionadas a alinhamento ocorreram porque o *software* Trados WinAlign não pôde alinhar todos os segmentos corretamente (cf. *Coleta de dados para o corpus* → *Ajuda* na página 134 do capítulo *Procedimentos metodológicos*). Essas ocorrências estão todas na ajuda, visto que este foi o único *subcorpus* que demandou esse tipo de alinhamento. Segue um exemplo:

Entrada do glossário	bootable CD / CD inicializável
Segmento original	The download contains an easy-to-use wizard that enables you to download the wizard to your computer or create a bootable CD , DVD, or USB device.
Segmento traduzido	É necessário que você possa iniciar seu computador com um dispositivo USB para usar a versão USB do Norton Bootable Recovery Tool. Antes de iniciar, feche todos os programas abertos em seu computador e certifique-se de que você esteja conectado à Internet.

Ao olhar para os segmentos acima, fica visível que as traduções nada têm a ver com os originais. As traduções correspondentes estão em outras partes, como se vê nos pares a seguir (o fundo escuro destaca a divisão dos segmentos alinhados):

If the installation of your product fails, you can download the Norton Bootable Recovery Tool.

Se a instalação de seu produto falhar, você poderá fazer o download da Norton Bootable Recovery Tool. A Norton Bootable Recovery Tool ajuda a verificar se seu computador apresenta os riscos que podem impedir uma instalação bem-sucedida.

The Norton Bootable Recovery Tool helps you to check your computer for the risks that can prevent successful installation.

O download contém um assistente fácil de usar que permite que você faça o download do assistente em seu computador ou crie um CD inicializável, DVD ou dispositivo USB.

The download contains an easy-to-use wizard that enables you to download the wizard to your computer or create a bootable CD, DVD, or USB device.

É necessário que você possa iniciar seu computador com um dispositivo USB para usar a versão USB do Norton Bootable Recovery Tool. Antes de iniciar, feche todos os programas abertos em seu computador e certifique-se de que você esteja conectado à Internet.

Esse tipo de caso já era esperado, visto que o WinAlign, embora seja muito eficiente no alinhamento, não consegue processar perfeitamente 100% dos casos. Este problema diz respeito à coleta do *corpus*; muito provavelmente, os arquivos bilíngues localizados não apresentavam esse tipo de problema. Por esse motivo, estes casos foram descartados da análise.

5.3 Glossário

As ocorrências referentes ao glossário compreendem, em sua maior parte, na representação inadequada das unidades no campo reservado para a tradução no glossário. O problema compreende as seguintes unidades:

- browse / procurar (*)
- community usage / uso pela comunidade*
- desktop / área de trabalho (*)
- double-click / clicar duas vezes (*)
- download / download (*)
- incoming / de entrada (*)
- intrusion prevention / prevenção contra intrusões (*)
- keystroke logger / keylogger (*)
- manager / gerenciador (*)
- network storage device / dispositivo de armazenamento em rede (*)

A lista acima é a representação fiel das unidades. Ou seja, os asteriscos entre parênteses (ou o asterisco individual, no caso de *community usage*) de fato constam na tradução. Como o *software* de QA busca a tradução presente no glossário na verificação terminológica, os caracteres incorretos fazem com que todos os casos para os termos listados acima sejam colocados na lista de validação – a menos que a tradução do texto apresente os asteriscos, o que é bastante improvável, especialmente nos casos de (*). A solução seria simplesmente remover os caracteres desnecessários dessas entradas.

Há ainda um caso especial:

- personally identifiable information / informações que identificam pessoalmente

Neste caso, há apenas uma ocorrência na tradução que corresponde à forma apresentada no glossário. Não por acaso, a unidade glossariada está mais para um sintagma livre que para uma unidade terminológica. As ocorrências traduzidas se dividem em:

- informação de identificação pessoal
- informação que o identifique pessoalmente

Essas particularidades levam a crer que a unidade tenha sido usada primeiro para algum fim específico e, por ter resolvido talvez algum problema de tradução, o tradutor ou terminólogo optou por registrá-la em glossário, sem antes validar a real necessidade de fazê-lo, nem de avaliar as prováveis situações de uso.

5.4 Análise fraseológica

A análise fraseológica tratada nesta seção compreende as características das UFE que foram vistas na revisão teórica – como o grau de fixação e a presença de UT (cf. páginas 51 e 52, respectivamente). Conforme a análise apresentada aqui demonstra, essa categorização não é significativa para a geração de falsos positivos, embora as UFE em geral apresentem outras características que têm influência no ruído, como flexão verbal e concordância de número (mais detalhes em *Outras causas de ruído relacionadas à fraseologia* na página 164).

Com relação às características das UFE, a análise revela que, das seis características apresentadas neste trabalho, três são determinantes para a ocorrência de falsos positivos no QA: a estabilidade sintática, a estabilidade semântica e o MT. Destas, as duas primeiras se sobrepõem a algumas das classificações usadas para analisar variação e o MT se sobrepõe a uma classificação empregada na análise da tradução.

Embora a categorização fraseológica usada neste trabalho não represente um número tão significativo no ruído, as UFE em geral contribuem para essas ocorrências devido a outras classificações, como as morfológicas (mais detalhes sobre essa e outras classificações em *Análise linguística* na página 216). Além disso, talvez esse fato decorra da baixa frequência com que tais unidades são registradas em glossário ou ainda de seu carácter sintagmático, o qual dificulta sua detecção pelo *software*.

No *corpus* foram identificadas 22 UFE, segundo os critérios de caracterização apresentados na revisão teórica (p. 48). Destas, 15 não resultam em ruído no QA. Um motivo para

isso pode ser que essas 15 não tenham sido detectadas pelo *software* (por exemplo, em função da inserção de elementos ou de flexão verbal);

Não há um pareamento entre as entradas do glossário com relação à UFE. Algumas são unidades simples no original e UFE na tradução. Além disso, parece não haver um critério explícito para a presença de UFE no glossário. Para as unidades verbais, há indício de que o uso de empréstimos contribui para o registro de UFE no glossário. Para as nominais, em geral são registradas apenas como sintagmas nominais (SN).

Com relação a essa baixa frequência de registro no glossário, a considerar os sintagmas verbais (SV) registrados para o inglês (todos em formato de *phrasal verb*, como *back up*), parece que as fraseologias inseridas no glossário em português são resultados de ação tradutória *ad hoc*, e não sistemática, por parte de quem mantém esse tipo de base terminológica. Em inglês, não há nenhuma ocorrência de UFE registrada como SV, apenas como SN. Algumas das UFE registradas como SV em português decorrem do uso de empréstimos. É o caso dos pares *back up / fazer backup* e *download / fazer o download*.

No que diz respeito à prescritividade atribuída ao setor da localização em geral, esta análise demonstra que ela de fato tem influência nas traduções, que procuram seguir à risca o glossário, mas está longe de ser sistemática a ponto de adicionar complexidade ao texto. Com as UFE, pôde-se observar que a variação lexical é usada para resolver problemas de tradução (cf. *Estabilidade semântica* na página 168 para detalhes).

As seções a seguir apresentam algumas observações a respeito dessa análise fraseológica, incluindo a interpretação de alguns números e um resumo das características das UFE que influem no ruído.

5.4.1 Fraseologia em números

Nesta seção, serão apresentados alguns cruzamentos dos dados obtidos na análise, como a relação entre as ocorrências de cada um dos *subcorpora* e as ocorrências em função do tipo de sintagma.

A seguinte tabela apresenta uma síntese dos resultados encontrados para UFE, segmentado para cada *subcorpora*. Os resultados apresentados consideram apenas a influência sobre o ruído que provém das características das UFE, como a estabilidade sintática e semântica.

Tabela 27 - Resultados gerais para UFE

<i>Subcorpus</i>	Falso positivo	
	#	%
Ajuda	50	94,34%
UI	3	5,66%
Total Geral	53	100,00%

Fonte: Elaborado pelo autor

Esta tabela apresenta as seguintes informações:

- **Subcorpus:** indica se a ocorrência foi encontrada na ajuda ou na UI;
- **Falso positivo:** apresenta as ocorrências que o tradutor descarta ao revisar a lista de QA.

A seguinte tabela mostra os resultados separados por tipo de sintagma existente na tradução:

Tabela 28 – Relação entre falsos positivos e classificação sintagmática por *subcorpus* para UFE

<i>Subcorpus</i>	Falso positivo						Total	
	N		SN		SV		#	%
	#	%	#	%	#	%		
Ajuda	2	3,77%	21	39,62%	27	50,94%	50	94,34%
UI		0,00%	1	1,89%	2	3,77%	3	5,66%
Total Geral	2	3,77%	22	41,51%	29	54,72%	53	100,00%

Fonte: Elaborado pelo autor.

Esta tabela apresenta as seguintes informações:

- **Subcorpus:** indica se a ocorrência foi encontrada na ajuda ou na UI;
- **Falso positivo:** indica as ocorrências que o *software* errou ao incluir na lista de QA.

Estão segmentados por classe gramatical:

- **N:** nomes;
- **SN:** sintagmas nominais;
- **SV:** sintagmas verbais.

Com base nos números acima, é possível tirar algumas conclusões preliminares, as quais poderiam ser confirmadas ou não por uma análise mais aprofundada e abrangente que contemplasse outros *corpora* de outras empresas.

Primeiro, *nenhuma das ocorrências fraseológicas consiste em erro efetivo*. Por um lado, isso mostra a qualidade do material traduzido, que deve ter passado por várias etapas de análise, apresentando um número pequeno de erros. Por outro, demonstra que as características das UFE apresentam um potencial para introduzir ruído no QA.

Em segundo lugar, *a proporção de ruído na ajuda é superior à da UI em relação ao volume de cada subcorpora*. O volume de palavras do *subcorpus* de ajuda é quase o triplo da UI. Contudo, o volume de falsos positivos na ajuda é quase sete vezes maior. Uma possível causa para isso pode estar nas características de uma UI, que normalmente traz verbos no infinitivo em opções de menu. Seguem alguns exemplos de ocorrências encontradas na UI para UFE *colocar em quarentena*, indicando o contexto das ocorrências da interface:

Tabela 29 - Ocorrências para a UFE colocar em quarentena na UI

Ocorrência	Contexto
<i>Colocar</i> arquivo em quarentena	Opção de menu ou caixa de diálogo
<i>Colocar em quarentena</i> manualmente	Opção de menu ou caixa de diálogo
Falha ao <i>colocar em quarentena</i>	Mensagem de erro
Não é possível <i>colocar em quarentena</i> arquivos em compartilhamento de rede.	Mensagem de erro
Se existirem arquivos com nível de confiança Baixo ou Ruim, clique no ícone de cruz vermelha (x) para <i>colocá-lo em quarentena</i> .	Ajuda contextual
Essa categoria inclui os riscos à segurança ou os arquivos infectados que o Norton AntiVirus não conseguiu reparar, <i>colocar em quarentena</i> ou remover.	Ajuda contextual

Fonte: Elaborado pelo autor.

Pode ser que a baixa variação do sujeito nessas frases ou o tipo de formulação que é usado propicie um índice maior de positivos do *software* na UI. Das ocorrências acima, é possível retirar formulações como:

- [SV_{infinitivo}] para opções de menu;
- *não é possível* [SV_{infinitivo}] para mensagens de erro.

Nesta segunda formulação, observa-se que o SV em questão é replicado nas mensagens de erro para determinada ação que o usuário quer fazer com o *software*. No caso de um

antivírus, o usuário pode ser um comando *Verificar*, para o qual há uma mensagem de erro *Falha ao verificar* ou *Falha na verificação*. Todo esse tipo de formulação, que faz parte de um *software* e é projetado em seu desenvolvimento, tende a contribuir para a presença de formas mais fixas. Embora essa seja a tendência, os exemplos acima mostram que esse nem sempre é o caso.

5.4.1.1 Outras causas de ruído relacionadas à fraseologia

Como foi observado antes, as ocorrências de falsos positivos analisadas até agora consideram apenas as características analisadas para fraseologia, como a estabilidade sintática e semântica. Ao incluir os demais aspectos que podem afetar a detecção de UFE (flexão verbal, por exemplo; mais detalhes em *Análise linguística* na página 216), a situação passa a pender muito mais para o ruído:

Tabela 30 - Total de incidência de falsos positivos para UFE ao considerar fatores relacionáveis

<i>Subcorpus</i>	Falso positivo	
	#	%
Ajuda	504	92,64%
UI	40	7,36
Total Geral	544	100%

Fonte: Elaborado pelo autor.

Observa-se que o volume de ocorrências sobre de 53 para 544, representando um aumento de 1000%. Embora as ocorrências de falsos positivos em função de fraseologia representem uma porcentagem baixa do *corpus*, vale dizer que as UFE também estão relacionadas a outros elementos causadores de ruído, como, por exemplo, a flexão verbal em:

Entrada do glossário	download / fazer o download
Segmento original	When the Download Insight Notifications option is set to Risks Only, Download Insight notifies only when you <i>download</i> an infected or a suspicious executable file.
Segmento traduzido	Quando a opção Notificações do Download Insight estiver configurada como Apenas riscos, você será notificado pelo Download Insight somente quando <i>fizer download</i> de um arquivo executável infectado ou suspeito.

Aqui, além da flexão verbal (*fizer*), a ausência do artigo definido *o* também impede o *software* de detectar a unidade na tradução. Outro exemplo para outro aspecto:

Entrada do glossário	memory usage / uso da memória
Segmento original	Norton AntiVirus monitors the overall system CPU usage and <i>memory usage</i> and the Norton-specific CPU usage and memory usage.
Segmento traduzido	O Norton AntiVirus monitora o <i>uso</i> geral da CPU e <i>da memória</i> do sistema, assim como o uso da CPU e da memória do Norton.

Aqui, ainda se observa uma elipse lexical anafórica, já que o referencial de *uso* é recuperado do sintagma nominal inteiro (uso da CPU; uso da memória). Observa-se ainda um caso de inserção neste exemplo (uso geral...).

5.4.2 Sumário das ocorrências em relação às características das UFE

Três características das UFE foram identificadas como determinantes para a ocorrência de falsos positivos, a saber, estabilidade sintática, estabilidade semântica e MT. Também é possível apontar o carácter sintagmático como causador indireto, já que historicamente pesquisas em PLN e linguística computacional têm apontado para as dificuldades de identificar e processar unidades compostas. Já as demais características – inclusão de pelo menos uma UT, não-composicionalidade e uso em âmbito específico – serviram para determinar que unidades são UFE no momento da análise, mas não servem para explicar as ocorrências de falsos positivos observadas. A tabela a seguir traz um resumo dos falsos positivos que são atribuídos às

UFE no *corpus*, seguida de comentários a respeito dessas três características causadoras de ruído.

Tabela 31 - Identificação de característica causadora de ruído e representação formal para as UFE do *corpus*

Termo	Característica classificatória	Segmento original	Segmento traduzido	Representação formal
download / fazer o download	estabilidade sintática – inserção	Norton AntiVirus automatically downloads definition updates regularly, scans your computer, and secures your computer from all types of viruses and unknown threats.	O Norton AntiVirus faz automaticamente o download de atualizações de definições regularmente, verifica o computador e protege-o contra todos os tipos de vírus e ameaças desconhecidas.	fazer [adv] o download
download / fazer o download	estabilidade sintática – inserção	Check and download the latest version of the product by using Norton Update Center.	Selecionar e fazer download da versão mais recente do produto usando o Norton Update Center.	fazer download
quarantine / colocar em quarentena	estabilidade sintática – inserção	Quarantine File	Colocar arquivo em quarentena	colocar [N] em quarentena
quarantine / colocar em quarentena	estabilidade sintática – inserção	Quarantine this file	Colocar esse arquivo em quarentena	colocar [Art+N] em quarentena
quarantine / colocar em quarentena	estabilidade sintática – inserção	You can click on the red cross (x) icon and quarantine the suspicious file.	Você pode clicar no ícone de cruz vermelha (x) e colocar o arquivo suspeito em quarentena .	colocar [Art+N+A] em quarentena
quarantine / colocar em quarentena	estabilidade sintática – inserção	If the Ask me what to do option is disabled Norton AntiVirus automatically quarantines the detected worm and notifies you.	Se a opção Perguntar o que fazer estiver desativada, o Norton AntiVirus vai colocar o worm detectado em quarentena automaticamente e notificar você.	colocar [Art+N+A] em quarentena
memory usage / uso da memória	estabilidade sintática – inserção	Displays a graph that represents overall memory usage and Norton-specific memory usage.	Exibe um gráfico que representa o uso geral da memória e o uso da memória pelo Norton.	uso [A] da memória
password protection / proteção por senha	estabilidade sintática – inserção	Settings Password Protection	Proteção das configurações por senha	proteção [Prep+Art+N] por senha
back up / fazer backup	estabilidade sintática – inserção	By using these Norton products, you can back up important files and folders and restore them.	Através do uso dos produtos Norton, é possível fazer o backup de arquivos e pastas importantes e restaurá-los.	fazer [Art] backup
community	estabilidade	This window displays information	Essa janela exibe informações como	uso [Prep+Art+N]

Termo	Característica classificatória	Segmento original	Segmento traduzido	Representação formal
usage / uso pela comunidade	sintática – inserção	such as origin and lineage details, activity, confidence level, and <i>community usage</i> of the file.	os detalhes de origem e de linhagem, a atividade, o nível de confiança e o <i>uso</i> do arquivo <i>pela comunidade</i> .	pela comunidade
download / fazer o download	estabilidade semântica e MT	Checks all files that you <i>download</i> from the Internet, including Java Applets and ActiveX controls	Verifica todos os arquivos <i>obtidos por download</i> da Internet, inclusive applets Java e controles ActiveX	obter por download
quarantine / colocar em quarentena	MT	If you do not send an email message at that time, then it is probably a worm and you should <i>quarantine</i> the file.	Se você não estava enviando mensagens naquele momento, é provável que o arquivo contenha um worm e deva ser <i>colocado em quarentena</i> .	colocado em quarentena
reputation analysis / análise de reputação	MT	If you do not want Download Insight to process any file that you download for <i>reputation analysis</i> , you can turn it off.	Caso não queira que o Download Insight <i>analise a reputação</i> dos arquivos obtidos por download, você poderá desativá-lo.	

Fonte: Elaborado pelo autor.

5.4.2.1 Estabilidade sintática – inserção

Nas ocorrências de falsos positivos decorrentes de estabilidade sintática, os seguintes padrões foram encontrados⁷⁷ para o português:

- NE[A] NT[Prep+N]
- NE[N] ins[A] NT[Prep+N]
- NE[N] ins[Prep+Art+N] NT[Prep+N]
- NE[V] ins[N] NT[Prep+N]
- NE[V] ins[Adv] NT[Prep+N]
- NE[V] ins[Adv] NT[N]
- NE[V] ins[Art+N] NT[Prep+N]
- NE[V] ins[Art+N+A] NT[Prep+N]
- NE[V] NT[N]
- NE[V] NT[Prep+N]

⁷⁷ NE indica o núcleo eventivo, NT é o núcleo terminológico e ins corresponde aos elementos inseridos na unidade.

Como se observa, os elementos de inserção são em geral adverbiais, nominais, sintagmas preposicionais (SP) ou SN, dependendo da estrutura superficial que está presente. Por exemplo, para uma UFE em forma de SV, a inserção consiste em um advérbio ou SN, como em *faz automaticamente o download* ou *colocar o worm detectado em quarentena*. Também pode estar na presença ou ausência de artigo no NT, como em *fazer o backup*. Este último caso é sobreposto, figurando como uma questão de variação também. Já para uma UFE em forma de SN, a inserção é um adjetivo ou SP que se liga à unidade, como se vê em *uso geral da memória* ou *proteção das configurações por senha*.

5.4.2.2 Estabilidade semântica

Há um caso específico de estabilidade semântica, que é da tradução de *download* por *obter por download*. Existe um termo em português, de uso bastante comum, que é *baixar* para traduzir *download* como verbo. É possível encontrar ocorrências para *baixar* em *sites* de suporte do Google, da VMware e da Microsoft – esta última registra o termo em sua base terminológica. O próprio *site* da Symantec apresenta o termo em algumas seções publicadas pela empresa, como o *site* de suporte.

Ainda assim, não foi encontrada nenhuma ocorrência para o termo no *corpus* selecionado. O termo *baixar* seria uma solução prática para ocorrências do termo como adjetivo ou particípio no inglês. Limitar-se à forma *fazer o download* gera alguma dificuldade tradutória para esses casos, como nestes exemplos:

Checks all files that you **download** from the Internet, including Java Applets and ActiveX controls.

Verifica todos os arquivos **obtidos por download** da Internet, inclusive applets Java e controles ActiveX.

In this case, the Download Insight window appears whenever you try to launch any **downloaded** file that has an unknown reputation score.

Nesse caso, a janela Download Insight será exibida sempre que você tentar iniciar qualquer arquivo **obtido por download** que tenha uma classificação de reputação desconhecida.

Para estes casos, a tradução mais correta em um sentido prescritivo seria:

Verifica todos os arquivos *dos quais é feito download* da Internet, inclusive applets Java e controles ActiveX.

Nesse caso, a janela Download Insight será exibida sempre que você tentar iniciar qualquer arquivo *do qual tenha sido feito download* que tenha uma classificação de reputação desconhecida.

Nota-se que tais formas trazem uma complexidade desnecessária às traduções. O mesmo também poderia ser dito de *obter por download*, mas talvez essa tenha sido a melhor alternativa ao alcance dos tradutores em função de alguma preferência ou prescritividade no uso de *download* – e quem sabe até mesmo de alguma proibição no uso de *baixar*.

Além disso, observa-se no último caso que a ocorrência representa silêncio, já que *downloaded* não é identificado pelo *software* de QA. As mesmas observações procedem para a forma verbal do termo *backup*, a qual também não é identificada, como se vê a seguir:

You can access or restore the *backed up* data from any computer that is connected to the Internet.

É possível acessar ou restaurar os dados *salvos em backup* de qualquer computador com conexão à Internet.

Estes últimos exemplos demonstram algumas das possibilidades de aplicação da metodologia usada neste trabalho para identificar silêncio. Neste caso, além de aspectos morfológicos, é possível aplicar os critérios de MT e estabilidade semântica.

Assim como na estabilidade sintática, a estabilidade semântica também se apresenta como uma questão de variação, visto que o caso de *fazer o download* e *obter por download* consiste em um caso de variação lexical.

5.4.2.3 Modelo transformacional

Na condição de característica causadora de ruído, o MT está diretamente relacionado com uma técnica de tradução, a *transposição* (mais informações sobre a técnica em *Técnicas de tradução* na página 109). O caso de *reputation analysis* / *análise de reputação* demonstra

bem a aplicabilidade do MT para identificar ruído no QA e também o uso dessa técnica na tradução:

If you do not want Download Insight to process any file that you download for *reputation analysis*, you can turn it off.

Caso não queira que o Download Insight *analise a reputação* dos arquivos obtidos por download, você poderá desativá-lo.

O *exemplo*, além de revelar as estruturas superficiais sobre um eixo temático, em que *análise de reputação* e *analisar a reputação* representam uma estrutura subjacente, também demonstra a transposição, já que o sintagma nominal *reputation analysis* do original foi traduzido pelo sintagma verbal *analisar a reputação*.

Além disso, o critério de MT permite agrupar sob a mesma análise várias estruturas superficiais que são registradas em glossário para a mesma estrutura subjacente. É o caso das seguintes entradas, todas presentes no glossário:

- quarantine / quarentena
- quarantine / colocar em quarentena
- quarantined / em quarentena

Destas as duas primeiras são responsáveis pela geração de ruído desnecessário na verificação do QA. Isso acontece porque o *software* verifica todos os segmentos em que *quarantine* ocorre, causando redundância na verificação. É o que se vê nos casos a seguir:

Restores the detected security risk from the *quarantine* to your computer.

Restaura o risco à segurança detectado da *quarentena* para o computador.

Failed to restore item from *Quarantine*.

Falha ao restaurar item da *quarentena*.

Para esses casos, quando o *software* processa a unidade nominal acima, ele identifica que a tradução correspondente foi usada e segue adiante. Contudo, por considerar a estrutura verbal como uma unidade diferente, ele processa os mesmos segmentos novamente, identificando que a tradução correspondente, *colocar em quarentena*, não foi usada nesses casos,

gerando ruído. Como se verá mais adiante na seção *Análise linguística*, em *Modelo transformacional* (p. 224), isso vale também para as unidades individuais.

5.4.3 Outros resultados da análise

5.4.3.1 Unidades potencialmente fraseológicas

Foram observadas algumas unidades que têm as características de UFE, mas que não apresentaram nenhum falso positivo, inclusive levando-se em consideração características como estabilidade sintática e semântica na análise das ocorrências. Todas essas unidades estão registradas como SN no glossário, e os registros para SV e sintagmas adjetivos (SA) na tabela a seguir indicam as possíveis ocorrências para as estruturas superficiais relacionadas.

Tabela 32 - Unidades potencialmente fraseológicas

Original			Tradução		
SN	SV	SA	SN	SV	SA
<i>antivirus protection</i>	-	antivirus-protected	<i>proteção antivírus</i>	-	protegido por antivírus
<i>antivirus scan</i>	-	antivirus-scanned	<i>verificação antivírus</i>	-	verificado em busca de vírus
<i>disk usage</i>	use from disk	used from disk	<i>uso do disco</i>	usar do disco	usado do disco
<i>internet download</i>	download from internet	downloaded from internet	<i>download da Internet</i>	fazer o download da Internet	download feito da Internet
<i>intrusion prevention</i>	prevent against intrusion	prevented against intrusion	<i>prevenção contra intrusões</i>	prevenir contra intrusões	prevenido contra intrusões
<i>limited-time subscription</i>	subscribe for a limited time	subscribed for a limited time	<i>assinatura por tempo limitado</i>	assinar por tempo limitado	assinado por tempo limitado
<i>online backup</i>	back up online	backed up online	<i>backup online</i>	fazer backup online	salvo em backup online
<i>online storage</i>	store online	stored online	<i>armazenamento online</i>	armazenar online	armazenado online
<i>product upgrade</i>	upgrade product	upgraded product	<i>upgrade de produto</i>	fazer upgrade do	produto com upgrade

Original			Tradução		
				produto	
<i>risk alert</i>	alert to risk	risk-alerted	<i>alerta de riscos</i>	alertar sobre riscos	alertado para riscos
<i>security threat</i>	threat security	security-threatened	<i>ameaça à segurança</i>	ameaçar a segurança	com a segurança ameaçada
<i>SONAR Detection</i>	-	SONAR-detected	<i>deteção do SONAR</i>	-	detectado pelo SONAR
<i>system scan</i>	scan system	system-scanned	<i>verificação do sistema</i>	verificar o sistema	verificado no sistema
<i>virus infection</i>	-	virus-infected	<i>infecção por vírus</i>	-	infectado por vírus
<i>Web protection</i>	protect in the Web	Web-protected	<i>proteção da Web</i>	proteger na Web	protegido na Web

Fonte: Elaborado pelo autor.

Para os casos em que não há um SV marcado na tabela acima, a questão é que o SN original é composto por um nome deverbal e por um nome que é agente nas ocorrências em que a estrutura superficial é verbal. Por exemplo, para *antivirus protection / proteção antivírus*, as ocorrências de *protect / proteger* consistem no seguinte:

You should be always prepared in case a *virus infects* your computer.

Você sempre deve estar preparado caso um *vírus infecte* seu computador.

Embora esses elementos façam parte de um eixo temático, nesse caso, não se trata de uma unidade à parte, senão de elementos que compõem um sintagma livre. No sentido apresentado por Bevilacqua (2004), as UFE seriam compostas por uma núcleo terminológico e outro eventivo, caso da UFE [*degradação do meio ambiente, degradar o meio ambiente, meio ambiente degradado*], em que o núcleo terminológico é *meio ambiente* e o eventivo é [*degradação, degradar, degradado*].

Em *virus infection*, tem-se a impressão de uma situação parecida. Contudo, em *degradar o meio ambiente*, os elementos que compõem o SV são um núcleo verbal e um SN complementar. No outro caso, no lugar de um elemento complementar, *vírus* é o sujeito, o agente que infecta, a causa da infecção. Como lidar com esse tipo de caso é uma tarefa a ser realizada em trabalhos futuros.

Ao contrário, para *online backup / backup online*, o agente seria um usuário do *software*, como se observa no seguinte exemplo:

The vxsnap prepare and make tasks allow you to *back up* volumes *online* with minimal disruption to users.

Assim, esta última unidade poderia ser considerada UFE. Contudo, como esse tipo de ocorrência não existe no *corpus* selecionado para as unidades listadas acima, elas foram descartadas na classificação. O que não se descarta, contudo, é que estudos mais abrangentes possam afirmar o contrário.

5.4.3.2 Um problema contrastivo assíncrono entre idiomas

O aspecto que fica mais visível na análise das ocorrências fraseológicas é que não há uma relação de 1:1 entre a entrada original e sua tradução. A lista a seguir apresenta as sete unidades fraseológicas que foram identificadas e catalogadas:

- back up / fazer backup
- community usage / uso pela comunidade
- download / fazer o download
- memory usage / uso da memória
- password protection / proteção por senha
- quarantine / colocar em quarentena
- reputation analysis / análise de reputação

Três destas unidades não são UFE nos dois idiomas, mas apenas em português:

- back up / fazer backup
- download / fazer o download
- quarantine / colocar em quarentena

Em função disso, constatou-se que era necessário fazer uma classificação mais aprofundada das ocorrências inicialmente classificadas como linguísticas. Isso se faz necessário porque o MT que serviu de base para a análise fraseológica também se aplica em unidades menores.

5.4.3.3 Sobreposição de classificações

Há uma sobreposição entre a classificação variacional e a fraseológica. Exemplo:

Entrada do glossário	download / fazer o download
Segmento original	These Web sites detect and exploit the vulnerability of your Web browser to download malware programs to your system without your consent or knowledge.
Segmento traduzido	Estes sites detectam e exploram a vulnerabilidade de seu navegador da Web para fazer download de programas de malware em seu sistema sem seu consentimento ou conhecimento.

Em Fraseologia, observa-se a característica da inserção, já que a unidade pode incluir o artigo definido *o*. Em variação, trata-se de uma alteração morfossintática simples, em que a estrutura da unidade permanece inalterada, mas há a omissão do artigo.

Talvez essa sobreposição não seja boa para uma aplicação computacional, visto que poderia gerar esforços duplicados. Na verdade, o caso não parece bem assim. Há uma diferença na perspectiva para cada caso. Ao passo que a metodologia adotada para variação já busca uma tipologia mais concreta, a metodologia da Fraseologia provê uma aplicação mais flexível. Deve-se observar que a característica da inserção está sobreposta em alguns casos especí-

Entrada do glossário	quarantine / colocar em quarentena
Segmento original	If there are files with Poor or Bad trust level, click the red cross (x) icon to quarantine them.
Segmento traduzido	Se existirem arquivos com nível de confiança Baixo ou Ruim, clique no ícone de cruz vermelha (x) para colocá-lo em quarentena .

ficos na tipologia da variação (como este aqui), mas não em todos. Por exemplo:

Nesse caso, há a inserção de um pronome oblíquo entre a UFE *colocar em quarentena*. Contudo, esse caso especificamente não é coberto em variação.

Outro indício de que a sobreposição não é necessariamente ruim está nos casos em que a variação revela uma UFE, como no seguinte exemplo:

Entrada do glossário	download / fazer o download
Segmento original	Download Insight protects your computer against any unsafe file that you may run or execute after you <i>download</i> it using a supported Web browser.
Segmento traduzido	O Download Insight protege seu computador contra quaisquer arquivos não seguros que você venha a executar depois de <i>obter-los por download</i> com um navegador suportado da Web.

No segmento de tradução acima, há um caso de variação, já que a unidade verbal *download / fazer o download* foi traduzida por *obter por download*. A mesma unidade consiste em UFE e, além disso, revela a inserção de elementos no pronome oblíquo *los*.

Outro caso que pôde ser observado está na sobreposição entre a classificação tradutória e a fraseológica. Isso acontece em função do MT, o qual permite fazer o agrupamento de unidades superficiais sob a mesma estrutura de base. Esse agrupamento, quando representa uma mudança de classe gramatical entre o original e a tradução, também se configura como uma ocorrência tradutória. Por exemplo:

Entrada do glossário	reputation analysis / análise de reputação
Segmento original	If you do not want Download Insight to process any file that you download for <i>reputation analysis</i> , you can turn it off.
Segmento traduzido	Caso não queira que o Download Insight <i>analise a reputação</i> dos arquivos obtidos por download, você poderá desativá-lo.

Nesse caso, o nome *reputation analysis* no original passa a ser o sintagma verbal *analise a reputação* na tradução. Na perspectiva fraseológica adotada neste trabalho, *analisar a reputação* corresponde a *análise de reputação* por causa do MT. Já na perspectiva tradutória, o tradutor teria aplicado a técnica de transposição neste segmento.

5.4.4 Futuros caminhos para essa classificação

A metodologia aplicada neste trabalho para identificar e analisar UFE no *corpus* pode ser aplicada à detecção de silêncio no original. A estabilidade sintática e semântica, bem como o MT, definitivamente causa ruído na tradução. Deixando as diferenças entre os idiomas à

parte, não há motivos para pensar que essas características não venham a gerar silêncio no original também. A título de exemplo, os seguintes segmentos foram extraídos do *corpus*, cada um com a aplicação de um desses critérios:

MT:

Entrada do glossário	password protection / proteção por senha
Segmento original	Lets you <i>protect</i> Norton AntiVirus settings <i>with a password</i> .
Segmento traduzido	Permite <i>proteger</i> as configurações do Norton AntiVirus <i>com uma senha</i> .
Entrada do glossário	system scan / verificação do sistema
Segmento original	Programs that have the ability to <i>scan systems</i> or monitor activity and relay information to other computers or locations in cyberspace.
Segmento traduzido	Programas capazes de <i>verificar sistemas</i> ou monitorar informações sobre atividade e retransmissão para outros computadores ou locais no espaço cibernético.
Entrada do glossário	virus infection / infecção por vírus
Segmento original	Lets you run a Quick Scan to protect possible virus-infected areas of your computer.
Segmento traduzido	Permite executar uma Verificação rápida e proteger áreas possivelmente <i>infectadas por vírus</i> em seu computador.
Entrada do glossário	Internet download / download da Internet
Segmento original	Checks all files that you <i>download from the Internet</i> , including Java Applets and ActiveX controls
Segmento traduzido	Verifica todos os arquivos <i>obtidos por download da Internet</i> , inclusive applets Java e controles ActiveX

Estabilidade sintática:

Entrada do glossário	Internet download / download da Internet
Segmento original	To <i>download</i> the Norton Bootable Recovery Tool Wizard <i>from the Internet</i>
Segmento traduzido	Para <i>fazer download</i> do Assistente do Norton Bootable Recovery Tool <i>pela Internet</i>

A análise do silêncio poderia revelar a quantidade de itens que não são devidamente processados por um *software* de QA, ajudando a quantificar os riscos que há no processo de verificação terminológica hoje. Parece que se dedica muito tempo à conferência de falsos positivos, ao passo que a ferramenta não cobre de fato todas as ocorrências que deveria – quem sabe algumas delas não escondam erros efetivos, como o não uso da terminologia apontada no glossário ou o emprego de uma variante indesejada pelo cliente.

A análise do silêncio poderia revelar com que segurança é possível afirmar que a terminologia no glossário foi usada nas traduções.

5.5 Análise tradutória

A análise tradutória tratada nesta seção compreende o uso de técnicas de tradução vistas na revisão teórica – como generalização e transposição – e apresenta os poucos casos de erro efetivo apontados pelo *software* de QA. Conforme a análise apresentada aqui demonstra, em números absolutos, essa categorização não é significativa para a geração de ruído em unidades terminológicas ou fraseológicas. Ainda assim, as poucas ocorrências na amostra demonstram que as traduções da área de fato seguem um método interpretativo-comunicativo, contrariando o pressuposto do paradigma da localização de que, na área, os tradutores seriam conversores de códigos linguísticos (sugerindo uma tradução literal).

As próximas seções apresentam a interpretação de alguns números obtidos para ruído tradutório, descrevem em detalhes as técnicas que foram encontradas nos segmentos de tradução analisados e os casos de erro efetivo encontrados.

5.5.1 Tradução em números

A seguinte tabela apresenta uma síntese dos resultados encontrados para as ocorrências tradutórias, segmentados para cada *subcorpora*. Além do total de ocorrências, contrasta positivos e falsos positivos nos itens que o *software* incluiu na lista de QA. Os resultados apresentados consideram a influência que as técnicas de tradução têm sobre a ocorrência de falsos positivos e mostram os itens que de fato correspondem a erros de tradução.

Tabela 33 - Erros de tradução e ruído tradutório

<i>Subcorpus</i>	Falso positivo		Positivo		Total	
	#	%	#	%	#	%
Ajuda	68	69,39%	23	23,47%	91	92,86%
UI	5	5,10%	2	2,04%	7	7,14%
Total Geral	73	74,49%	25	25,51%	98	100,00%

Fonte: Elaborado pelo autor.

Esta tabela apresenta as seguintes informações:

- **Subcorpus:** indica se a ocorrência foi encontrada na ajuda ou na UI;
- **Falso positivo:** apresenta as ocorrências que o tradutor descarta ao revisar a lista de QA;
- **Positivo:** corresponde às ocorrências que de fato são erros de tradução. Idealmente, estas deveriam ser as únicas ocorrências que a lista de QA deveria gerar.

Conforme mostrado nas estatísticas da seção *Análise global dos resultados* (p. 152), os positivos (ou seja, os erros efetivos) correspondem a uma pequena porcentagem das ocorrências, somando 2,39% do total de ocorrências. Em relação às ocorrências exclusivas à tradução, estes representam 25,51% do total. A maior parte das ocorrências consiste em falsos positivos. A proporção é de três falsos positivos para cada positivo. Vale lembrar que os erros anotados aqui se referem a todas as categorias de análise, e não apenas à tradutória. Trata-se de outro aspecto que, desconsiderado pelo *software* de QA, retorna mais falsos positivos.

5.5.2 Classificação tradutória e sua influência no ruído

5.5.2.1 Técnicas de tradução

Para fins de comparação, a seguinte tabela mostra as técnicas introduzidas por Hurtado Albir na revisão teórica sobre tradução (p. 109), agora também indicando aquelas que foram encontradas na análise. Vale lembrar que o objeto dessa análise é limitado; o foco está em unidades de tradução menores, como nomes, verbos, sintagmas nominais, etc., que se encontram registrados em glossário. Frases como um todo não foram analisadas, e não se descarta

que uma análise detalhada do texto como um todo não venha a trazer à tona mais usos dessas técnicas.

Tabela 34 - Técnicas tradutórias envolvidas em ocorrências do QA

Técnica	Falso positivo		Probabilidade de ocorrência
	#	%	
Elisão	36	49,32%	Provável
Transposição	30	41,10%	Provável
Descrição	5	6,85%	Provável
Generalização	1	1,37%	Provável
Modulação	1	1,37%	Provável
Adaptação	-	-	Improvável
Ampliação linguística	-	-	Improvável
Amplificação	-	-	Provável
Calco	-	-	Improvável
Compactação linguística	-	-	Improvável
Criação discursiva	-	-	Improvável
Empréstimo	-	-	Provável
Equivalente consagrado	-	-	Improvável
Particularização	-	-	Provável
Substituição	-	-	Improvável
Tradução literal	-	-	Improvável
Variação	-	-	Parte de outra análise
Total Geral	73	100,00%	

Fonte: Elaborado pelo autor.

Esta tabela apresenta as seguintes informações:

- **Técnica:** nome da técnica de tradução;
- **Falso positivo:** total de ocorrências de ruído (em valores absolutos e porcentagens);
- **Probabilidade de ocorrência:** hipótese formulada na revisão teórica quanto à probabilidade de ocorrência de determinada técnica – provável ou improvável –, levando-se em consideração as particularidades dos textos produzidos para localização.

Com relação à frequência, as técnicas que apresentam mais ocorrências são transposição e elisão; juntas, somam 83,56% dos resultados tradutórios. Acompanhando o prognóstico

da revisão teórica, a transposição figura de fato como uma técnica de aplicação bastante provável. Há também um volume considerável de ocorrências que não foram previstas segundo a classificação de técnicas adotada, os quais serão discutidos mais adiante. As demais técnicas, descrição, generalização e modulação, apresentam ocorrências pontuais. Mais detalhes sobre os resultados para cada técnica estão em subseções logo mais adiante (pp. 181-188).

Com relação à probabilidade de ocorrência, amplificação e empréstimo, embora indicadas como possíveis, não ocorrem. Com relação a empréstimos, a hipótese era que não fossem existir ocorrências, especialmente porque o empréstimo estaria glossariado. O item que chama mais a atenção é a ampliação linguística, visto ter sido marcada como não relevante em primeiro lugar.

A tabela a seguir apresenta as técnicas que foram identificadas como causadoras de ruído na lista de QA separadas por *subcorpus*:

Tabela 35 - Técnicas causadoras de ruído por *subcorpus*

<i>Subcorpus</i> / Técnica	Falso positivo	
	#	%
Ajuda	68	93,15%
Elisão	32	43,84%
Transposição	29	39,73%
Descrição	5	6,85%
Modulação	1	1,37%
Generalização	1	1,37%
UI	5	6,85%
Elisão	4	5,48%
Transposição	1	1,37%
Total Geral	73	100,00%

Fonte: Elaborado pelo autor.

Esta tabela apresenta as seguintes informações:

- ***Subcorpus*/Técnica:** nome da técnica de tradução, agrupada por *subcorpus*;
- **Falso positivo:** total de ocorrências de ruído (em valores absolutos e porcentagens).

Observa-se que os casos pontuais de descrição, modulação e particularização ocorrem no *subcorpus* de ajuda. A UI contém apenas as técnicas de maior ocorrência na ajuda. As seguintes subseções falam de cada uma das técnicas encontradas na análise, apresentando exemplos para os casos encontrados.

5.5.2.1.1 Técnicas mais frequentes

As técnicas usadas com mais frequência são as de elisão e transposição. Os próximos parágrafos apresentam mais detalhes sobre estes casos.

Com relação às ocorrências encontradas para a técnica de elisão, sua maioria está concentrada em uma única entrada do glossário: *instant messenger / software de mensagens instantâneas*. São 31 das 36 ocorrências encontradas. Há indícios de que esta unidade não esteja formulada da maneira mais adequada no glossário ou que ela esteja inserida em alguma transformação (MT). Isso demanda um refinamento um pouco maior nos números referentes a essa unidade. Para isso, a seguinte tabela traz as ocorrências de positivos e falsos positivos para *instant messenger / software de mensagens instantâneas* em relação ao tipo de sintagma encontrado no original e na tradução:

Tabela 36 – Positivos e falsos positivos para *instant messenger / software de mensagens instantâneas*

Posição da unidade no texto	Positivo	Falso positivo	Classificação da ocorrência de ruído	Total Geral
Outra unidade	-	1	Linguística	1
Complemento de outra unidade	6	31	Tradutória, técnica de elipse	37
Mesma unidade	2	15	Variacional	17
Total Geral	8	47		55

Fonte: Elaborado pelo autor.

Esta tabela apresenta as seguintes informações:

- **Posição da unidade no texto:** indica se a unidade do glossário faz parte de um sintagma nominal maior ou não;
- **Positivo:** indica o total de ocorrências incluídas corretamente na lista pelo *software* de QA;
- **Falso positivo:** indica o total de ocorrências que o *software* de QA não deveria ter incluído na lista;
- **Classificação da ocorrência de ruído:** classificação registrada na análise para o falso positivo em questão.

Das 55 ocorrências da unidade no *corpus*, apenas oito trazem *software de mensagens instantâneas* na tradução (aquelas que de fato foram acertadas pelo *software* de QA). Em todas as ocorrências relacionadas a ruído, a unidade faz parte de um sintagma nominal; para os positivos, em seis dos oito casos a unidade também aparece como parte de um sintagma nominal. Os 15 falsos positivos em que se encontra a mesma unidade do glossário, na verdade, se referem à variação⁷⁸, e não à elipse em tradução.

Como se vê, em todos os falsos positivos referentes à técnica de elipse, a unidade *instant messenger / software de mensagens instantâneas* consiste em um sintagma nominal que complementa outros. Nas ocorrências encontradas no *corpus*, a parte omitida do sintagma é *software de*, como nos casos a seguir:

Antispyware options let you choose which categories of risk you want Norton AntiVirus to detect for manual, email, and *instant messenger* scanning.

As opções de Antispyware permitem que você escolha as categorias de risco que o Norton AntiVirus deverá detectar para a verificação manual, de e-mail e *de mensagens instantâneas*.

Scans for and detects viruses, spyware, and other risks in *instant messenger* attachments.

Verifica e detecta vírus, spyware e outros riscos em anexos de *mensagens instantâneas*.

Os demais casos para *instant messenger* são todos semelhantes. Como se vê, as traduções mudam o referencial de *software de mensagens instantâneas* para as mensagens em si. Do ponto de vista funcional, a mensagem desejada permanece, já que em última análise um anexo de um *software* de mensagens instantâneas é transferido via mensagem instantânea e uma verificação antivírus voltada para um *software* desses é feita sobre as mensagens instantâneas que ele envia. Do ponto de vista do paradigma da localização, a constatação soa como uma incoerência, a considerar a perspectiva prescritivista em termos tradutórios e terminológicos atribuída à localização, contrariando a noção de que o tradutor atua praticamente como um conversor de códigos linguísticos (produzindo traduções possivelmente literais) e segue cegamente a terminologia imposta. Claro que não é possível afirmar isso com total segurança, já que são desconhecidos os processos reais de produção desse *software* localizado nem afirmar que essa tradução final não tenha sido revisada ou finalizada por outro profissional, fora da área da tradução.

⁷⁸ O termo *programa de mensagens instantâneas* foi usado no lugar de *software de mensagens instantâneas*.

As ocorrências de outras entradas do glossário também mostram que as traduções não se limitam a transcodificar, mas sim demonstram que o tradutor considera contexto e finalidade para fazer seu trabalho. Segue um exemplo:

Entrada do glossário	content / conteúdo
Segmento original	Programs that facilitate delivery of <i>advertising content</i> to the user through their own window, or by utilizing another program's interface.
Segmento traduzido	Programas que facilitam a exibição de <i>anúncios</i> ao usuário em uma janela própria ou por meio da interface de outro programa.

Aqui, o referente muda do *conteúdo dos anúncios* para os anúncios em si. No fim das contas, a tradução em si soa mais concreta e menos pedante que o original (basta pensar em *exibição de anúncios* versus *distribuição de conteúdo publicitário*).

Por fim, para as ocorrências da técnica de transposição, apenas duas direções foram encontradas:

- de unidade verbal para nominal;
- de unidade nominal para verbal.

A passagem de unidade verbal para nominal consiste na grande maioria dos casos. A lista a seguir mostra as entradas que envolvem a técnica:

- download / fazer o download
- log on / fazer logon
- manage / gerenciar
- turn on / ativar
- renewal / renovação
- reputation analysis / análise de reputação

Embora duas das unidades verbais apresentem características evidentes de nominalização, por se tratarem de UFE com empréstimos, a entrada *turn on / ativar* corresponde a 60% das ocorrências. Mesmo descontando os muitos segmentos repetidos, o número de ocorrências exclusivas corresponde a quase 50% dos casos. Seguem alguns exemplos de unidade verbal para nominal:

Entrada do glossário	download / fazer o download
Segmento original	In this case, you must turn off Download Insight so that Norton AntiVirus lets <i>you download</i> the file and does not remove it from your computer.
Segmento traduzido	Nesse caso, você deve desativar o Download Insight para que o Norton AntiVirus permita <i>o download</i> do arquivo e não o remova de seu computador.
Entrada do glossário	turn on / ativação
Segmento original	Norton AntiVirus supports the following Media Center applications to <i>turn on</i> Quiet Mode:
Segmento traduzido	O Norton AntiVirus oferece suporte para a <i>ativação</i> do Modo silencioso automático nos seguintes aplicativos do Media Center:

Alguns casos, como no exemplo de *turn on* acima, parecem decorrer de certa dificuldade em traduzir o original. Percebe-se uma formulação que de fato dificulta um pouco a compreensão do original, visto que, pelo que se vê na tradução, ela quer dizer algo como o seguinte:

Norton AntiVirus enables you can turn on Quiet Mode for the following Media Center applications:

A seguir, um exemplo de transposição de unidade nominal para verbal:

Entrada do glossário	reputation analysis / análise de reputação
Segmento original	If you do not want Download Insight to process any file that you download for <i>reputation analysis</i> , you can turn it off.
Segmento traduzido	Caso não queira que o Download Insight <i>analise a reputação</i> dos arquivos obtidos por download, você poderá desativá-lo.

Este demonstra a dificuldade de traduzir outra unidade de tradução, a qual envolve *reputation analysis*. Trata-se de *to process* <objeto a processar> *for* <ação a ser executada no objeto>. Esse tipo de padrão é comum na área, e, a considerar o exemplo acima, pode envolver o uso dessa técnica para solucionar o problema. É o que se vê também no seguinte caso (que também poderia figurar nessa análise, mas que ficou de fora do recorte):

Scan selected items *for risks*.

Fazer a verificação de riscos nos itens selecionados.

Em geral, os casos de transposição parecem estar associados a situações pontuais de trabalho, as quais requerem uma solução *fora do convencional*, na perspectiva da localização. Não parece haver uma motivação que vise a simplificar o texto traduzido. Por exemplo, o segmento abaixo parece caminhar no sentido de simplificar:

Entrada do glossário	renewal / renovação
Segmento original	Follow the on-screen instructions to complete your subscription <i>renewal</i> .
Segmento traduzido	Siga as instruções exibidas na tela para <i>renovar</i> sua assinatura.

Já o seguinte caso vai na direção oposta, ao transpor vários verbos por nomes, incluindo o termo *manage* / *gerenciar*:

It also contains links to scan your computer, view the history of protection events, and *manage* quarantined items.

Ele também contém links para a verificação do computador, a visualização do histórico de eventos de proteção e *gerenciamento* de itens em quarentena.

5.5.2.1.2 Casos pontuais

A análise tradutória também revelou alguns casos pontuais para o uso das técnicas de descrição, generalização e modulação. Mais detalhes são apresentados a seguir.

Para a técnica de descrição, as ocorrências encontradas envolvem apenas uma entrada do glossário, *risk program* / *programa de risco*. A considerar as ocorrências encontradas, parece se tratar de um caso de inconsistência terminológica ou de variação. No *corpus* como um todo, o termo *programa de risco* coocorre com traduções descritivas para *risk program*. No entanto, a coocorrência não é observada dentro do mesmo contexto (por exemplo, em um

mesmo tópico da ajuda do *software*). . Isso sugere, por exemplo, que vários tradutores, editores e revisores textuais podem ter se envolvido no projeto ou que parte do material já estava traduzida e não foi revisada na execução do projeto, sendo apenas incorporada ao conteúdo final. Os exemplos a seguir mostram dois casos em que *risk program* foi traduzido por uma paráfrase descritiva e outro em que o termo foi efetivamente usado:

Norton AntiVirus needs to close a suspected *security risk program* to resolve it.

O Norton AntiVirus precisa fechar um *programa suspeito de ser um risco à segurança* para resolvê-lo.

In this case, you can allow the security *risk program* remain on your computer or restore the security *risk program* if Spyware Protection removes it.

Nesse caso, você pode permitir que o *programa que coloca em risco a segurança* permaneça no computador ou restaurar *esse programa* se a Proteção contra spyware o tiver removido.

In this case, you can restore the security risk to regain the functionality of a program that requires the *risk program* to run.

Nesse caso, você poderá restaurar o risco à segurança para recuperar a funcionalidade de um programa que necessite desse *programa de risco* para ser executado normalmente.

Os dois últimos segmentos são semelhantes em sua estrutura, sugerindo a inconsistência denominativa do termo no material localizado. Neste caso específico, seria válido questionar a marcação usada neste trabalho, de ruído para as ocorrências, como se fossem aceitáveis. A opção por tratar como aceitável se baseia exclusivamente no fato de, após todo um ciclo de produção, passando por revisões possivelmente pautadas pela consistência terminológica, o texto ainda apresente esse tipo de ocorrência para o termo. Nas ocorrências aqui analisadas, não parece um caso em que a presença de variante causa grandes prejuízos na compreensão do material. Inclusive, apesar de menores em número, as traduções com paráfrases descritivas parecem mais simples de compreender.

Já o caso único de generalização, em que um termo mais genérico é usado, ocorre no seguinte segmento:

Entrada do glossário	task / tarefa
Segmento original	In this case, Idle Time Scans do not run because of the CPU-intensive <i>task</i> .
Segmento traduzido	Nesse caso, a Verificação durante tempo ocioso não é executada devido à <i>atividade</i> intensa da CPU.

Neste, a unidade *task / tarefa* foi traduzida por *atividade*. O termo *task / tarefa* em informática geralmente se refere a um programa ou rotina que está em execução. No exemplo acima, este é o caso, como se depreende do parágrafo em que o segmento acima se encontra:

Some of the predefined parameters that Norton AntiVirus checks are CPU idle time, disk usage, and type of electric current you use to operate your computer.

For example, you set the Idle Time Out duration as 10 minutes and *watch an online video for 11 minutes without any intervention*. In this case, Idle Time Scans do not run because of the CPU-intensive task.

Alguns dos parâmetros predefinidos que o Norton AntiVirus verifica são tempo ocioso da CPU, uso do disco e tipo de corrente elétrica que você usa para operar seu computador.

Por exemplo, você pode definir o Tempo limite de ociosidade em 10 minutos e *assistir a um vídeo online durante 11 minutos sem intervenções*. Nesse caso, a Verificação durante tempo ocioso não é executada devido à atividade intensa da CPU.

Contudo, na tradução, optou-se por generalizar a referência, reduzindo a associação direta do ato de assistir a um vídeo *on-line* como uma tarefa executada pela CPU. Do ponto de vista da localização e internacionalização, a tradução para o português brasileiro estaria dentro do desejável, visto que reduz a dependência que o segmento de tradução tem do contexto (facilitando, assim, a tarefa de adaptar seu reuso em outros textos no futuro). Também não se descarta a possível influência de isolamento contextual na atividade tradutória, em função de como uma tradução é processada nessa área.

Por fim, há um caso isolado de modulação:

Entrada do glossário	schedule / agendar
Segmento original	Norton AntiVirus prompts you that the Idle Time Scans option gets turned off when you <i>schedule</i> a Full System Scan.
Segmento traduzido	O Norton AntiVirus solicita que a opção Verificações durante tempo ocioso seja desativada quando uma Verificação completa do sistema <i>for agendada</i> .

Como se vê, o foco da ação é movido do agente (o usuário leitor, retratado por *you*) para o objeto que sofre a ação (a *Verificação completa do sistema*). Esse caso envolve uma unidade que compreende estruturas superficiais (*agendar, agendamento, agendado*). A situação é semelhante ao que ocorre com todas as outras ocorrências para o modelo transformacional, conforme será tratado mais adiante na análise linguística, em *Modelo transformacional* (p. 224).

5.5.2.2 Erros de tradução

Como já foi falado em *Tradução em números* (p. 177), os erros efetivos representam a menor parte das ocorrências tradutórias. A seguinte tabela apresenta os erros efetivos encontrados, classificados segundo dois grandes grupos propostos por Hurtado Albir para a classificação de erros, a saber, compreensão do original e expressão na língua de chegada⁷⁹:

⁷⁹ Cf. *Erro de tradução* (p. 112).

Tabela 37 - Ocorrências para erros de tradução

<i>Subcorpus/</i> Erro de tradução	Positivo	
	#	%
Ajuda	23	92,00%
Compreensão do original -> Referência extralinguística malso- lucionada	3	12,00%
Compreensão do original -> Sentido incorreto	2	8,00%
Expressão na língua de chegada -> Léxico	1	4,00%
Expressão na língua de chegada -> Ortografia e pontuação	17	68,00%
UI	2	8,00%
Compreensão do original -> Referência extralinguística malso- lucionada	1	4,00%
Expressão na língua de chegada -> Ortografia e pontuação	1	4,00%
Total Geral	25	100,00%

Fonte: Elaborado pelo autor.

Esta tabela apresenta as seguintes informações:

- **Subcorpus/Erro de tradução:** mostra o tipo de ocorrência de erro de tradução, juntamente com sua categoria, para cada *subcorpus*;
- **Positivo:** total de ocorrências de erros efetivos de tradução em valores absolutos e porcentagens.

Dado o grande volume de informações do *corpus*, essa pequena quantidade de erros de tradução (25 no total) permite afirmar que o trabalho apresenta um alto nível de qualidade tradutória do ponto de vista terminológico. As próximas seções trazem um apanhado de cada um dos erros encontrados.

5.5.2.2.1 Compreensão do original

Os erros de compreensão do original estão todos relacionados ao não uso das traduções apresentadas no glossário, mas a grande maioria parece casos de variação. Os parágrafos a seguir descrevem os casos encontrados e discutem por que tais ocorrências não são consideradas falsos positivos.

Em primeiro lugar, há alguns casos de referências linguísticas malsolucionadas. Todos os casos consistem em variantes lexicais ou morfossintáticas. Uma pequena sutileza na vari-

ante morfossintática encontrada permite tratá-la com mais certeza como um erro, como se vê a seguir:

Entrada do glossário	Insight Network Scan / Verificação da Rede Insight
Segmento original	<i>Insight Network Scan</i>
Segmento traduzido	<i>Verificação de rede Insight</i>

Neste exemplo, a tradução soa como uma variante morfossintática, em que a única diferença está na caixa-baixa de *rede* em *rede Insight*. Contudo, *Rede Insight* consiste em um referente externo bem delimitado. Constam as seguintes entradas no glossário envolvendo *Insight Network*:

- Insight Network / Rede Insight
- Norton Insight Network / Rede do Norton Insight
- Insight Network Scan / Verificação da Rede Insight

Há indícios no *corpus* de que *Insight Network* e *Norton Insight Network* tenham o mesmo referencial, o qual é uma tecnologia que registra a reputação de arquivos transferidos via Internet. Nesse caso, essa tecnologia é tratada como um produto, um nome próprio, nos moldes de Norton AntiVirus ou Microsoft Word.

Já nos casos das variantes morfossintáticas, não é possível afirmar o erro com tanta certeza por causa da diferença de datação do glossário e do *software* localizado:

Entrada do glossário	Symantec Store / Loja online da Symantec
Segmento original	If you downloaded the product from the <i>Symantec Store</i>
Segmento traduzido	Se você fez o download do produto pela <i>Symantec Store</i>
Segmento original	The Product Key is stored on your computer as part of the download process and is included in the confirmation email from the <i>Symantec Store</i> .
Segmento traduzido	O código do produto estará armazenado no seu computador como parte do processo de download e incluído no e-mail de confirmação enviado pela <i>Symantec Store</i> .

As duas ocorrências para *Symantec Store* acima são as únicas no *corpus* inteiro. Sendo assim, não é possível afirmar que o termo não tenha sido glossariado como *Symantec Store* /

Symantec Store, tendo sofrido uma alteração posterior à execução desse projeto de localização – ou idas e vindas de *Symantec Store* para *loja online da Symantec*. Outra possibilidade é que o termo tenha sido colocado no glossário mais tarde e, até o momento de sua inserção, tenha cabido a tradutores diferentes ou ao mesmo tradutor em momentos diferentes solucionar a tradução.

Os demais erros referentes à compreensão do original se referem a traduções com sentido incorreto. A falta de uso das entradas glossariadas acusa duas traduções que modificam o sentido. Estes são os únicos casos envolvendo problema de compreensão que ocorrem na análise. Segue o primeiro caso:

Entrada do glossário	Norton AntiVirus / Norton AntiVirus
Segmento original	In the Norton AntiVirus main window Control Center, click Fix Now.
Segmento traduzido	Na janela principal do centro de controle, clique em Corrigir agora.

A unidade de tradução *Norton AntiVirus main window Control Center*, embora soe um tanto extensa para um nome composto em inglês, deveria ser traduzida por *o centro de controle da janela principal do Norton AntiVirus*. Não se trata de uma janela autônoma para esse centro de controle, mas sim de uma parte da janela principal do antivírus. Aqui, a falta de *Norton AntiVirus* até seria aceitável, a considerar o contexto geral. Nesse caso, o usuário está consultando a referência do *software* Norton AntiVirus, e a seção em questão já deixava isso claro:

Figura 24 - Referência do *software* Norton AntiVirus

You can respond to the System Status indicators directly from the main window.	Você pode responder aos indicadores de Status do sistema diretamente da janela principal.
To respond to System Status indicators from the main window	Para responder aos indicadores de Status do sistema na janela principal
1. In the Norton AntiVirus main window Control Center, click Fix Now .	1. Na janela principal do centro de controle, clique em Corrigir agora .
2. Follow the on-screen instructions.	2. Siga as instruções exibidas na tela.

Fonte: *Subcorpus* de ajuda (p. 123).

Logo, o erro não está diretamente ligado à ausência da unidade, mas sim se trata de uma coincidência, foi detectado por ação fortuita. Já no segundo caso, a ausência da unidade está diretamente ligada ao erro:

Entrada do glossário	reputation / reputação
Segmento original	technology wherein a remote Symantec server on the Web stores the latest reputation information.
Segmento traduzido	, na qual um servidor remoto da Symantec armazena as definições de vírus mais recentes.

Aqui, *reputation information* foi traduzido por *definições de vírus*. No contexto do *software*, embora as informações sobre a reputação de um arquivo podem ter relação com as definições de vírus, a informação passada não está totalmente precisa. A reputação de um arquivo, por exemplo, pode mostrar que ele é seguro, ou seja, não teria nenhuma definição de vírus associada.

5.5.2.2.2 Expressão na língua de chegada

Com relação a esta categoria, a grande maioria das ocorrências consiste em questões de espaçamento ou ortografia. Assim como os casos descritos em *Referência extralinguística malsolucionada* acima, todas as ocorrências a seguir podem ser consideradas casos de variação. Os parágrafos a seguir descrevem os casos encontrados e discutem por que tais ocorrências, apesar de variantes, não são consideradas ruído.

Com relação a léxico, há uma ocorrência em que a tradução apresentada no glossário muda em termos lexicais:

Entrada do glossário	risk / risco
Segmento original	You choose what to do at each phase of the <i>risk</i> response process.
Segmento traduzido	Você define o que fazer a cada fase do processo de resposta ao <i>vírus</i> .

Nesse caso, a unidade *risk / risco* foi traduzida por *vírus*. O caso poderia ser considerado uma particularização, visto que um vírus consiste em um risco nessa área específica.

Contudo, nesse caso, a unidade *risk* também se refere a outras ameaças que podem causar riscos a um computador ou ao usuário, como *worms*, *spyware* e *hackers*. O segmento de tradução acima ocorre dentro de uma tabela, na seguinte seção, especificamente (o grifo em negrito e itálico foi inserido para facilitar a visualização do segmento):

Figura 25 - Contexto para segmento de tradução envolvendo risk / risco

<p>How to increase protection</p>	<p>You can increase the email protection using the following options:</p> <ul style="list-style-type: none"> • Scan outgoing messages for suspected worms Turns on scanning to detect suspected worms in your outgoing email messages and alerts you if a threat is detected • How to respond when an outbound threat is found If Norton AntiVirus detects a threat in an outgoing email message, you have the following options: <ul style="list-style-type: none"> ○ Automatically remove [recommended] Provides you the greatest protection by automatically removing detected suspected email worms from your computer without interacting with you ○ Ask me what to do Provides you the greatest control when a suspected email worm is detected. <i>You choose what to do at each phase of the risk response process.</i> A summary window appears detailing all of the actions that were taken
<p>Como aumentar a proteção</p>	<p>Você pode aumentar a proteção de e-mail usando as seguintes opções:</p> <ul style="list-style-type: none"> • Verificar suspeitas de worm em mensagens enviadas Ativa a verificação para detectar suspeitas de worm nas mensagens de e-mail enviadas e alerta você se uma ameaça for identificada • Como responder ao encontrar uma ameaça de saída Se o Norton AntiVirus detectar uma ameaça em uma mensagem de e-mail enviada, você terá as seguintes opções: <ul style="list-style-type: none"> ○ Remover automaticamente [recomendado] Fornece maior proteção ao remover automaticamente worms de e-mail suspeitos e detectados do computador sem interagir com você ○ Perguntar o que fazer Fornece maior controle quando uma suspeita de worm de e-mail é detectada. <i>Você define o que fazer a cada fase do processo de resposta ao vírus.</i> Uma janela de resumo será exibida com os detalhes de todas as ações tomadas

Dependendo do ponto de vista, esse caso também pode ser considerado como um problema de compreensão. Nos dois casos, o resultado final é um acerto por parte do *software* de QA.

Já com relação aos casos referentes à ortografia e pontuação, um se destaca dos demais, pois representa a grande maioria e envolve todas as unidades do glossário que contêm o termo *on-line*. Primeiro, seguem os demais casos:

Entrada do glossário	File Insight / File Insight
Segmento original	The Details & Settings link in the notification alert lets you view additional details about the resource consumption by the program in the <i>File Insight</i> window.
Segmento traduzido	O link Detalhes e configurações no alerta de notificação permite exibir detalhes adicionais sobre o consumo de recurso pelo programa na janela File Insight .

Aqui, falta um espaço entre *janela* e *File*. Esse tipo de problema também é facilmente detectado por um recurso de verificação ortográfica, como o encontrado em editores de texto.

Entrada do glossário	Web page / página da Web
Segmento original	When you click the Scan Facebook Wall option, Norton AntiVirus takes you to the Facebook login <i>Web page</i> .
Segmento traduzido	Quando você clicar na opção Verificar mural do Facebook, o Norton AntiVirus levará você à <i>página da Web</i> de login do Facebook.

Já nesse caso, a palavra *página* está escrita sem acento. Outra ocorrência detectável via verificação ortográfica.

Com relação a *on-line*, como foi dito acima, as ocorrências envolvem várias unidades presentes no glossário, a saber:

- online backup / backup online
- online help / ajuda online
- online storage / armazenamento online
- online tutorial / tutorial online

Nas traduções, a ortografia usada para *on-line* é “*on-line*”, como se vê a seguir:

Depois de configurar sua conta, você poderá acessá-la pelo Centro de controle e acessar o status de seu backup *on-line*.

Você não pode acessar a Ajuda *on-line* durante a desinstalação. Portanto, imprima o tópico da Ajuda referente à desinstalação do Norton AntiVirus antes de continuar.

Para obter mais informações sobre os produtos da Symantec que oferecem armazenamento *on-line*, vá para:

Além de visualizar detalhes, também é possível usar o link Saiba mais para acessar a página do tutorial *on-line* e obter mais informações sobre o arquivo.

Isso sugere que, na época em que o material foi localizado, a grafia corrente determinada pela empresa seria *on-line*, fato que teria mudado mais tarde, a considerar a grafia atual encontrada no glossário. Sendo assim, para a época, a proporção total de erros poderia ser ainda menor, desconsiderando a maioria das ocorrências de erro por ortografia, já que *on-line* provavelmente não teriam sido assinalado no momento em que este projeto foi executado. Ainda assim, esse caso representa um fato que não é tão raro na área: atualização pontual – por vezes *ad hoc* – da terminologia. Uma breve incursão no *site* da Symantec dá uma amostra da questão:

Tabela 38 - Grafias de *on-line* ao longo do tempo no *site* da Symantec

Grafia de <i>on-line</i>	Data da publicação da página	URL
on-line	2006	http://www.symantec.com/pt/br/about/news/release/article.jsp?prid=20061019_01
online	2008	http://www.symantec.com/pt/br/about/news/release/article.jsp?prid=20081125_01
on-line	2010	https://support.norton.com/sp/pt/br/home/current/solutions/20080702134123BR?entsrc=redirect_pubweb&product=home&pvid=f-home&version=1
online	2011	https://support.norton.com/sp/pt/br/norton-management/current/solutions/v1027402_NIS_OEM_2012_pt_br?ispid=&product=Norton+Management&pvid=f-nm&version=current
online	2013	http://buy.norton.com.gtm.symantec.com/pt-br/mf/premiumServices

Fonte: *Sites* da Symantec (2006, 2008, 2010, 2011b, 2013c).

Em última análise, o caso também demonstra que, para atualizações ortográficas pontuais, envolvendo unidades que não sofrem alteração morfológica de nenhum tipo (*on-line* em português não flexiona em número nem gênero), o *software* de QA dá conta do recado.

5.6 Análise variacional

A análise variacional tratada nesta seção compreende a catalogação das ocorrências entre alterações simples e complexas segundo Freixa. Conforme a análise apresentada aqui demonstra, em números absolutos, essa categorização não é significativa para a geração de ruído em unidades terminológicas ou fraseológicas. Ainda assim, as ocorrências na amostra demonstram que as traduções da área de fato apresentam o uso de variantes, contrariando o pressuposto do paradigma da localização de que, na área, os tradutores seguiriam o glossário em todos os casos, mantendo a consistência terminológica prescrita pelo cliente.

As próximas seções apresentam a interpretação de alguns números obtidos para ruído variacional na análise, descrevem em detalhes os casos de variação na tipologia de Freixa (2002) que são encontrados nos segmentos de tradução analisados e discorrem sobre as causas para variação, também dentro da perspectiva da autora.

5.6.1 Variação em números

Nesta seção, são apresentados alguns cruzamentos dos dados obtidos na análise de ruído decorrente de variação terminológica, discorrendo sobre a incidência de falsos positivos em relação aos positivos.

A seguinte tabela apresenta uma síntese dos resultados encontrados para as ocorrências variacionais, segmentados para cada *subcorpora*. Além do total de ocorrências, traz a relação entre os positivos e falsos positivos encontrados na lista de QA.

Tabela 39 - Ruído decorrente de variação por *subcorpus*

<i>Subcorpus/</i> Tipo de variação	Falso positivo		Positivo		Total	
	#	%	#	%	#	%
Ajuda	195	81,25%	24	10,00%	219	91,25%
Alteração simples	90	37,50%	22	9,17%	112	46,67%
Alteração complexa	105	43,75%	2	0,83%	107	44,58%
UI	19	7,92%	2	0,83%	21	8,75%
Alteração simples	7	2,92%	2	0,83%	9	3,75%
Alteração complexa	12	5,00%		0,00%	12	5,00%
Total Geral	214	89,17%	26	10,83%	240	100,00%

Fonte: Elaborado pelo autor.

Esta tabela apresenta as seguintes informações:

- **Subcorpus/Tipo de variação:** indica o *subcorpus* e a classificação variacional mais geral das ocorrências;
- **Falso positivo:** apresenta as ocorrências de variação que representam ruído;
- **Positivo:** corresponde às ocorrências que de fato são erros de tradução. Idealmente, estas deveriam ser as únicas ocorrências que a lista de QA deveria gerar.

Como a tabela mostra, a maior distribuição de variantes está na ajuda. Isso poderia sugerir, por exemplo, que há uma preocupação maior em manter a consistência dos termos na UI. Há indícios disso nas ocorrências de unidades que apresentam variação somente na ajuda. O seguinte quadro mostra as ocorrências de variação partilhadas entre UI e ajuda e as ocorrências de ruído exclusivas a cada uma:

Quadro 3 - Variantes por *subcorpus*

Compartilhadas	Exclusivas à ajuda	Exclusivas à UI
<ul style="list-style-type: none"> • desktop / área de trabalho (*) • download / fazer o download • incoming / de entrada (*) • Insight Network Scan / Verificação da Rede Insight • instant messenger / <i>software</i> de mensagens instantâneas • online backup / backup online • outgoing / de saída (*) • rescan / reavaliação 	<ul style="list-style-type: none"> • disk usage / uso do disco • online help / ajuda online • online storage / armazenamento online • online tutorial / tutorial online • product upgrade / upgrade de produto • scan history / histórico da verificação • Self-Service Options / Opções de autoatendimento • Symantec Store / Loja online da Symantec • viral / viral • Web Protection / Proteção na Web • idle time / tempo ocioso • location / local • override / sobrepor • quarantine / colocar em quarentena • report / relatar • state / estado • support / suportar • support / suporte • Symantec product / produto da Symantec • task / tarefa • Web page / página da Web 	<ul style="list-style-type: none"> • risk alert / alerta de riscos • system scan / verificação do sistema

Fonte: Elaborado pelo autor.

Com relação às ocorrências gerais, também se vê que a maior parte dos casos de variação consiste em ruído na lista do *software* de QA. Poucos casos são positivos, ou seja, erros efetivos (para detalhes sobre as ocorrências de erro efetivo, consulte *Erros de tradução* na página 188).

Para finalizar, segue um mapa da tipologia de Freixa, indicando em negrito mais itálico os casos de variação detectados nesta análise:

Tabela 40 - Ocorrências variacionais no *corpus* segundo a tipologia de Freixa

Ocorrência variacional	Tipo	Subtipo	Exemplo ⁸⁰
I - Alteração gráfica	1. termo e forma artificial	a) termo e símbolo	-
		b) termo e fórmula química	-
		c) termo e outra fórmula artificial	-
	2. termo e abreviação	d) termo e sigla	-
		e) termo e abreviatura	system scan / verificação do sistema Verific. completa do sist. Verific. completa do sistema
	3. alteração ortográfica		online backup / backup online backup on-line
II - Alteração morfossintática	1. permanência da estrutura	a) alteração/presença de artigo	Quarantine / colocar em quarentena colocar na quarentena
		b) alteração de número	scan history / histórico da verificação histórico de verificações
		c) alteração de preposição	Web page / página da Web página na Web
		d) alteração de gênero	màxima absoluta / màxim absolut
	2. alteração na estrutura	e) [N+A] / [N+SP]	idle time / tempo ocioso tempo de ociosidade
		f) monolexical / polilexical	producte ecològic / ecoproducte
	g) outras alterações na estrutura	matèria primera / primera matèria	
III - Redução	1. reduções na extensão	a) Tipo [N+X] = [N]	
		[N+A] = [N]	desktop PC / PC desktop PC
		[N+SP] = [N]	Web page / página da Web página

⁸⁰ Para os casos de variação detectados, os exemplos são retirados do *corpus*. A primeira tradução é a que consta em glossário; a segunda é variante encontrada nas traduções.

	$[N_1+N_2] = [N_1]$	-
	b) Tipo $[N+X+Y] = [N+X]$	-
	$[N+SP+A] = [N+SP]$	-
	$[N+A+SP] = [N+A]$	-
	$[N+A_1+A_2] = [N+A_1]$	-
	$[N+SP_1+SP_2] = [N+SP_1]$	-
	c) outras reduções na extensão	-
	d) $[N+A] = [A]_N$	-
2. reduções na base	e) $[N_1+N_2] = [N_2]$	-
	f) $[N_1+SP (de+N_2)] = [N_2]$	-
3. outras reduções		-
IV - Alteração lexical	1. unidades monolexicais	override / sobrepor substituir
	a) alteração na base	download / fazer o download obter por download
	$[N+SP (prep + art + N)]$	Self-Service Options / Opções de autoatendimento opções de autosserviço
	$[N+A]$	outgoing / de saída enviado
	outras alterações na base	-
	b) alteração na extensão	-
	$[N+SP (prep + art + N)]$	-
2. unidades polilexicais		

		[N+A]	-
		outras alterações na extensão	-
V - Diversas alterações complexas	1. com semelhança formal	a) monolexical / polilexical	rescan / reverificação repetir verificação
		b) polilexical / polilexical	-
	2. sem semelhança formal	c) monolexical / polilexical	viral / viral de vírus
		d) polilexical / polilexical	Symantec Store / Loja online da Symantec Symantec Store

Fonte: Elaborado pelo autor com base em Freixa (2002).

Como se vê, uma grande variedade dos tipos indicados existe efetivamente no *corpus*. A tabela a seguir traz os números para as ocorrências específicas, destacando os tipos e subtipos:

Tabela 41 - Ocorrências por subtipo de variação

Subtipo	Falso positivo		Positivo		Total	
	#	%	#	%	#	%
Alteração simples	97	40,42%	24	10,00%	121	50,42%
Alteração gráfica → Alteração ortográfica		0,00%	22	9,17%	22	9,17%
Alteração gráfica → Termo e abreviação → Termo e abreviatura	3	1,25%		0,00%	3	1,25%
Alteração morfossintática → Alteração na estrutura → [N+A] / [N+SP]	11	4,58%		0,00%	11	4,58%
Alteração morfossintática → Permanência da estrutura → Alteração de nome	7	2,92%		0,00%	7	2,92%
Alteração morfossintática → Permanência da estrutura → Alteração de preposição	10	4,17%		0,00%	10	4,17%
Alteração morfossintática → Permanência da estrutura → Alteração/presença de artigo	66	27,50%	2	0,83%	68	28,33%
Alteração complexa	117	48,75%	2	0,83%	119	49,58%
Alteração lexical → Unidades monolexicais	16	6,67%		0,00%	16	6,67%
Alteração lexical → Unidades polilexicais → Alteração na base → [N+SP (prep + art + N)]	16	6,67%		0,00%	16	6,67%
Alteração lexical → Unidades polilexicais → Alteração na base → Outras alterações na base	28	11,67%		0,00%	28	11,67%
Alteração lexical → Unidades polilexicais → Alteração na extensão → [N+SP (prep + art + N)]	1	0,42%		0,00%	1	0,42%
Diversas alterações complexas → Com semelhança formal → Monolexical / polilexical	11	4,58%		0,00%	11	4,58%
Diversas alterações complexas → Sem semelhança formal → Monolexical / polilexical	41	17,08%		0,00%	41	17,08%
Diversas alterações complexas → Sem semelhança formal → Polilexical / polilexical	1	0,42%	2	0,83%	3	1,25%
Redução → Reduções na extensão → [N+A]= [N]	1	0,42%		0,00%	1	0,42%
Redução → Reduções na extensão → [N+SP] = [N]	2	0,83%		0,00%	2	0,83%
Total Geral	214	89,17%	26	10,83%	240	100,00%

Fonte: Elaborado pelo autor.

Esta tabela apresenta as seguintes informações:

- **Ocorrência variacional / Tipo e subtipo de variação:** indica os tipos e subtipos de variação que ocorrem, separados por alterações complexas e simples;
- **Falso positivo:** apresenta as ocorrências de variação que representam ruído;
- **Positivo:** corresponde às ocorrências que de fato são erros de tradução. Idealmente, estas deveriam ser as únicas ocorrências que a lista de QA deveria gerar.

A seção a seguir discute cada uma das ocorrências apontadas nessa tabela, trazendo exemplos e apontando para as possíveis causas.

5.6.2 Ocorrências de variação e sua influência no ruído

5.6.2.1 Alteração simples

Das ocorrências para alterações simples, os dois casos ocorrem no *corpus*: alteração gráfica e morfossintática. Do primeiro, os tipos encontrados consistem na coocorrência de termo e abreviatura e de alteração ortográfica. Com relação à ortográfica, trata-se do caso já tratado como erro na análise tradutória, das unidades do glossário que incluem o termo *on-line*. Com relação às morfossintáticas, a maior parte está atrelada a possíveis UFE. As seções a seguir tratam dos casos em detalhes.

A alteração gráfica referente a termo e abreviatura se refere à seguinte unidade e aos seguintes segmentos:

Entrada do glossário	system scan / verificação do sistema
Segmento original	Full System Scan
Segmento traduzido	Verific. completa do sist.
	Verific. completa do sistema
Segmento original	Run Full System Scan
Segmento traduzido	Verific. completa do sist.

Todos os segmentos acima ocorrem na UI. Provavelmente as abreviaturas foram inseridas por limitação de caracteres. Trata-se de um problema comum ao texto de UI, que muitas vezes está limitado à largura de um menu ou ao tamanho de uma caixa de diálogo. A causa provável dessa variação é, portanto, de economia linguística e sua origem provável está em parâmetros definidos para o projeto de localização e, nesse sentido, é funcional.

Conforme dito acima, a alteração ortográfica se refere às unidades do glossário que incluem o termo *on-line*. Nas traduções, ocorre exclusivamente a grafia *on-line*, mas o glossário assinala *online* como a grafia a ser seguida. Uma possível razão para isso, conforme discorrido anteriormente, seria uma possível mudança na grafia posterior à conclusão do projeto de localização referente ao *corpus*. A causa, portanto, é dialetal e cronológica. Se fosse possível comprovar tal mudança, isso demonstraria a tendência da área de promover alterações terminológicas em carácter *ad hoc*.

Para mais detalhes e exemplos sobre essa alteração, consulte a seção *Ortografia e pontuação* (p. 194) da análise tradutória.

As alterações morfossintáticas representam boa parte das encontradas no *corpus*, especial em função das alterações do subtipo *alteração/presença de artigo*, os quais somam quase 1/3 das ocorrências totais de variação. Para a maioria dos casos de alteração morfossintática, a alteração mantém a estrutura da unidade glossariada, mas 4,58% dos casos consistem em alteração na estrutura também. Nos próximos parágrafos, segue uma síntese dos casos encontrados.

Para as alterações com permanência na estrutura, os primeiros casos se referem à alteração/presença de artigo. A seguinte tabela apresenta as unidades encontradas e suas variantes:

Quadro 4 - Variantes com estrutura idêntica

Unidade do glossário	Variante
download / fazer o download	fazer download
Insight Network Scan / Verificação da Rede Insight	Verificação de Rede Insight
product upgrade / upgrade de produto	upgrade do produto
quarantine / colocar em quarentena	mover para a quarentena ⁸¹
<i>scan history</i> / histórico da verificação	histórico de verificações ⁸²

Fonte: Elaborado pelo autor.

Um primeiro olhar pode sugerir que alguns casos, como o de *product upgrade* ou *scan history* sejam, na verdade, unidades registradas de maneira inadequada no glossário. Os exemplos de *scan history* sugerem algo nesse sentido:

When your computer is idle, Norton AntiVirus runs a scan depending on the *scan history* of your computer.

Quando seu computador está ocioso, o Norton AntiVirus executa uma verificação dependendo do *histórico da verificação* de seu computador.

Idle Time Scans detect the time when you do not use your computer and intelligently run scans depending on the *scan history* of your computer.

A Verificação durante tempo ocioso detecta quando você não usa seu computador e executa de maneira inteligente verificações, dependendo do *histórico de verificações* do computador.

Os segmentos são muito semelhantes e aparecem em contextos semelhantes, em seções estruturadas, porém autônomas. E as traduções contêm variantes aparentemente desnecessárias. Na perspectiva da localização, a considerar que o usuário não acessa necessariamente os mesmos blocos de informação, não haveria motivos, por exemplo, discursivos para tal variante. Em última análise, com relação à tradução, ambas se referem ao registro das ocasiões em que o antivírus é executado no computador.

⁸¹ Esse caso consiste ainda em uma alteração complexa, visto que envolve também uma variante lexical, de *colocar* para *mover*.

⁸² Esse caso consiste ainda em uma alteração na palavra, já que *verificação* está no plural.

Já em *product upgrade*, a questão está atrelada à situação discursiva, ao contexto, como se vê nos seguintes casos:

These differ from *product upgrades*, which are newer versions of the entire product.

Elas são diferentes dos *upgrades de produto*, que são versões mais recentes do produto completo.

If you do not want to get *product upgrade* alerts, you can turn off this option.

Se não desejar receber alertas de *upgrade do produto*, você poderá desativar essa opção.

Nesse caso, a primeira ocorrência se refere a *upgrades* em geral, enquanto a segunda fala especificamente dos casos de *upgrade* do produto Norton AntiVirus. Sendo assim, de certa maneira, a causa para esse caso poderia ser apontado como cognitiva, de imprecisão conceitual.

Uma característica que se destaca em todas as unidades apontadas acima. Todas têm algo em comum, são (potencialmente) fraseológicas⁸³ e, nessa perspectiva, estariam mais propensas a aceitar a inserção de elementos, como a presença de um artigo.

Agora, com relação à alteração de número, a seguinte tabela apresenta as unidades encontradas e suas variantes:

Quadro 5 - Variantes com alteração de número

Unidade do glossário	Variante
download / fazer o download	fazer downloads
risk alert / alerta de riscos	alerta de risco
scan history / histórico da verificação	histórico de verificações

Fonte: Elaborado pelo autor.

Todos os casos parecem relacionados a uma questão de imprecisão conceitual, assim como *product upgrade* apresentado acima. Contudo, a unidade *alerta de riscos* não aparece no *corpus*, e o caso de *histórico de verificações*, como se viu acima, parece mais uma questão de incoerência na tradução. Outra vez, as unidades todas são possivelmente fraseológicas.

⁸³ Salvo *download* e *quarantine*; mas suas traduções também são fraseológicas.

Por fim, a respeito de alteração de preposição, a seguinte tabela apresenta as unidades encontradas e suas variantes:

Quadro 6 - Variantes com alteração de preposição

Unidade do glossário	Variante
quarantine / colocar em quarentena	mover para a quarentena
Web page / página da Web	página na Web
Web Protection / Proteção na Web	proteção da Web

Fonte: Elaborado pelo autor.

No primeiro caso, pelo que se pôde observar, a variante também existe no original, pois o termo *move to quarantine* coocorre com o verbo *quarantine*. No caso detectado pelo *software*, trata-se de *quarantine / mover para a quarentena*, no entanto. Aqui, não se descarta uma causa cronológica, de traduções executadas em momentos distintos (mesmo que o período que as separa seja de, por exemplo, poucos meses). Outra possibilidade é que a causa seja dialetal de variação social; aqui, a ocorrência poderia ser explicada, por exemplo, por dois tradutores distintos (ou duas equipes de tradução) executando o projeto, um a cargo do *software* e o outro, da ajuda.

Com relação aos casos de *Web page*, talvez estes possam ser explicados por causas dialetais de variação social:

Lets you access the Symantec **Web page** to get more information about the file.

Permite acessar a **página** da Symantec **na Web** para obter mais informações sobre o arquivo.

In some cases, the Symantec Security Response **Web page** provides steps on removing the non-viral threat from your computer.

Em alguns casos, a **página** do Symantec Security Response **na Web** apresenta as etapas para a remoção da ameaça não viral do computador.

Talvez, em termos de adequação, um sintagma encadeado como o indicado acima seria preferível a, por exemplo:

If you forget your Norton Account password, you can get a temporary password by clicking the Forgot your password link in the Norton Account sign-in *Web page*.

Se você esquecer a senha da Norton Account, será possível obter uma senha temporária clicando no link Esqueceu sua senha, na *página da Web* de acesso da Norton Account.

As ocorrências de *página da Web* e *página na Web* são parênteses (nove para sete, respectivamente). As ocorrências da última, contudo, parecem mais adequadas estilisticamente. Um tradutor experiente, com bom conhecimento da língua portuguesa, optaria pelo uso de *página na Web* tal qual apresentado nos primeiros exemplos, mesmo que o glossário aponte para uma preposição específica, por saber que a regência de um termo como *Web*, por se tratar de uma localidade, permite a preposição *em*.

Dois outros comentários a respeito desse caso: primeiro, *página na Web* deflagra casos de inserção que não correspondem a UFE (conforme visto em *Estabilidade sintática – inserção* na página 167); segundo, isso se opõe à lógica da consistência terminológica difundida na localização.

Já a alteração na estrutura trata de duas unidades específicas, e o padrão encontrado é o mesmo para ambas: [N+A] / [N+SP]. Apenas a direção, como se vê na tabela a seguir, é diferente para cada uma:

Quadro 7 - Variantes com alteração de estrutura

Unidade do glossário	Variante
idle time / tempo ocioso	tempo de ociosidade
Symantec product / produto da Symantec	produto Symantec

Fonte: Elaborado pelo autor.

No primeiro caso, a grande maioria das ocorrências é para *tempo ocioso*. Com relação aos casos da variante *tempo de ociosidade*, não parece haver um padrão específico para sua ocorrência:

The Norton Tasks window lets you view and monitor all Norton-specific background tasks that Norton AntiVirus runs during *idle time*.

A janela Tarefas do Norton permite que você visualize e monitore todas as tarefas em segundo plano específicas do Norton executadas pelo Norton AntiVirus durante o *tempo de ociosidade*.

The details helps you decide whether to start a background task or wait for Norton AntiVirus to run the job during *idle time*.

Os detalhes ajudam você a decidir se deve iniciar uma tarefa em segundo plano ou esperar até que o Norton AntiVirus a execute durante o *tempo ocioso*.

Os dois segmentos acima coocorrem, e a variante é a primeira a ocorrer na seção em questão. A grande maioria é da forma glossariada (3:1 em quatro ocorrências no total).

Já a segunda unidade, *Symantec product*, apresenta um número de ocorrências maior para *produto Symantec*. Em marketing, é prática comum não empregar preposição + artigo em sintagmas que vinculam um produto ou empresa a algo, usando o nome diretamente, como se fosse adjetivo; talvez essa seja a origem dessa variante.

Os motivos para todos os casos discutidos, tal qual apontados acima para as alterações de permanência na estrutura, poderiam ser cronológicos ou dialetais. Outra vez, a aceitação e publicação dessas variantes contraria a lógica da consistência terminológica da localização.

5.6.2.2 Alteração complexa

Das ocorrências para alterações complexas, os três casos ocorrem no *corpus*: redução, alteração lexical e diversas alterações complexas. Do primeiro, o único tipo encontrado consiste em redução na base. Com relação à alteração lexical, todos os tipos foram encontrados, ou seja, alteração em unidades monolexicais e polilexicais. Quanto aos subtipos, para o segundo grupo, há casos de alterações na base e na extensão (as unidades monolexicais não estão classificadas em nenhum subtipo). Por fim, para o tipo *diversas alterações complexas*, o único subtipo que não ocorre é *polilexical / polilexical* nas alterações com semelhança formal. Os parágrafos a seguir trazem os detalhes.

Na amostra analisada, os casos de redução estão em menor número que as demais alterações complexas. Como foi afirmado, trata-se de reduções na extensão da unidade. O quadro a seguir apresenta as ocorrências encontradas, incluindo o padrão de redução envolvido:

Quadro 8 - Variantes com redução de extensão

Unidade do glossário	Variante	Padrão da variante
desktop / área de trabalho (*)	PC ⁸⁴	[N+SP]= [N]
Web page / página da Web	página	[N+SP] = [N]

Fonte: Elaborado pelo autor.

Ambos os casos acima parecem se enquadrar como questões discursivas de economia linguística. O primeiro caso pode decorrer de alguma limitação de caracteres, visto que o segmento, mostrado abaixo, ocorre na UI:

DESKTOP PC

PC

O mesmo segmento original ocorre também na ajuda. Lá, contudo, a tradução é diferente:

DESKTOP PC

PC DESKTOP

Aqui, o motivo para tal variação, assim como muitos outros já discutidos, pode ser cronológico ou dialetal. E mais uma vez, o fato de essas variantes terem sido aceitas contraria o paradigma da localização e a venda de consistência terminológica como parâmetro de garantia de qualidade (QA).

Com relação às alterações lexicais, estas representam a metade das alterações complexas. Muito disso se deve a apenas duas unidades do glossário, *download / fazer o download* e *instant messenger / software de mensagens instantâneas*.

⁸⁴ Esse caso ainda apresenta uma alteração lexical; a unidade completa, encontrada no *corpus*, é *PC desktop*, variante de *PC de área de trabalho*.

Com relação às alterações em unidades monolexicais, a tabela a seguir mostra as unidades do glossário e suas respectivas variantes:

Quadro 9 - Variantes monolexicais

Unidade do glossário	Variante
location / local	Localização
override / sobrepor	substituir; sobrescrever
report / relatar	transmitir; reportar
state / estado	Status
support / suporte	Ajuda
task / tarefa	Atividade

Fonte: Elaborado pelo autor.

Parte dessas alterações parece ser consequência da busca por uma melhor adequação contextual; talvez os envolvidos na tradução tenham considerado as alternativas ao glossário mais adequadas de acordo com o contexto. Por exemplo, para *location*:

É possível visualizar detalhes, como o ícone padrão do programa, o nome do programa e a *localização* do programa.

O System Insight oferece um *local* centralizado onde você pode visualizar e monitorar as atividades do sistema.

Percebe-se que, no primeiro caso, o efeito de localizar está mais presente que no segundo, o qual aponta para algo mais concreto. No entanto, assim como outros casos de variação vistos antes, aqui também há incoerências. Há muitas ocorrências de *local* cuja situação assemelha-se ao caso de *localização* acima, como a seguinte:

Navegue até o *local* do arquivo cujos detalhes deseja verificar.

Logo, a análise retorna outra vez ao ponto em que não é possível afirmar ao certo a causa da variação. Pode ser cronológica (traduções de segmentos para o mesmo projeto em momentos distintos) ou dialetal (traduções para o mesmo projeto feitas por equipes distintas ou geograficamente separadas). Seja qual for a causa específica, isso vem a reforçar a dificul-

dade de se analisar alguns aspectos específicos aos projetos de localização em função da lógica da distribuição presente na área (cf. *Distribuição* na página 98). Embora fora de escopo, essa variante também envolve uma alteração conceitual, visto que, embora relacionáveis, as unidades *local* e *localização* não se referem exatamente ao mesmo referencial em cada caso.

A situação das demais unidades na tabela acima não é diferente. É possível especular as mesmas causas, e os exemplos são parecidos, como neste caso de *override*:

Se você não desejar *sobrescrever* suas configurações manuais das conexões de rede, desative essa opção.

Também é possível *substituir* sua configuração e desativar manualmente o Modo silencioso a qualquer momento.

Aqui, é praticamente a situação de *local* e *localização*. Com relação a *transmitir*, tem-se o mesmo:

Os cookies de rastreamento podem então *transmitir* a terceiros as informações obtidas.

Os cookies de rastreamento podem *reportar* essas informações a terceiros.

Já com relação às unidades polilexicais, como afirmado acima, a maior parte se deve a apenas duas unidades. Uma é notadamente fraseológica em português: *download* / *fazer o download*; a outra é *instant messenger* / *software de mensagens instantâneas*, a qual foi discutida na análise de tradução como um caso de elisão (p. 181). Ao contrário das alterações monolexicais, aqui há menos unidades do glossário. Um motivo provável consiste nas limitações do *software*, visto que detecta apenas as formas registradas no glossário. A seguinte tabela traz as unidades que ocorrem, suas variantes, e o padrão relacionado:

Quadro 10 - Variantes polilexicais

Unidade do glossário	Variante	Subtipo	Padrão da variante
disk usage / uso do disco	consumo de disco	Alteração na base	[N+SP (prep + art + N)]
instant messenger / <i>software</i> de mensagens instantâneas	instant messenger program / programa de mensagens instantâneas	Alteração na base	[N+SP (prep + art + N)]
download / fazer o download	obter por download	Alteração na base	Outras alterações na base
quarantine / colocar em quarentena	mover para a quarentena	Alteração na base	Outras alterações na base
Self-Service Options / Opções de autoatendimento	opções de autosserviço	Alteração na extensão	[N+SP (prep + art + N)]

Fonte: Elaborado pelo autor.

As variantes encontradas aqui também apresentam um padrão muito parecido com as monolexicais. O caso de *disk usage*, por exemplo, assemelha-se aos de *location*, *override* e *report* acima:

Se você executar um programa importante que apresente alto **consumo** de CPU, memória ou disco, será possível adicioná-lo à lista Exclussões de programas.

Você pode adicionar os programas que apresentam alto **uso** de CPU, memória e disco à lista Exclussões de programas.

Ou seja, a situação poderia se explicar por uma causa dialetal ou cronológica, mais do que alguma questão discursiva de qualquer tipo.

Já *instant messenger*, além de estar envolvida em elipses como foi visto na análise tradutória, também apresenta uma variante. Talvez, nesse caso, a variante seja resultado da tradução usada no glossário para o termo. No glossário, o termo é apresentado como o sintagma nominal consolidado *instant messenger / software de mensagens instantâneas*. Contudo, no *corpus*, o termo nunca aparece em isolamento. Em todos os casos, ele faz parte da extensão de outro sintagma nominal. Logo, parece que a variante *instant messenger program* encontrada

no *corpus* é, na verdade, o único termo que se referiria ao *software* de mensagens instantâneas. Claro que, considerando a área de computação como um todo, *instant messenger* e *instant messenger software/program* são variantes. Seja como for, esse termo parece apresentar bastante volatilidade nas traduções, embora esteja bastante difundido no meio (vide MSN Messenger, Live Messenger, Skype, etc.).

Há ainda o caso igualmente numeroso da unidade *download / fazer o download*. Como se viu ao longo dessa análise, as UFE têm apresentado muito potencial para causar ruído, das mais variadas maneiras. Este caso foi discutido como uma ocorrência de estabilidade semântica na análise fraseológica (cf. *Estabilidade semântica* na página 168). A questão de *quarantine*, embora também envolva uma UFE, é mais semelhante aos casos observados acima, se enquadrando na situação de *disk usage*. Ou seja, trata-se do uso de uma variante lexical (*mover* no lugar de *colocar*) que não apresenta uma explicação concreta, mas que também não representa nenhuma queda de qualidade no material localizado. Por fim, há um único caso de alteração na extensão, envolvendo *Self-Service Options*. Essa mudança também está alinhada à variante de *quarantine*: *self-service* foi traduzido por *autosserviço* em vez de *autoatendimento*.

Por fim, há o caso das diversas alterações complexas. Essa categoria também apresenta um número razoável de ocorrências, igualmente em função de apenas duas unidades do glossário ambas peculiarmente relacionadas: *incoming / de entrada* e *outgoing / de saída*. A seguinte tabela mostra os casos específicos:

Quadro 11 - Variantes que envolvem diversas alterações complexas

Unidade do glossário	Variante	Tipo
rescan / reverificação	repetir verificação; nova verificação; verificar novamente; executar nova verificação	Com semelhança formal
incoming / de entrada (*)	Recebido	Sem semelhança formal
outgoing / de saída (*)	Enviado	Sem semelhança formal
support / suportar	ter suporte para	Com semelhança formal
viral / viral	de vírus	Com semelhança formal
idle time / tempo ocioso	tempo de ociosidade	Com semelhança formal
Symantec Store / Loja online da Symantec	Symantec Store	Sem semelhança formal

Fonte: Elaborado pelo autor.

O primeiro dos casos apresenta uma quantidade considerável de variantes. De fato, algumas ocorrências para o termo estão em sua forma verbal, demandando uma expressão diferente de *reverificação*. Mas também há o uso de uma unidade nominal distinta, *nova verificação*. Na verdade, a unidade glossariada é minoria no *corpus*, aparecendo apenas uma vez, em sua forma verbal, na UI.

Depois, os casos de *incoming* e *outgoing* são relacionáveis e, conforme dito acima, representam a maioria das ocorrências. Um exemplo:

You can configure Email Antivirus Scan to scan all **incoming** email messages and **outgoing** email messages.

É possível configurar a Verificação antivírus de e-mail para verificar todas as mensagens de e-mail **recebidas** e **enviadas**.

Todas estão relacionadas ao termo *e-mail* ou *mensagem*. Não há ocorrências para *outgoing / de saída*, mas no caso de *incoming / de entrada*, todas as ocorrências têm relação com tráfego de rede, como no seguinte exemplo:

AutoBlock blocks all *incoming* traffic from the attacking computer for a limited time, even if the incoming traffic is not a recognized attack method.

O Bloqueio automático interrompe automaticamente toda a comunicação *de entrada* do computador invasor por um determinado período, mesmo que essa comunicação não seja um método de ataque conhecido.

Talvez por uma questão de adequação ou de uso, exista essa variante para cada caso, a qual não ocorre no original. Não seria de se descartar que, nesse caso, a forma *incoming* se refira a duas unidades terminológicas adjetivas diferentes.

Por fim, a unidade *Symantec Store* foi tratada como um erro tradutório (cf. *Referência extralinguística malsolucionada* na página 189). Ainda assim, não é possível afirmar que o termo apresentado no glossário não faça parte de uma mudança terminológica recente, representando uma variação de causa cronológica.

5.7 Análise linguística

A análise linguística tratada nesta seção agrega todos os aspectos que são normalmente ignorados (ou melhor, tomados por pressuposto) em Estudos de Terminologia e Tradução. Compreende aspectos discursivos, como o uso de elipses, ou morfológicos, como a flexão verbal, os quais são abordados de maneira mais explícita em trabalhos de PLN e linguística computacional, visto que estes precisam lidar diretamente com aquilo que acontece na superfície de um texto.

Conforme a análise apresentada aqui demonstra, em números absolutos, essa categorização é a mais significativa de todas para a geração de ruído em unidades terminológicas ou fraseológicas. Vale lembrar, no entanto, que essas ocorrências também estão associadas a UFE, variantes ou questões tradutórias.

As próximas seções apresentam a interpretação de alguns números obtidos para ruído linguístico na análise e descrevem em detalhes os casos encontrados nos segmentos de tradução analisados.

5.7.1 Análise linguística em números

Esta seção traz alguns cruzamentos dos dados obtidos na análise de ruído linguístico, falando sobre os positivos e falsos positivos encontrados.

A seguinte tabela sumariza os resultados encontrados para as ocorrências linguísticas, segmentados para cada *subcorpora*. Além do total de ocorrências, traz os falsos positivos encontrados na lista de QA.

Tabela 42 - Resultados para ocorrências linguísticas por *subcorpus*

<i>Subcorpus</i>	Falso positivo	
	#	%
Ajuda	1141	94,30%
UI	69	5,70%
Total Geral	1210	100,00%

Fonte: Elaborado pelo autor.

Esta tabela apresenta as seguintes informações:

- ***Subcorpus***: indica se a ocorrência foi encontrada na ajuda ou na UI;
- **Falso positivo**: apresenta as ocorrências que o tradutor descarta ao revisar a lista de QA.

As ocorrências listadas compreendem exclusivamente aspectos classificados como linguísticos. Vale dizer, contudo, que todas as unidades glossariadas que apresentaram ruído no *corpus* em alguma categoria (fraseológica, etc.) envolvem pelo menos uma ocorrência linguística também.

A seção a seguir discute cada um dos tipos de ocorrência linguística catalogados, trazendo exemplos e apontando para as possíveis causas.

5.7.2 Ocorrências linguísticas

As seguintes subseções apresentam em detalhes os casos encontrados para as ocorrências linguísticas. As ocorrências aqui presentes estão mais alinhadas com os aspectos trabalhados anteriormente nos itens sobre as categorias *Análise fraseológica* (p. 160), *Análise tra-*

*duto*ria (p. 177) e *Análise variacional* (p. 196) da análise. A tabela a seguir apresenta em detalhes as ocorrências encontradas:

Tabela 43 - Classificação para as ocorrências consideradas erro pelo *software* de QA

Subtipo de falso positivo	UI		Ajuda		Total	
	#	%	#	%	#	%
Anáfora		0,00%	1	0,07%	1	0,07%
Elipse lexical anafórica	1	0,07%	45	3,01%	46	3,08%
Flexão Verbal	15	1,00%	89	5,96%	104	6,96%
Gênero	2	0,13%	44	2,95%	46	3,08%
Inserção	2	0,13%	18	1,20%	20	1,34%
MT	42	2,81%	242	16,20%	284	19,01%
Não traduzível	11	0,74%	1	0,07%	12	0,80%
Número	12	0,80%	284	19,01%	296	19,81%
Outra unidade	26	1,74%	421	28,18%	447	29,92%
Referência a UI		0,00%	238	15,93%	238	15,93%
Total Geral	111	7,43%	1383	92,57%	1494	100,00%

Fonte: Elaborado pelo autor.

Esta tabela apresenta as seguintes informações:

- **UI e Ajuda:** indicam o *subcorpus* em que as ocorrências aparecem;
- **Subtipo de falso positivo:** apresenta as ocorrências que representam ruído, separadas para cada subtipo catalogado.

As seções a seguir descrevem cada um dos subtipos na tabela.

5.7.2.1 Anáfora

Há um caso individual no *corpus* inteiro para anáfora, encontrado neste segmento:

Entrada do glossário	risk / risco
Segmento original	Definitions in the Cloud provide a generic name for the <i>risk</i> detected, but the local definitions provide the specific name for the <i>risk</i> detected.
Segmento traduzido	As definições da Nuvem fornecem um nome genérico para os <i>riscos</i> detectados, enquanto as definições locais fornecem nomes específicos para <i>eles</i> .

Consiste, portanto, no uso de um pronome para a segunda ocorrência de *risk*, a qual de fato consiste no sintagma *risk detected* em ambos os casos. Como foi dito, é um caso único. Sendo assim, não é possível tirar qualquer conclusão a respeito do uso de pronomes em substituição a alguma unidade glossariada, menos ainda especular se isso ocorreria também entre segmentos distintos, como no exemplo a seguir (a área sombreada destaca o local em que o *software* segmenta os pares de tradução):

Entrada do glossário	Security History / Histórico de segurança
Segmento original	<i>Security History</i> lets you monitor the security tasks that your product performs in the background. In addition, the alerts that you receive can be reviewed at any time in <i>Security History</i> .
Segmento traduzido	O <i>Histórico de segurança</i> permite o monitoramento das tarefas de segurança que o produto executa em segundo plano. <i>Ele</i> permite também que os alertas recebidos possam ser analisados a qualquer momento.

Na segunda ocorrência, no segundo segmento, o tradutor optou por usar um pronome, aproveitando a proximidade com o referente. Esse exemplo foi retirado do *corpus*, mas não faz parte desta análise por ter ficado de fora do recorte que foi feito. Ele demonstra que nem sempre os princípios da internacionalização (cf. *GILT e outros termos comuns à área* na página 24) são seguidos nas traduções; também mostra que estudos nesse sentido poderiam trazer resultados mais palpáveis.

5.7.2.2 Elipse lexical anafórica

Os casos de elipse lexical representam 3,08% dos falsos positivos nas marcações de erro linguístico. Todos os casos de elipse envolvem sintagmas e, em sua maioria, ocorrem tanto no original quanto na tradução, embora isso não seja sistemático. Seguem alguns exemplos:

Entrada do glossário	risk / risco
Segmento original	The View Details link provides additional details about the risk such as the <i>risk</i> impact and the origin of the <i>risk</i> .
Segmento traduzido	O link Exibir detalhes oferece detalhes adicionais sobre os <i>riscos</i> , como o impacto e a origem dos <i>riscos</i> .

Aqui, a tradução agrupa *risk impact* e *origin of the risk* sob o mesmo sintagma nominal, *o impacto e a origem dos riscos*.

Entrada do glossário	vírus / vírus
Segmento original	About <i>virus</i> and security risk protection features
Segmento traduzido	Sobre os recursos da proteção <i>antivírus</i> e contra riscos à segurança

Neste caso, a elipse que causa ruído está no original. Segundo o glossário, a unidade *virus protection* é traduzida por *proteção antivírus*. A tradução do glossário foi aplicada, mas o *software* considerou apenas a ocorrência de *virus* individualmente.

5.7.2.3 Flexão Verbal

A flexão verbal representa uma porcentagem razoável do total de falsos positivos em marcações de erro (6,96%), sendo que o volume está distribuído igualmente entre verbos e sintagmas verbais. Seguem alguns exemplos de traduções corretas que foram marcadas como erros pelo *software* em função de verbos flexionados:

Entrada do glossário	copy / copiar scan / verificar
Segmento original	In this case, you might want to create a folder into which you <i>copy</i> and <i>scan</i> those files before using them.
Segmento traduzido	Neste caso, você poderá criar uma pasta na qual <i>copiará</i> e <i>verificará</i> esses arquivos antes de utilizá-los.
Entrada do glossário	quarantine / colocar em quarentena
Segmento original	If a file has Poor or Bad trust level, Norton AntiVirus lets you <i>quarantine</i> the file.
Segmento traduzido	Se o arquivo apresentar nível de confiança Baixo ou Ruim, o Norton AntiVirus permitirá que você o <i>coloque em quarentena</i> .
Entrada do glossário	sign in / acessar
Segmento original	<i>Sign in</i> to Norton Online Backup and check your online backup and storage details.
Segmento traduzido	<i>Acesse</i> o Norton Online Backup e verifique seu backup on-line e os detalhes de armazenamento.

Em números absolutos, trata-se de 104 ocorrências, sendo 85,58% destas na ajuda. Como foi observado em *Fraseologia em números* (p. 161), é possível que isso tenha relação com um número mais limitado de sujeitos na UI.

5.7.2.4 Gênero

As ocorrências de gênero representam uma proporção pequena do total (3,08%). Todas estão atreladas a unidades adjetivas (individuais ou sintagmáticas). Das nove unidades adjetivas que constam no glossário, quatro têm potencial para introduzir variação de gênero com base na tradução registrada no glossário, a saber:

- fragmented / fragmentado
- hidden / oculto
- low / baixo
- signature-based / baseado em assinaturas

Destas, apenas *hidden* / oculto não apresenta ruído. Há apenas quatro ocorrências para *hidden* no *corpus* inteiro, todas no masculino. Além disso, duas outras unidades também trazem casos de concordância de gênero, visto que a tradução usada é uma variante do glossário:

- *incoming* / de entrada (*)
- *outgoing* / de saída (*)

As traduções que ocorrem são, respectivamente, *recebido* e *enviado*. Os segmentos a seguir exemplificam os casos descritos para gênero:

Entrada do glossário	signature-based / baseado em assinatura
Segmento original	SONAR is quicker than the traditional <i>signature-based</i> threat detection techniques.
Segmento traduzido	O SONAR é mais rápido que as tradicionais técnicas de detecção de ameaças <i>baseadas em assinaturas</i> .
Entrada do glossário	low / baixo
Segmento original	Shows the files that have very <i>low</i> user prevalence.
Segmento traduzido	Mostra os arquivos que têm prevalência de usuários muito <i>baixa</i> .

O primeiro caso, além do gênero, também apresenta uma diferença de número em relação à unidade registrada no glossário.

5.7.2.5 Inserção

Há uns poucos casos de inserção, como aqueles apresentados na análise fraseológica (cf. *Sumário das ocorrências em relação às características das UFE* na página 165), mas, ao contrário daqueles, os quais estavam fundamentados sobre a proposta fraseológica de Bevilacqua, as ocorrências aqui presentes nada têm a ver com unidades derivadas de verbo. Um exemplo:

Entrada do glossário	limited-time subscription / assinatura por tempo limitado
Segmento original	Your Symantec product includes a complimentary, <i>limited-time subscription</i> for definition update services.
Segmento traduzido	O produto Symantec inclui uma <i>assinatura</i> complementar <i>por tempo limitado</i> para os serviços de atualização de definições.

Talvez seja possível explicar a ocorrência desses casos por outros motivos, como, por exemplo, o registro inadequado de unidades no glossário. Conforme foi visto na página 159 (seção *Glossário* desta análise), na localização é comum que os glossários sejam feitos para suprir situações específicas e até mesmo pontuais de trabalho. Nesse sentido, não se descarta que uma unidade tenha sido registrada para solucionar um problema de tradução específico e, sendo assim, a definição dos equivalentes não teria sido muito estudada. Outra possibilidade é que o equivalente encontrado ou estudado não seja tão consolidado em termos formais quanto o original. O seguinte caso traz indícios dessa segunda hipótese:

Entrada do glossário	disk space / espaço em disco
Segmento original	It also provides you free <i>disk space</i> of about 20-30MB.
Segmento traduzido	Ela também fornece <i>espaço</i> livre <i>em disco</i> de aproximadamente 20 a 30 MB.

A unidade *disk space* em inglês, dentro do *corpus*, não apresenta variações nem inserções (como, por exemplo, *space in disk* ou *disk unused space*). Outros casos parecidos:

Entrada do glossário	Web page / página da Web
Segmento original	In addition, you can select a threat name to open the Symantec Security Response <i>Web page</i> where you can learn more about the threat.
Segmento traduzido	Além disso, é possível selecionar um nome de ameaça a ser aberta na <i>página</i> do Symantec Security Response <i>na Web</i> , na qual é possível saber mais sobre a ameaça.

Entrada do glossário	risk alert / alerta de riscos
Segmento original	High <i>Risk Alert</i>
Segmento traduzido	<i>Alerta de alto risco</i>

Esse caso, inicialmente imprevisto, demonstra a necessidade de se investigar mais a fundo a questão da estabilidade sintática em traduções. Um dos aspectos relacionados seria a assimetria dessa estabilidade entre o original e sua tradução. Em última análise, esse tipo de característica poderia decorrer tanto de diferenças linguísticas entre os idiomas para determinado léxico quanto da volatilidade da área (como, nesse caso, a computação ou informática).

5.7.2.6 Modelo transformacional

O MT representa um volume considerável das ocorrências totais para falsos positivos linguísticos (19,01%). Como foi afirmado acima, o caso foi tratado na seção *Modelo transformacional* (p. 169) do capítulo sobre os resultados fraseológicos. Aqui são analisadas especificamente as unidades que não são contempladas na análise fraseológica, como as unidades simples e alguns casos específicos para unidades sintagmáticas deverbais.

Entrada do glossário	quarantine / colocar em quarentena
Segmento original	Please remove all user data including my stored passwords & <i>quarantine</i> contents.
Segmento traduzido	Remover todos os dados do usuário, inclusive minhas senhas armazenadas e conteúdo <i>em quarentena</i> .

Nesse caso, embora o sintagma nominal do inglês apresente a unidade nominal de *quarantine*, a tradução traz a equivalente adjetiva *em quarentena*. Se o *software* pudesse agrupar todas as formas superficiais dessa unidade, esse caso poderia ter sido omitido da lista do QA. Outro exemplo semelhante, desta vez sobre uma unidade simples:

Entrada do glossário	report / relatar
Segmento original	You can also use the Cancel option if you do not want to send the <i>report</i> to Symantec.
Segmento traduzido	Além disso, você poderá clicar na opção Cancelar se não quiser enviar o <i>relatório</i> à Symantec.

A tradução está correta, e o glossário também inclui a unidade nominal *report / relatório*. Logo, seria desnecessário processar novamente o segmento com a forma *report / relatar*. Se ambas estivessem agrupadas de alguma maneira, isso reduziria ainda mais a lista do QA. Os demais casos são todos semelhantes a estes, e na grande maioria das vezes, pelo menos duas formas superficiais são registradas para a mesma unidade, tal qual nos casos acima.

5.7.2.7 Não traduzível

Esse subtipo tem raiz em um fato comum a projetos de localização. Às vezes, alguns segmentos são colocados à tradução, quando na verdade não deveriam ser traduzidos. Em geral, os tradutores conseguem identificá-los ou recebem instruções para não traduzir. O *software* de QA, alheio aos arranjos processuais que as pessoas fazem acerca do material traduzido, os processam da mesma maneira. Alguns exemplos:

Entrada do glossário	support / suportar
Segmento original	https://www-secure.symantec.com/home_homeoffice/support/special/embedded_support/index.jsp
Segmento traduzido	https://www-secure.symantec.com/home_homeoffice/support/special/embedded_support/index.jsp
Entrada do glossário	instant messenger / software de mensagens instantâneas
Segmento original	AOL <i>Instant Messenger</i> -AIM
Segmento traduzido	AOL <i>Instant Messenger</i> -AIM

Embora a proposta de uma solução para o problema esteja fora do escopo deste trabalho, esse aspecto especificamente poderia ser tratado com facilidade. As ferramentas de QA geralmente apresentam um recurso que identifica quando uma tradução contém exatamente o mesmo conteúdo que o original em um segmento. Sendo assim, o *software* poderia remover segmentos nessa condição da verificação terminológica, assim economizando mais algumas linhas da lista de QA. Em um QA, o foco que o tradutor dá para esse tipo de segmento é diferente; como não há nada traduzido, normalmente a análise está em julgar se esse é o caso ou não, e não se algum termo foi usado conforme o glossário etc. Alguém poderia argumentar que, ainda assim, a ferramenta seria útil, já que um tradutor poderia ter deixado um segmento parcialmente traduzido ou tenha inserido um ou uns poucos caracteres inadvertidamente em um segmento não traduzido, o qual deveria ter sido traduzido. No entanto, para esses casos, alguns produtos de *software* de QA conseguem identificar segmentos parcialmente não traduzidos.

5.7.2.8 Número

Outro subtipo que apresenta um volume considerável de ocorrências (19,81% do total), a marcação de número está distribuída em dois grupos principais: a concordância, em adjetivos e verbos no particípio, e o uso de nomes simples e compostos no plural, tanto em casos em que não seria relevante manter a unidade glossariadas no singular quanto em situações na qual a unidade glossariada faz parte da extensão de um sintagma nominal. É importante destacar que a grande maioria das ocorrências de número consiste em nomes simples do glossário que fazem parte de um sintagma nominal, correspondendo a cerca de 13,5% dos 19,81% totais.

Com relação às ocorrências de concordância de número, seguem alguns exemplos:

Adjetivo:

Entrada do glossário	hidden / oculto
Segmento original	Always keep this feature turned on to prevent any <i>hidden</i> programs that affect the operating system of your computer.
Segmento traduzido	Mantenha sempre este recurso ativado para impedir que programas <i>ocultos</i> afetem o sistema operacional do computador.

Sintagma adjetival:

Entrada do glossário	signature-based / baseado em assinaturas
Segmento original	SONAR Protection is quicker than the traditional <i>signature-based</i> threat detection techniques.
Segmento traduzido	A Proteção do SONAR é mais rápida que as tradicionais técnicas de detecção de ameaças <i>baseadas em assinaturas</i> .

Sintagma verbal:

Entrada do glossário	download / fazer o download
Segmento original	Download Intelligence protects your computer against any unsafe file that you <i>download</i> from the supported portals
Segmento traduzido	A opção Download inteligente protege seu computador contra arquivos não seguros <i>obtidos por download</i> de portais

Quanto a este último caso, trata-se do único sintagma verbal identificado, e todas as ocorrências consistem em uma variante (obter por download), conforme visto em *Alteração complexa* (p. 209) na análise de variação.

Sobre os casos em nomes simples e compostos, trata-se de três situações específicas:

- o termo é usado apenas no singular no original, mas pode ser usado das duas maneiras na tradução;
- é indiferente manter ou não o nome no singular;
- a unidade faz parte da extensão de um sintagma.

Para os dois primeiros casos, das entradas do glossário envolvidas, todas são nominais, e a maior parte consiste em unidades simples. As entradas do glossário que mais se envolveram nessa situação foram:

- damage / dano
- risk / risco
- threat / ameaça

No primeiro caso, a palavra *damage* em inglês, no sentido de ferir, causar dano, não é usada no plural. No *corpus*, ela nunca aparece no plural no original. Na tradução, por sua vez, ela nunca aparece no singular. Talvez, nesse caso específico, o termo devesse ser registrado no plural em português no glossário, mas seria necessária uma análise mais aprofundada para chegar a essa conclusão em um âmbito maior (como no da terminologia oficial da empresa ou da área de segurança de dados digitais). O exemplo a seguir sugere uma decisão tradutória de manter estruturas paralelas:

Indicates that your computer and activities are protected from *threats, risks, and damage*.

Indica que o seu computador e as atividades estão protegidos contra *ameaças, riscos e danos*.

No entanto, a hipótese não se sustenta em todos os casos:

Most viruses just replicate; many also do *damage*.

A maioria dos vírus não apenas se replica, mas também causa *danos*.

A program with the potential to cause *damage* to a computer by destruction, disclosure, modification of data, or denial of service.

Programa com potencial para causar *danos* a um computador por meio da destruição, divulgação, modificação de dados ou negação de serviço. à segurança.

De qualquer forma, não há dados suficientes para tirar uma conclusão nesse sentido. Já as outras duas unidades de maior frequência, *risk / risco* e *threat / ameaça*, não apresentam tal particularidade no inglês e, além disso, são usadas no original e na tradução tanto no singular quanto no plural. Para esses casos, a decisão de optar pelo plural parece estar associada a um referencial não bem definido no original:

An active status of your Norton AntiVirus is necessary for protecting your computer against *any* security *threat* or *risk*.

Para que seu computador seja protegido contra quaisquer *ameaças* ou *riscos* à segurança, é necessário que o Norton AntiVirus tenha o status ativo.

Not all actions are available for *every type of* email *threat*.

Nem todas as ações estão disponíveis para todos os tipos de *ameaças* de e-mail.

Nos exemplos apresentados, o original apresenta uma formulação que já está generalizada e que permite a reformulação de várias maneiras: *qualquer*, *quaisquer*, *todos os*. Ainda há casos em que a relação entre a unidade de tradução e outro referente possibilitaria o uso de singular e de plural na tradução:

Entrada do glossário	threat / ameaça
Segmento original	In the case of unsafe files, Norton AntiVirus identifies them as <i>threat</i> and removes them.
Segmento traduzido	No caso de arquivos não seguros, serão identificados como <i>ameaças</i> e removidos pelo Norton AntiVirus.

Aqui, o referente que é identificado como ameaça está no plural, permitindo a formulação de duas maneiras: *serão identificados como ameaças* ou *como (uma) ameaça*.

Definitions in the Cloud provide a generic name for the *risk* detected, but the local definitions provide the specific name for the risk detected.

As definições da Nuvem fornecem um nome genérico para os *riscos* detectados, enquanto as definições locais fornecem nomes específicos para eles.

As demais entradas envolvidas são:

- boot sector / setor de inicialização
- copy / cópia
- device / dispositivo
- download / download (*)
- infection / infecção
- internet download / download da Internet
- notification / notificação
- reputation level / nível de reputação
- Symantec product / produto da Symantec
- task / tarefa
- Trojan horse / Cavalo de Troia

Os casos encontrados se repetem. Basicamente, trata-se de segmentos em que não faz diferença no ponto de vista da significação manter a unidade de tradução no singular ou colocá-la no plural, como nos seguintes exemplos:

Entrada do glossário	boot sector / setor de inicialização
Segmento original	A program that infects another program, <i>boot sector</i> , partition sector, or document by inserting itself or attaching itself to that medium.
Segmento traduzido	Programa que infecta outros programas, <i>setores de inicialização</i> , setores de partição ou documentos por meio de inserção ou anexação.

Entrada do glossário	copy / cópia
Segmento original	A worm can send itself or send a <i>copy</i> of itself in an email message without any interaction with you.
Segmento traduzido	Os worms podem enviar a si mesmos ou <i>cópias</i> suas em mensagens de e-mail sem intervenção humana.

Há ainda um caso em que a diferença se dá por um erro no original:

Entrada do glossário	device / dispositivo
Segmento original	You must use the same port number for each of the <i>device</i> that is connected to your Network Security Map.
Segmento traduzido	Você deve usar o mesmo número de porta em todos os <i>dispositivos</i> conectados ao Mapa de segurança da rede.

Aqui, o segmento do original apresenta um erro, pois a formulação correta gramaticalmente é *each of the devices*. Isso traz à tona um aspecto da localização. Como o texto não é um objeto estático, em uma futura atualização do *software* ou da ajuda, na qual o erro é corrigido do original, isso possivelmente não exigirá nenhuma modificação na tradução.

Já com relação aos casos em que a unidade glossariada compõe a extensão de um sintagma, segue um exemplo:

Entrada do glossário	device / dispositivo
Segmento original	This category includes Windows System files (.sys), dynamic link library files (.dll), and <i>device</i> driver files (.drv).
Segmento traduzido	Essa categoria inclui arquivos do sistema do Windows (.sys), arquivos DLL (.dll) e arquivos de driver de <i>dispositivos</i> (.drv).
Entrada do glossário	Norton security product / produto de segurança Norton
Segmento original	The collective efforts from all <i>Norton security product</i> users help in quick identification of solutions for these threats and risks.
Segmento traduzido	O esforço coletivo de todos os usuários de <i>produtos de segurança Norton</i> ajuda na rápida identificação de soluções para essas ameaças e riscos.

5.7.2.9 Outra unidade

Esta causa está presente em 27,27% do total de falsos positivos. Aqui, as ocorrências envolvem não só nomes próprios, mas também unidades comuns (terminológicas ou não), que partilham da mesma forma encontrada no glossário para dada unidade. Os nomes próprios compreendem a grande maioria das ocorrências (82,33% do total). Seguem alguns exemplos:

Entrada do glossário	low / baixo
Segmento original	You have <i>low</i> storage for online backup. Please add more storage.
Segmento traduzido	Você tem <i>pouco</i> armazenamento disponível para o backup on-line. Adicione mais armazenamento.
Entrada do glossário	Insight / Insight
Segmento original	This window provides <i>insight</i> on the file in an organized way.
Segmento traduzido	A janela proporciona <i>insights</i> sobre o arquivo, de maneira organizada.

Nesses casos, a forma das unidades corresponde diretamente, mas a tradução é diferente por se tratarem de unidades distintas. No primeiro caso, as duas traduções para *low* de certa forma têm relação uma com a outra por indicarem uma quantidade inferior de determi-

nada medida (tamanho de disco, armazenamento). Um caso em que foi usado *baixo*, conforme o glossário, é o seguinte:

This category mostly includes the *low*-level risks that require your attention for a suitable action.

Essa categoria inclui principalmente os riscos de *baixo* nível que exigem sua atenção para uma ação adequada.

É possível observar que, por particularidades das respectivas traduções, *armazenamento* e *nível* requerem referentes distintos. Por questões de adequação, *baixo armazenamento* ou *pouco nível* não resultam em traduções corretas. Já no segundo caso, o *Insight* do glossário se refere a um nome de produto (nome próprio), mas na tradução, é um nome comum, traduzível por *informações* ou *insight*, conforme o segmento traduzido apresenta.

Há também casos em que o *software* detecta parte de outra unidade, a qual nada tem a ver com a entrada do glossário em questão. Exemplos:

Entrada do glossário	antivirus scan / verificação antivírus
Segmento original	A Norton <i>AntiVirus scan</i> can identify threats on your computer in the following ways:
Segmento traduzido	Uma verificação do Norton AntiVirus pode identificar ameaças em seu computador das seguintes maneiras:
Entrada do glossário	antivirus protection / proteção antivírus
Segmento original	It does not delay Norton <i>AntiVirus protection</i> .
Segmento traduzido	Isso não adiará a proteção do Norton AntiVirus.
Entrada do glossário	download / download (*)
Segmento original	If you do not want <i>Download</i> Insight to prompt you to select a suitable action for any type of file, move the Download Insight Full Report switch to the Never position.
Segmento traduzido	Se você desejar que o <i>Download</i> Insight solicite a você uma ação adequada para qualquer tipo de arquivo, mova a chave do Relatório completo do Download Insight para a posição Nunca.

Nos três casos acima, o *software* detecta partes de dois nomes próprios (Norton AntiVirus e Download Insight), registrando as ocorrências como possíveis erros de consistência terminológica. Como foi afirmado antes, não é possível precisar que tais ocorrências ainda

carreguem o mesmo referente que a unidade do glossário aponta (respectivamente, antivírus e download). O certo é que os nomes próprios não são tão variáveis e deveriam ser tratados de maneira diferente pelo *software*, conforme sugerem Sag et al (2002).

Uma grande parcela dessas ocorrências se deve a uma única entrada do glossário: *download*. Na verdade, a duas, a considerar que esta está registrada duas vezes, em sua forma nominal e verbal. Com relação à nominal, o motivo para as ocorrências de ruído está em um problema em como a forma traduzida foi registrada no glossário:

- download (*)

Seja qual for o motivo para terem incluído o asterisco entre parênteses na entrada, o fato é que esses caracteres extras fazem com que todas as ocorrências de *download* em sua forma nominal sejam incluídas na lista de QA. Além disso, a forma verbal do termo (*fazer o download*) faz com que essas ocorrências sejam dobradas, já que todos os segmentos, como os indicados no exemplo acima, são processados outra vez.

Claro que poderia se afirmar que esses casos são específicos, uma vez que o *corpus* agrupa informações sobre produtos cujos nomes trazem termos usados na área ou subárea de atuação da empresa (antivírus, download, entre outros). Contudo, há indícios suficientes que levam a crer que este aspecto não possa ser ignorado, como, por exemplo, a presença de nomes de produtos de outras empresas (no caso, o Internet Explorer, da Microsoft, que foi encontrado mais acima).

5.7.2.10 Referência a UI

Estas ocorrências foram identificadas por acaso, decorrendo de outros subtipos (como *Número e Flexão Verbal*, por exemplo), demonstrando a necessidade de manter algum registro ou algum tipo de marcação que indique o que é referência à UI especialmente na ajuda. Além de ajudar no processamento do QA, isso também facilitaria a manutenção da ajuda, visto que mudanças na interface poderiam ser rastreadas mais facilmente se existisse algum dispositivo que as identificasse nos segmentos de tradução. Seguem exemplos:

Entrada do glossário	Poor Trust / Baixa confiança
Segmento original	For example, if you want to view all files that have <i>Poor</i> trust level, click on the relevant section in the Trust Level graph.
Segmento traduzido	Por exemplo, se desejar ver todos os arquivos que tenham o nível de confiança <i>Baixo</i> , clique na seção relevante no gráfico Nível de confiança.

Neste caso, percebe-se que a detecção do *software* intersecciona uma referência a UI, em que apenas *Poor* se refere à UI, ao passo que *trust level* compõe outro elemento textual.

Entrada do glossário	report / relatar
Segmento original	If you do not want Norton AntiVirus to display the Monthly <i>Report</i> automatically, you can select the Do not display monthly reports automatically option in the Norton Monthly <i>Report</i> window.
Segmento traduzido	Caso não queira que o Norton AntiVirus exiba o <i>Relatório</i> mensal automaticamente, selecione a opção Não exibir relatórios mensais automaticamente, na janela <i>Relatório</i> mensal do Norton.

Nesse caso, o que está em jogo é tratar as referências à UI de maneira diferente. O exemplo a seguir justifica por que estes devem receber tratamento diferente:

Entrada do glossário	clear / limpar
Segmento original	You can use the <i>Clear</i> All User Trust option next to the drop-down list to remove the user trust.
Segmento traduzido	Você pode usar a opção <i>Remover</i> a classificação de confiança do usuário ao lado da lista suspensa para remover a confiança do usuário.

Aqui, poderia se argumentar que *remover* é uma variante de *limpar* na tradução de *clear*, e isso pode ser o caso de fato. No entanto, pressupondo que:

- a tradução na UI esteja correta (*Clear All User Trust / Remover a classificação de confiança do usuário*);
- o uso de *limpar* tenha sido imposto prescritivamente.

Ainda assim, nem sempre o tradutor terá a opção de modificar a tradução na UI e, nesse caso, tampouco poderá modificá-la na ajuda, sob o risco de dificultar o entendimento do usuário, o qual, ao buscar a opção *Limpar a classificação de confiança do usuário*, não encontrará o texto na UI. Isso não significa que o usuário não encontrará a opção para suprir sua necessidade de uso do *software*, mas passará um trabalho, pequeno ou grande, dependendo do

usuário, mas certamente desnecessário. Esse seria um dos motivos para se tratar essas referências de maneira especial – e, nesse caso, sendo uma razão totalmente enraizada na prática da localização.

Essas constatações sobre aspectos linguísticos normalmente abordados em PLN e linguística computacional encerram a análise das ocorrências encontradas na lista de QA. Como se pôde observar em um âmbito geral, uma lista de QA pode conter muito ruído, e as classificações feitas neste trabalho contribuem para visualizar melhor os tipos de problema envolvidos no processamento computacional. Algumas das observações feitas neste capítulo como um todo serão retomadas nas conclusões e considerações finais que vem a seguir, incluindo classificações fraseológicas, variacionais, tradutórias e linguísticas, bem como as classificações novas para cobrir erros de glossário e de alinhamento do *corpus*.

5.7.2.11 Possíveis causas para ocorrências de ruído gerado por aspectos linguísticos

Durante a análise linguística, um aspecto foi identificado como possível causador para alguns dos tipos de ruído observados. Tal aspecto consiste no fato de a unidade glossariada fazer parte de outro sintagma e pode explicar o porquê da ocorrência de um determinado falso positivo, mas sem estar diretamente relacionado na superfície com a ocorrência em si.

Por exemplo, como se viu acima nos casos de número (p. 226), a maioria dos termos simples que ocorrem no plural fazem parte da extensão de outra unidade. Os seguintes exemplos demonstram isso:

Entrada do glossário	infection / infecção
Segmento original	Computer Scans lets you specify different types of scans to detect and prevent any <i>virus infection</i> on your computer.
Segmento traduzido	As verificações do computador permitem especificar tipos diferentes de verificações para detectar e impedir <i>infecções por vírus</i> em seu computador.
Entrada do glossário	content / conteúdo
Segmento original	Programs that facilitate delivery of <i>advertising content</i> to the user through their own window, or by utilizing another program's interface.
Segmento traduzido	Programas que facilitam a exibição de <i>anúncios</i> ao usuário em uma janela própria ou por meio da interface de outro programa.
Entrada do glossário	viral / viral
Segmento original	In case of <i>non-viral threats</i> , you can configure Norton AntiVirus to exclude the security risk from being detected in the future scans.
Segmento traduzido	Em caso de <i>ameaças não virais</i> , você pode configurar o Norton AntiVirus para excluir o risco à segurança de ser detectado em futuras verificações.

Como se vê, em nenhum dos casos, o fato de fazer parte de outra unidade é a causa direta do ruído. Nos dois primeiros, as ocorrências são tradutórias, tratando-se de uma generalização e uma elisão, respectivamente. O último caso, além de envolver uma unidade que está em oposição a *viral (non-viral)*, trata-se de um caso de número, já que o sintagma envolvido é *non-viral threats*, em que a base está no plural, demandando concordância de número da extensão.

Sendo assim, seria possível afirmar que um *parsing do corpus* ao processar o QA terminológico seria um passo importante para melhorar a qualidade das listas de QA. A partir da identificação de sintagmas, o *software* poderia passar prever melhor o comportamento de determinadas unidades (em especial para os casos em que a unidade glossariada faz parte da extensão do sintagma).

Esse aspecto também é extensível ao *Modelo transformacional* (visto acima na página 224), como se vê a seguir:

Entrada do glossário	support / suportar support / suporte
Segmento original	Following is an overview of our <i>support</i> offerings:
Segmento traduzido	A seguir, uma visão geral de nossas ofertas de <i>suporte</i> :

Neste caso, *support* está em sua forma nominal, mas o *software* processa as duas unidades; tanto a nominal quanto a verbal. Enquanto a tradução nominal é excluída da lista, a verbal acaba sendo incluída.

Entrada do glossário	download / download (*) download / fazer o download
Segmento original	For example, you want to <i>download</i> an unsafe file.
Segmento traduzido	Por exemplo, caso queira <i>fazer o download</i> de um arquivo não seguro.

Na via inversa, a unidade no segmento é verbal tanto no original quanto na tradução. Para a verbal, o *software* exclui da lista, mas para a nominal, inclui. Observa-se que, aqui, o erro é em função do registro inadequado da unidade (*download (*)*). Ainda assim, o que importa nos exemplos é demonstrar que o *software* processa ambas as unidades e, se estas fossem consideradas apenas uma, isso não ocorreria.

A situação não se limita às unidades que têm mais de uma estrutura superficial registrada no glossário. É o que demonstra este exemplo:

Entrada do glossário	boot / inicializar
Segmento original	A program that infects another program, <i>boot</i> sector, partition sector, or document by inserting itself or attaching itself to that medium.
Segmento traduzido	Programa que infecta outros programas, setores de <i>inicialização</i> , setores de partição ou documentos por meio de inserção ou anexação.

Aqui, apenas a forma verbal *inicializar* consta em glossário. Como se observa no segmento, *boot* está na forma nominal e foi traduzida também no equivalente nominal *inicialização*.

6 CONCLUSÕES

O objetivo geral da presente pesquisa foi o de prover insumos linguísticos que colaborem para a elaboração de um modelo linguisticamente motivado para o *software* de QA terminológico (no par de idiomas aqui estudado), contemplando aspectos linguísticos fundamentados em Terminologia e Tradução, com a meta final de reduzir a incidência de falsos positivos nas listas de QA que o tradutor revisa. Nesse quesito, este trabalho espera ter revelado alguns pontos que possam ser aproveitados. Os parágrafos a seguir trazem mais detalhes sobre as conclusões obtidas a partir da análise, relacionando os aspectos fraseológicos, variacionais, tradutórios e linguísticos estudados.

Antes de passar às conclusões tiradas da análise, contudo, vale mencionar a complexidade envolvida em desenvolver uma pesquisa nos moldes aqui presentes, em uma abordagem multidisciplinar e multidimensional que demandou esforço considerável, primeiro na coleta de informações e na elaboração da revisão teórica, passando pela organização dos materiais para análise e, finalmente, na análise em si, a qual exigiu equilíbrio entre o ato de observar fenômenos diferentes com parâmetros de análise diversos e a necessidade de apresentar uma classificação consistente para esses parâmetros. Em um trabalho que envolve tantos assuntos que ora se tangenciam, ora se entrecruzam, não é difícil de perder a linha fundamental que deveria permear todos os pontos de análise. Espera-se que este trabalho tenha cumprido esse objetivo sem grandes desvios e sem cair em contradições.

Ficou demonstrado que o *software* de QA gera um volume considerável de falsos positivos, o qual resulta em um trabalho oneroso e maçante para o tradutor. Claro está que revisar uma unidade específica de cada um desses segmentos não é tão trabalhoso quanto traduzir o mesmo volume. Mas, ainda assim, resulta em muito trabalho, e em boa medida, trabalho desnecessário. Já com relação às classificações específicas, se por um lado a análise demonstrou o impacto que o uso de técnicas de tradução, variantes e fraseologias têm na incidência de falsos positivos, por outro, deixou visível também a necessidade de incorporar aspectos linguísticos mais básicos, como a morfologia e o uso de elipses, que já são bastante trabalhados em PLN e linguística computacional e que normalmente ficam implícitos em trabalhos sobre Tradução ou Terminologia.

Outra finalidade proposta para este trabalho era descrever, em linhas gerais, o cenário da localização, sua história e suas práticas. Espera-se que tanto o panorama geral apresentado

em *Localização* (p. 22) quanto os questionamentos do paradigma da localização (p. 96), bem como seu entrelaçamento com terminologia (p. 34) e tradução (p. 102), possam ajudar a nortear o desenvolvimento de modelos linguísticos e aplicações computacionais para a área e a compreender o papel do tradutor no processo como um todo. No que diz respeito à prescritividade atribuída ao setor da localização em geral, o estudo demonstrou que ela de fato tem influência nas traduções, que procuram seguir à risca o glossário, mas está longe de ser sistemática a ponto de restringir os tradutores a optar por traduções literais ou estilisticamente inadequadas no texto.

Além dos insumos mencionados acima, outro objetivo da análise foi prover informações que ajudem na tomada de decisões para o desenvolvimento de melhorias no *software* de QA. Nesse sentido, as ocorrências mais frequentes, se solucionadas, trariam mais benefícios para o tradutor. Como se pode observar, a classificação linguística é a que apresenta a grande maioria dos casos (71,62%), seguida de variacional (10,26%), tradutória (3,50%) e, por fim, fraseológica (2,54%). A análise revelou ainda duas outras categorias de classificação, *Alinhamento* e *Glossário*. Para estas, observa-se que problemas no glossário representam um percentual significativo das ocorrências (9,25%). Finalmente, os problemas na segmentação do *corpus* representam 2,83% dos casos. Abaixo estão alguns pontos de destaque da análise.

Como foi dito acima, duas categorias de análise inicialmente não previstas tiveram de ser criadas em função do ruído gerado: *Alinhamento* e *Glossário*. Dessas duas, contudo, o caso do alinhamento não é relevante para o ruído estudado aqui, já que ele não representa o alinhamento real dos segmentos do *corpus*. Por alinhamento real entende-se aquele encontrado, hipoteticamente, nos arquivos que a empresa proprietária do conteúdo coletado no *corpus* manteria em algum formato de bilíngue. Logo, essa classificação indica uma falha na montagem do *corpus* aqui analisado. Com relação aos demais segmentos presentes no *corpus*, é possível afirmar com relativa segurança que eles correspondem de fato ao alinhamento feito nos arquivos bilíngues originais.

O caso da categoria *Glossário*, por sua vez, demonstra talvez que a metodologia aplicada pelo cliente deva ser revista ou aprimorada. Por exemplo, um passo nessa direção seria remover os asteriscos ou parênteses desnecessários no campo que indica a forma glossariada dos termos. A questão do MT, em unidades com mais de uma estrutura superficial que tiveram duas ou mais de suas formas de superfície registradas no glossário, também poderia ter sido classificada nesta categoria.

Com relação à análise fraseológica, conclui-se que três características são determinantes para causar ruído: a estabilidade sintática, a estabilidade semântica e o MT. As demais ajudaram a identificar que unidades são UFE, bem como quais teriam potencial para ser UFE. Um próximo passo seria trabalhar em maneiras de formalizar cada uma das características para uma implementação em PLN. O desafio, aqui, não envolveria apenas o trabalho de descrição formal, mas também processual. Por exemplo, se uma UFE precisa ter necessariamente uma UT, esta também deveria estar listada no glossário. No caso de *quarantine / colocar em quarentena*, a UT *quarentena* também está no glossário. Pelo menos a considerar as unidades glossariadas que foram estudadas no recorte, as unidades terminológicas presentes nas UFE também estão listadas individualmente no glossário. Ainda na esteira da questão UFE-UT, o problema de distinguir essas unidades também deve ser discutido e trabalhado futuramente, bem como sua representação e organização em um glossário.

Outro ponto importante sobre a análise fraseológica está no registro aparentemente fortuito das UFE no glossário. Foi possível apurar dois motivos para isso. Primeiro, parte das UFE encontradas decorrem de empréstimos, como em *fazer o download*. O uso no original dessas unidades como estrutura argumentativa de uma sentença resulta na necessidade de definir uma estrutura verbal em português. O outro motivo encontrado também está relacionado à reprodução da estrutura argumentativa do original. A unidade *quarantine* foi traduzida para *colocar em quarentena* porque provavelmente o verbo *quarentenar*⁸⁵ não seja usual na área de segurança de dados digitais.

O trabalho ainda lista os padrões de UFE encontrados, inclusive os elementos que se inserem nessas unidades. Longe de ser uma lista definitiva, o exercício serviu para mostrar um caminho possível para desenvolver uma ferramenta que processe as UFE. No lugar de uma abordagem exaustiva, isso seria feito de maneira progressiva. Por exemplo, o *software* de QA também poderia ir, ao longo do tempo, incorporando *input* do usuário no assunto. Por exemplo, durante o QA terminológico, o usuário pode indicar que determinada ocorrência é falso positivo, mas que tem elementos de inserção. O usuário então marcaria os elementos de inserção na tela, permitindo ao *software* aprender padrões novos. Posteriormente, o *software* poderia inclusive tentar aplicar regras de uma UFE a outras. Para aprender por padrões, essa sugestão implica que, o *software* já teria o *corpus* etiquetado ou analisado previamente.

A estabilidade semântica tem vínculo direto com variação. Contudo, desvendar se a UFE tem ou não relação direta com a possibilidade de variação ocorrer é tema para trabalhos

⁸⁵ Verbo registrado por grandes dicionários como Houaiss e Aurélio.

futuros. Uma vez mais, a variação pode decorrer do uso de empréstimos. Como o tradutor recorre a outros verbos para elaborar a estrutura argumentativa de uma frase, pode ser que o primeiro verbo escolhido não cubra todos os casos. Como foi mencionado nos resultados, a UFE *fazer o download* não traduz muito adequadamente o particípio *downloaded*. Em vez de montar frases mais complexas e difíceis de ler, os tradutores recorreram a *obter por download*, surgindo daí a variação na UFE.

Por fim, a análise mostrou que o MT influi nas ocorrências fraseológicas e, mais do que isso, também explica a técnica de transposição na tradução e se aplica a unidades simples. Revela-se, portanto, como outra característica básica que permeia as propostas usadas como base neste trabalho.

Em última análise, pôde-se ver que não existe relação de 1:1 entre a estrutura da unidade original e da tradução. O original *download* é uma unidade simples e, até onde se viu, não apresenta variação nos segmentos. Sua tradução (*fazer o download*) é UFE e apresenta variação (*obter por download*). Ou seja, dentro da unidade, cada uma das faces (original / tradução) deve ser tratada de maneira independente. Além disso, conforme ilustra a técnica de transposição, não existe relação de 1:1 entre a classe gramatical do original e da tradução. Elas são alteradas do original para a tradução, mas algo subjacente permanece e justifica a relação de equivalência entre original e tradução em um segmento.

Já com relação à análise tradutória, observa-se que o *corpus* apresenta um número pequeno de erros efetivos, o que demonstra o rigor aplicado à tradução do conteúdo.

Com relação à aplicação das técnicas de tradução e ao volume de ocorrências encontrado, é possível afirmar que a área não segue totalmente a premissa, apontada no paradigma da localização, de que a função do tradutor é servir de conversor de códigos linguísticos. Um ponto importante é que a análise trata exclusivamente de unidades glossariadas. Isso significa que, em geral, elas desempenhariam um papel importante dentro do *corpus*. Logo, se essas próprias unidades podem ser omitidas ou mescladas (como se viu em elisão), talvez uma análise que extrapole os limites dos termos (e das fraseologias e variantes) possa revelar que a aplicação das técnicas de tradução seja mais frequente do que se viu aqui, além de demonstrar que tradução literal não explica o que acontece de fato na localização.

Ainda sobre a aplicação das técnicas, foi possível observar que se usa muito transposição, elisão e generalização. Contudo, não se pode descartar totalmente as outras técnicas sem realizar análises mais abrangentes. O estudo ainda aponta para algumas possíveis causas para

o uso dessas técnicas, como particularidades de um idioma para o outro (*damage* sempre no singular em inglês versus *danos* geralmente no plural em português, por exemplo).

Na superfície, a análise tradutória demonstra que:

- unidades originais no singular podem ser traduzidas no plural;
- unidades originais em determinada classe gramatical podem ser traduzidas em outra classe gramatical.

Como ficou demonstrado, a variação existe. Todas as grandes categorias propostas por Freixa estão representadas. A considerar as categorias maiores (alterações simples e alterações complexas), constata-se que ambas são equiparáveis em termos de incidência de falsos positivos. Dentro das alterações simples, tanto as gráficas quanto as morfossintáticas representam um percentual significativo das ocorrências e têm relação com UFE em boa parte dos casos. Com relação às gráficas, é marcante o caso das grafias de *on-line*, que sugere o envolvimento de causas cronológicas. É comum na área a terminologia usada na tradução mudar em intervalos curtos de tempo. Geralmente, essa terminologia é definida de maneira diferente que a do original, a qual se mostra mais estável. Já sobre as alterações morfossintáticas, por exemplo, constata-se que o problema da presença/ausência de artigo requer estudos mais aprofundados. Os casos ilustrados na análise variacional sugerem que seria interessante pelo menos marcar esses casos de maneira especial, para que o tradutor possa distinguir entre as ocorrências em que o *software* não consegue detectar a unidade daquelas em que identifica um caso simples de variação morfossintática. É importante lembrar que essa variação pode ser relevante ou não com relação à equivalência tradutória, o que corrobora com a sugestão de criar uma classificação especial para esses casos no *software* de QA, em vez de, por exemplo, simplesmente omitir as ocorrências da lista.

Vale lembrar também que algumas das escolhas parecem desnecessárias quanto à ausência de artigo. Segmentos em contextos idênticos foram traduzidos de maneira diferente nesse quesito. Isso pode sugerir que tradutores diferentes fizeram os segmentos, ou o mesmo tradutor em dia ou situação diferente (e sem revisar o que fez), ou algum outro motivo – como um segmento já existente e bloqueado que o tradutor não pôde alterar ou um segmento omitido do tradutor e incluído diretamente no material final, sem que o tradutor tivesse acesso. Essas situações não podem ser tratadas como ruído. Ainda sobre alterações morfossintáticas, aconteceram casos com alteração na estrutura, envolvendo duas unidades específicas. Um motivo para isso pode estar em variação cronológica, bem como na falta de sincronia entre todos os segmentos de tradução, em que alguns são mais antigos, não revisados, e outros no-

vos foram recém-traduzidos usando alguma terminologia nova, que acaba de ser definida, ou que não existia na época em que os outros segmentos foram traduzidos. Não se descarta também uma falha durante o TEP + QA, em que os envolvidos não conseguiram identificar o fato. Vale dizer que, nos casos específicos, os erros ou inadequações, se esse for o caso, estão longe de representar algum risco ou serem graves a ponto de dificultar a compreensão.

Sobre as alterações complexas, estas também são responsáveis por uma quantidade significativa de ocorrências. Aqui, com relação a caminhos futuros, um problema a resolver é como tratar tais casos. Registrar todas as unidades no glossário? Tanto as reduções quanto as alterações lexicais? Vincular variantes? Estas são algumas questões que estudos futuros poderiam responder.

Dentro dessa classificação, a maior parte das ocorrências consiste em alterações lexicais e em diversas alterações complexas. Conforme afirmado acima sobre as alterações lexicais, algumas decorrem da necessidade de expressar a estrutura argumentativa do original, a qual não pode ser concretizada com o uso do verbo originalmente escolhido para acompanhar uma UFE. Quanto à redução, os casos envolvem poucas unidades e são menos frequentes. Além disso, todos envolvem redução na extensão; a base permanece sempre. Os números não têm expressividade suficiente para se tirar conclusões assertivas, contudo.

Além disso, foi possível listar algumas causas da variação que teriam relação com a localização:

- discursiva, de economia linguística, especialmente para UIs, em que são feitas alterações gráficas para respeitar o limite de tamanho das *strings*;
- dialetal e cronológica, em que ocorre alteração ortográfica, alteração lexical e morfológica;
- cognitiva / imprecisão conceitual, envolvendo alteração morfossintática;
- dialetal / variação social, que compreende alteração lexical e morfológica.

Dessas causas, destacam-se as cronológicas ou dialetais, uma vez que as metodologias e as práticas aplicadas em localização favorecem isso:

- muda-se a terminologia muitas vezes sem se considerar o impacto da alteração em conteúdos já localizados;
- os projetos podem ser estruturados de tal maneira que alguns segmentos estejam bloqueados para modificação ou sequer sejam passados para o tradutor. No primeiro caso, pode ser que o tradutor ou editor não aponte para a variante por: não revisar o segmento em questão; por se deparar com o suposto problema, mas ignorar, afinal, isso está

fora do escopo de seu trabalho; por revisar e avisar o cliente, que ignora o ocorrido por qualquer motivo. No segundo caso, o segmento seria preenchido só na finalização do projeto – ou ainda o tradutor recebe um pacote de tradução que inclui apenas os segmentos com conteúdo novo a traduzir;

- mais de um tradutor trabalha no projeto; dependendo do volume de palavras a traduzir, mais de um editor também. Duas cabeças revisoras pensam de maneiras diferentes. Problemas de tradução são resolvidos de maneiras diversas e as pessoas têm pontos de vista diferentes ao conferir a terminologia do glossário e contrastá-la com aquilo que foi traduzido.

Tudo isso sugere que, além de se preocupar com a parte tecnológica e computacional do processo de QA terminológico, também é necessário repensar a maneira com que se faz localização dentro das empresas, por parte de tradutores e terminólogos internos, e além delas também, por parte dos MLV e tradutores, editores, revisores textuais, terminólogos autônomos.

Enfim, ainda na esteira de apontar possíveis estudos e aspectos que podem e devem ser aprofundados com relação ao presente trabalho, são lançadas questões:

- é necessário incluir alteração ortográfica na lista ou não, em todos os casos? Casos como as variantes *online* e *on-line* sugerem que esse tipo de alteração deveria aparecer na lista de qualquer maneira;
- foram detectados casos de abreviação bem específicos e bastante comuns em UI (em função de limitação de caracteres). Teria algum jeito, nesse caso específico, de tratar o problema? Parece o tipo de caso que deveria vir à tona sempre; em última análise, o volume constatado não representa muito e não seria ônus para o tradutor revisar;
- também é preciso ter em mente que as variações vistas aqui podem ocorrer juntas, o que pode gerar certa complexidade ao tratamento dos casos.

Por fim, mas nem um pouco menos importante, os aspectos agrupados neste trabalho como linguísticos se revelaram os maiores causadores de ruído. Um possível motivo para isso é que os aspectos nessa categoria, como alguns morfológicos (concordância de gênero e número, flexão verbal), estão em uma camada de análise mais baixa ou básica que aqueles vistos em terminologia e tradução.

Entre os casos observados estão anáforas, elipses lexicais, flexão verbal, concordância de gênero, concordância de número, elementos não traduzíveis e unidades diferentes com a

mesma forma (unidades homógrafas). Os parágrafos a seguir trazem uma síntese dos resultados.

Com relação às anáforas, pôde-se observar que os tradutores usam esse recurso nas traduções, especialmente quando ocorrem dentro do mesmo segmento de tradução (geralmente do tamanho de uma frase). Na análise, dois casos foram encontrados. Um dentro do mesmo segmento e outro fora. Análises mais abrangentes poderiam dizer se o uso de anáforas causaria ruído suficiente para justificar o esforço de trabalhar esse aspecto no *software* ou não.

Sobre as elipses lexicais anafóricas, há algumas poucas ocorrências, representando 3,08% do ruído linguístico. Embora não seja das mais recorrentes, representa um número considerável de ocorrências. O desafio computacional para esses casos implicaria que o *software* de QA conseguisse identificar o vínculo da unidade, já que ela fica encadeada com outras, como em:

Entrada do glossário	memory usage / uso da memória
Segmento original	Provides the historical data of the CPU usage or <i>memory usage</i> of the last one day
Segmento traduzido	Fornece os dados de histórico do <i>uso</i> da CPU ou <i>da memória</i> no último dia

Contudo, constatou-se que o problema também estava presente no original, e poderia gerar silêncio na detecção, como no exemplo abaixo:

Displays the CPU, *memory*, and hard disk *usage* details of the process.

Exibe os detalhes do processo quanto ao *uso* da CPU, *da memória* e do disco rígido.

Até onde se pôde apurar, o problema da elipse está em unidades nominais. Portanto, um bom ponto de partida seria trabalhar essas unidades nominais, no sentido de contemplar esse tipo de elipse na detecção do *software* de QA. Para fins de estudo, um caminho interessante seria investigar essas elipses em outras unidades, de outras classes gramaticais também, como nesse caso:

Entrada do glossário	download / fazer o download
Segmento original ⁸⁶	The best way to ensure that your purchased product is installed is to log into Norton and <i>download</i> the product that has a current subscription.
Segmento traduzido ⁸⁷	A melhor maneira de garantir que o produto adquirido está instalado é <i>fazer</i> o login no Norton e em seguida <i>o download</i> do produto que tem a assinatura atual.

Seguindo adiante, os casos de flexão verbal também influenciam bastante (6,96% das ocorrências linguísticas). Unidades compostas, como as UFE, representam um desafio a mais aqui, visto que o *software* teria de saber qual dos componentes do sintagma é verbal. Sobre a concordância de gênero, quando houver UT adjetivas ou verbais (em função do particípio) no glossário, é importante que seja contemplada, já que a concordância de gênero é influenciada pela unidade (nome ou sintagma nominal) ao qual modifica. Os adjetivos também influem em número, embora não sejam os únicos. Falando na concordância de número, há dois pontos importantes a considerar. Além da concordância para adjetivos e particípios, existem os casos de generalização (vistos acima em tradução) e os casos em que o termo está na extensão de outra unidade (esses são os casos que mais causam ruído).

Para finalizar essa síntese de resultados, há alguns pontos importantes sobre a análise linguística que poderiam ser levados adiante em estudos futuros. Os parágrafos a seguir discorrem sobre essas ideias.

Quando uma unidade glossariada faz parte de outra unidade em um segmento de tradução, isso poderia influir no ruído, embora não seja uma causa direta. Por exemplo, se a unidade é adjetiva, seu gênero e número dependem do nome ou sintagma nominal do qual faz parte. Quando a unidade glossariada está na extensão de um sintagma no segmento de tradução, pôde-se observar que muitas acabam indo no plural por causa disso. Pela análise, a ocorrência maior desses casos se dá quando tal unidade está em um sintagma nominal. É necessário investigar mais para saber se há influência em outros sintagmas, como os verbais, por exemplo.

Em geral, as unidades compostas apresentam mais ruído (falsos positivos). Sobre elas, como se viu no ponto acima, é possível tirar alguns padrões: quando a unidade glossariada está na extensão de outra unidade no segmento de tradução, há mais chances de que ela apareça no plural do que quando ela aparece fora de um sintagma nominal. A unidade, isolada,

⁸⁶ Fonte: Symantec (2013d).

⁸⁷ Fonte: Symantec (2013e).

também pode aparecer no plural, mesmo que esteja no singular no original. Esse caso, no entanto, talvez seja mais importante de se manter na lista de QA para que o tradutor valide se a mudança é justificável ou não.

O MT também influi em unidades simples. Aplicado a unidades menores, é um dos grandes causadores de falsos positivos. Esse aspecto da língua vale a pena ser estudado mais a fundo na localização.

Há nomes próprios no glossário, e existem referências à UI do *software* que contêm UT ou UFE. Uma referência a UI pode ocorrer tanto no *software* quanto na ajuda. No *software*, a referência é quando um segmento aponta para outro item da interface, o qual está em outra *string*, ou seja, em outro segmento de tradução. As seguintes strings ilustram o caso:

Clique em 'Avançar' para continuar

Avançar

Trata-se de dois segmentos, mas a referência existe apenas no primeiro caso. Isso acontece de maneira semelhante na ajuda também. Esses casos deveriam ser todos ignorados pelo *software*, o qual deveria identificar essas referências de alguma maneira. Uma possibilidade é processual, ou seja, adiciona-se parâmetros de detecção ao *software*, como alguns nomes clássicos de referência, como janela, opção, botão, etc. Outra possibilidade é metodológica: na redação da ajuda e do conteúdo textual do *software*, os redatores começariam a usar algum tipo de marcação (ou *tag*) para identificar as referências a elementos de UI. Estes, no momento do processamento, seriam ou ignorados, ou o programa buscaria as ocorrências do próprio *software* para validar se estão corretas na tradução. Enfim, as referências a UI demonstram a necessidade de manter algum registro ou algum tipo de marcação que indique o que é referência à UI especialmente na ajuda. Além de ajudar no processamento do QA, isso também facilitaria a manutenção da ajuda, visto que mudanças na interface poderiam ser rastreadas mais facilmente se existisse algum dispositivo que as identificasse nos segmentos de tradução.

6.1 Considerações finais

Em geral, uma série de conclusões, hipóteses ou apontamentos futuros poderia ser descrita aqui em algum nível de detalhe. Por questões de espaço, infelizmente, não é possível tratar de todos. Os próximos parágrafos apresentam um breve resumo de alguns pontos que poderiam ser investigados com mais profundidade em trabalhos futuros.

A relação entre incidência de falsos positivos e os *subcorpora* é um deles. O volume de ocorrências entre a ajuda e a UI mostra que a última corresponde a mais ou menos 10% da primeira em todos os casos, embora a interface de usuário componha quase 27% do volume de texto total do *corpus*. Qual seria a explicação para isso? Um estudo sobre a densidade terminológica desses *subcorpora* poderia explicar o motivo. Outra possibilidade de estudo reside no padrão de frase encontrado na UI, que talvez apontem para um uso limitado de formas verbais e, conseqüentemente, um número reduzido de ruído em função disso.

Outro ponto importante reside em como fazer o *software* identificar determinados padrões estruturais. Muito se falou ao longo do trabalho de classes gramaticais, flexão verbal, sintagmas nominais, adjetivais, etc. Para processar essas informações da maneira mais adequada possível, dentro dos limites da tecnologia, o *software* de QA precisa identificar esses aspectos. Nesse sentido, executar o *parsing* do *corpus* ao processar a verificação terminológica seria um passo importante para melhorar a qualidade das listas de QA. A partir da identificação de sintagmas, o *software* poderia passar a prever melhor o comportamento de determinadas unidades (em especial para os casos em que a unidade glossariada faz parte da extensão do sintagma).

Futuramente, será necessário equacionar todos esses fatores para evitar redundâncias ao fazer a representação formal dos dados e resultados obtidos neste trabalho. Alguns pontos específicos já começaram a ser delineados, por exemplo, na esquematização dos padrões de inserção vistos na análise fraseológica (p. 167) e na catalogação das variantes com base na tipologia adotada (pp. 196-216).

Este é um passo importante na direção de um modelo eficaz para o processamento do QA terminológico, visto que, como se pôde observar, um único falso positivo pode apresentar várias causas, como no exemplo a seguir:

Entrada do glossário	download / fazer o download
Segmento original	<i>Download</i> away – we scan the files you download and eliminate online threats that we find.
Segmento traduzido	<i>Faça downloads</i> sem preocupação. Verificamos os arquivos que você obtém por download e eliminamos as ameaças online encontradas.

O caso envolve os seguintes aspectos:

- morfologia na flexão do verbo *fazer* da tradução;
- fraseologia, pois a unidade é UFE na tradução;
- variação morfossintática, visto que a unidade não apresenta o artigo “o” indicado no glossário, além de trazer *download* no plural.

Além desses pontos, há a representação formal da unidade, a qual revela a relação assíncrona dos idiomas mencionada anteriormente. A unidade do original é verbal e simples; na tradução, também é verbal, mas composta. Original e tradução devem ser tratados em separado por um *software* de QA. O fato de haver uma inserção, por exemplo, depende disso. Hoje, o cruzamento de padrões usado pelo *software* serve para o original e a tradução. No futuro, o caminho ideal para uma ferramenta linguisticamente motivada seria ter módulos em separado para cada idioma e, talvez, algumas funções conjuntas, partilhadas por idiomas específicos. Mas, em última análise, cada idioma seria tratado à parte.

Ainda nesse raciocínio de equacionar as categorias e os pontos de análise deste trabalho, na apresentação dos resultados foi brevemente levantado o problema de sobreposição na classificação aqui adotada (p. 174-175). Então, há classificações que se superpõem, como no seguinte caso:

fazer download

fazer o download

Como se viu no trabalho, a unidade glossariada é *fazer o download*. A presença ou ausência do artigo é prevista na tipologia de variação usada e explicada como um elemento de inserção em fraseologia. Outro exemplo que revela a questão é no MT aplicado à fraseologia versus técnica de transposição na tradução (p. 175).

Uma hipótese para o problema é uma possível diferença nas perspectivas adotadas para a classificação das ocorrências. No entanto, isso não parece muito provável, dada a simila-

ridade da visão das propostas. Nesse sentido, os aspectos duplicados sugerem que, na verdade, existe uma camada mais básica a considerar, como se viu na análise linguística. Todas as propostas adotadas por este trabalho usam uma teoria semântica semelhante (cognitiva, comunicativa, social, funcional, etc.), embora essa base não seja explicitada. Enfim, a questão que fica é como resolver isso. Talvez trabalhos futuros devam explicitar a base semântica sobre a qual assentar todos os casos analisados. Um passo nessa direção seria falar mais sobre essa fundamentação e separar os parâmetros que dela provêm. Ou seja, essa base semântica poderia ser adotada em um âmbito geral, derivando para os outros casos, assim eliminando as sobreposições.

Ainda na esteira das classificações sobrepostas, outro ponto que veio à tona na análise é o duplo processamento desnecessário que o *software* executa, visível em dois casos específicos:

- MT: o original tem a mesma forma para classes gramaticais, e o glossário registra as unidades em separado:

download / download (*)

download / fazer o download

- sintagmas compostos: a unidade simples ou complexa é registrada no glossário individualmente e também como parte de outra unidade glossariada:

rule / regra

rule list / lista de regras

Como se pôde constatar na análise, para ambos os casos, o *software* processa as duas unidades que têm a mesma forma em inglês, mesmo que já tenha identificado a ocorrência em questão como positivo ou falso positivo. Isso gera ruído e é desnecessário. Ou seja, o *software* teria de saber quando não deve processar mais determinada ocorrência. Esse problema deve ser levado em conta também em uma ferramenta linguisticamente motivada, na qual as unidades não dependeriam mais exclusivamente da forma para serem identificadas.

O caso de MT revela que os glossários precisam estar mais bem estruturados. É preciso indicar as unidades que têm relação umas com as outras – ou agrupá-las, indicando suas formas superficiais, em casos de unidades como *download*, cujas formas superficiais estão todas atreladas a uma estrutura de base.

Já o caso de sintagmas compostos mostra que o *software* teria de saber qual unidade processar primeiro (ou saber descartar alguma ocorrência da lista, caso mais tarde no processamento encontre outra unidade que se encaixe exatamente na detecção). Definir a ordem de processamento, no entanto, parece demandar alguma complexidade. Por exemplo, para alguns casos, como em *download* versus *Download Insight*, a prevalência poderia ser das unidades de maior estrutura para a de menor. Após processar essa unidade, o *software* poderia marcar a ocorrência de alguma maneira, deixando a unidade *Download Insight* detectada no segmento de lado nas rodadas de processamento seguintes. O problema seria para unidades com estruturas do mesmo tamanho, como neste caso:

Entrada do glossário	Norton AntiVirus / Norton AntiVirus antivirus scan / verificação antivírus
Segmento original	A <i>Norton AntiVirus scan</i> can identify threats on your computer in the following ways:
Segmento traduzido	Uma <i>verificação do Norton AntiVirus</i> pode identificar ameaças em seu computador das seguintes maneiras:

As duas unidades glossariadas têm o mesmo número de componentes e ambas são igualmente detectadas pelo *software*. Mas qual deveria prevalecer? Nesse caso, a unidade que de fato aparece é *Norton AntiVirus*. Talvez os nomes próprios tenham preferência nesse caso. Em última análise, fica nítido que seria necessário definir prioridades de processamento para o *software*.

Um ponto importante de destacar neste trabalho como apontamento futuro é o estudo do silêncio. Este trabalho analisa apenas as ocorrências geradas pela verificação do QA e, sendo assim, trata especialmente do ruído encontrado na parte traduzida dos segmentos do *corpus*. Algumas das características analisadas, como a inserção (pp. 167 e 222) e a flexão verbal (p. 220) também são aplicáveis ao original e sugerem que ocorrências são deixadas de fora da lista de QA. Por acaso, foi possível identificar alguns casos de silêncio durante a análise.

Certamente, investigar o silêncio é imperativo para uma análise definitiva da eficácia do *software* de QA. Esse trabalho pôde constatar que, das ocorrências detectadas pelo *software*, boa parte consiste em falsos positivos. Não há motivos para pensar diferente no sentido contrário, ou seja, que exista uma boa dose de silêncio, a qual impediria a detecção de ocorrências, quem sabe, inadequadas ou incorretas.

Um estudo sobre o silêncio poderia revelar o volume exato de ocorrências que um *software* de QA deixa de registrar. Uma hipótese é que os resultados poderiam ir de encontro a um dos argumentos de venda usados por empresas que comercializam esse tipo de *software* (ou que comercializam isso como um recurso de sua ferramenta CAT), já que o QA terminológico é vendido como uma ferramenta que ajudaria a assegurar a consistência da terminologia em um projeto de localização. Justiça seja feita, a ferramenta ajuda, mas, ao que parece, está longe de oferecer uma cobertura que dê a margem de segurança vendida por essas empresas.

Cabe aqui ressaltar outra vez o papel da etapa de QA no processo de localização como um todo. Trata-se de uma etapa de garantia, de ratificar que os demais processos, TEP, foram eficazes e produziram uma tradução adequada, coerente, correta e dentro dos padrões esperados.

O trabalho também traz uma série de questionamentos a respeito das atividades e do papel do tradutor no setor da localização, os quais também sugerem algumas linhas de investigação que, embora sejam tangenciais ao trabalho aqui apresentado, também poderiam enriquecer os estudos linguísticos com enfoque em localização, especialmente se o paradigma da localização for usado como base. Segue uma relação dos principais pontos que poderiam ser estudados:

- com a localização, o que muda no papel do tradutor técnico?
- com a localização, o tédio do tradutor técnico aumenta?
- quais são os desafios envolvidos em trabalhar uma tradução com contexto e cotexto limitados (ou, ainda, sem contexto nem cotexto)? (Algumas pistas são apontadas, como, por exemplo, acesso a bases de referência, memórias de tradução; busca de contexto na UI; interação com outros participantes no processo, como cliente, editores, revisores).

Para concluir, espera-se que todas essas sugestões de investigações futuras e o trabalho como um todo mostrem que os objetivos propostos foram atingidos, com a descrição do cenário da localização, bem como de suas práticas, com as incoerências, inconsistências que deveriam ser sanadas em glossários ou processos, já que, como se viu, não é apenas o processamento computacional que precisa ser aperfeiçoado. Espera-se também que o trabalho tenha contribuído com elementos que permitam progredir para o segundo passo, ou seja, trabalhar em uma representação formal que permita implementar uma ferramenta de QA terminológico linguisticamente motivada que facilite o trabalho do tradutor.

7 REFERÊNCIAS

- AGUADO DE CEA, G. A multiperspective approach to specialized phraseology: Internet as a reference corpus for phraseology. In: SANTIAGO POSTEGUILLO, M. J. E. M. L. G. V. **The Texture of Internet: Netlinguistics in Progress**. Newcastle: Cambridge Scholars Publishing, 2007. Cap. 9, p. 182-207.
- ARROJO, R. **Tradução, Deconstrução e Psicanálise**. Rio de Janeiro: Imago, 1993.
- BALLY, C. **Traité de stylistique française**. Paris: Klincksieck, v. 1, 1951.
- BASSNETT, S. **Translation Studies**. 3. ed. Londres e Nova York: Routledge, 2002.
- BÉJOINT, H.; THOIRON, P. Pour un index évolutif et cumulatif de cooccurrents en langue techno-scientifique sectorielle. **Meta: Journal des Traducteurs**, Montréal, v. 34, n. 4, p. 661-671, 1989.
- BELL, R. T. **Translation and Translating, Theory and Practice**. Londres e Nova York: Longman, 1991.
- BERBER SARDINHA, T. **Linguística de Corpus**. Barueri: Manole, 2004.
- BEVILACQUA, C. R. **A FRASEOLOGIA JURÍDICO-AMBIENTAL**. Porto Alegre: [s.n.], 1996. Dissertação de Mestrado em Estudos da Linguagem. Instituto de Letras. Universidade Federal do Rio Grande do Sul.
- BEVILACQUA, C. R. **Unidades Fraseológicas Especializadas**: Estado de la Cuestión y Perspectivas. Barcelona: Instituto Universitario de Lingüística Aplicada, Universitat Pompeu Fabra, 1999. Trabalho de Pesquisa (Doutorado em Lingüística Aplicada).
- BEVILACQUA, C. R. **Unidades Fraseológicas Especializadas Eventivas**: descripción y reglas de formación en el ámbito de la energía solar. BARCELONA - PORTO ALEGRE: [s.n.], 2004. Tese de Doutorado em Lingüística Aplicada (Léxico). Instituto Universitario de Lingüística Aplicada (IULA). Universidad Pompeu Fabra.
- BLAIS, E. Le phraséologisme. Une hypothèse de travail. **Terminologies Nouvelles**, Bélgica, n. 10, p. 50-56, 1993.
- CABRÉ, M. T. **Terminology: Theory, methods and applications**. Amsterdam/Philadelphia: John Benjamins Publishing Company, v. 1, 1998a.
- CABRÉ, M. T. Elementos para una teoría de la terminología: hacia un paradigma alternativo. **El llinguaraz**, Buenos Aires, 1, n. 1, 1998b. 59-77.
- CABRÉ, M. T. Una nueva teoría de la terminología: de la denominación a la comunicación. In: CABRÉ, M. T. **La terminología: representación y comunicación**. Barcelona: Universitat Pompeu Fabra, 1999b. p. 109-128.
- CABRÉ, M. T. Principios teóricos sobre la terminología, ámbito y unidades de estudio. Sumario de principios y consecuencias metodológicas. In: CABRÉ, M. T.; FELIU, J. **La**

terminología científico-técnica: reconocimiento, análisis y extracción de información formal y semántica. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada, 1999c. p. 173-187.

CABRÉ, M. T. Sumario de principios que configuran la nueva propuesta teórica. In: CABRÉ, M. T.; FELIU, J. **La terminología científico-técnica:** reconocimiento, análisis y extracción de información formal y semántica. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada, 2001. p. 17-26.

CABRÉ, M. T. Theories of Terminology: Their description, prescription and explanation. **Terminology**, Amsterdam, v. 9, n. 2, p. 163-199, 2003.

CABRÉ, M. T.; LORENTE, M.; ESTOPÀ, R. **Terminología y fraseología.** Actas del V Simposio de Terminología Iberoamericana. Ciudad de México: Colegio de México. 1996. p. 67-81.

CABRÉ, T. **La terminología:** representación y comunicación – elementos para una teoría de base comunicativa y otros artículos. Barcelona: Institut Universitari de Lingüística Aplicada, 1999a.

CASARES, J. **Introducción a la lexicografía moderna.** Madrid: Consejo Superior de Investigaciones Científicas, 1950. 354 p.

CATFORD, J. C. **A Linguistic Theory of Translation:** an Essay in Applied Linguistics. Londres: Oxford University Press, 1965.

CHOMSKY, N. **Aspects of the Theory of Syntax.** Cambridge: The MIT press, 1965.

CIAPUSCIO, G. E. Reformulación textual: El caso de las noticias de divulgación científica. **Revista Argentina de Lingüística**, Buenos Aires, v. 9, n. 1-2, p. 69-117, 1993.

CIAPUSCIO, G. E. Los científicos explican: la reformulación del léxico experto en la consulta oral. **Cadernos do IL**, Porto Alegre, v. 18, p. 37-47, 1997a.

CIAPUSCIO, G. E. Lingüística y divulgación de la ciencia. **Quark**, Barcelona, v. 7, p. 19-28, 1997b.

CIAPUSCIO, G. E. La terminología desde el punto de vista textual: selección, tratamiento y variación. **Organon**, Porto Alegre, v. 26, p. 43-65, 1998.

CIAPUSCIO, G. E. **Textos especializados y terminología.** Barcelona: IULA (UPF) Institut Universitari de Lingüística Aplicada, 2003.

COHEN, B. Méthodes de repérage et de classement des cooccurrents lexicaux. **Terminologie et traduction**, Bruselas, Bruselas, v. 2, n. 3, p. 505-511, 1992.

CORBEIL, J.-C. **Les terminologies devant Babel.** Actes du Colloque Terminologie et Technologies Nouvelles. La Défense, du 9 au 11 décembre 1985. Quebec: OLF. 1988. p. 49-62.

DELISLE, J. L'analyse du discours comme méthode de traduction. **Folia Linguistica**, v. 13, n. 1-2, p. 55-62, 1980, 2009.

DEPALMA, D. A. **Business without Borders**. Nova York: Wiley, 2002.

DESMET, I. **Pour une approche terminologique des sciences sociales et humaines. Les sciences sociales et humaines du travail en portugais et en français**. Paris: Université Paris-Nord (Paris XIII), v. II, 1995-1996. Tese de Doutorado.

DIAS-DA-SILVA, B. C. **A Face Tecnológica dos Estudos da Linguagem**: o processamento automático das línguas naturais. Araraquara: [s.n.], 1996. 274 p. Tese (Doutorado em Lingüística e Língua Portuguesa). Faculdade de Ciências e Letras. Universidade Estadual Paulista.

DIAS-DA-SILVA, B. C. O estudo Lingüístico-Computacional da Linguagem. **Letras de Hoje**, v. 41, n. 2, p. 103-138, Junho 2006.

DUNNE, K. J. A Copernican revolution. In: DUNNE, K. J. **Perspectives on Localization**. Amsterdam/Filadélfia: John Benjamins Publishing Company, v. XIII, 2006. Cap. 1, p. 1-11.

EMC. CLARiiON Host Encryption. **emc.com**, 2008a. Disponível em: <www.emc.com/products/detail/software/clariion-host-encryption.htm>. Acesso em: 21 Junho 2008.

EMC. CLARiiON Host Encryption. **brazil.emc.com**, 2008b. Disponível em: <<http://brazil.emc.com/products/detail/software/clariion-host-encryption.htm>>. Acesso em: 02 Maio 2013.

EMC. EMC CLARiiON Host Encryption. **emc.com**, 2010a. Disponível em: <www.emc.com/collateral/software/data-sheet/h7286-clariion-host-encryption-ds.pdf>. Acesso em: 2013 Maio 02.

EMC. emc.com. **EMC VNX Host Encryption**, 2013. Disponível em: <www.emc.com/collateral/software/data-sheet/h8571-ds-emc-powerpath-encryption.pdf>. Acesso em: 02 Maio 2013.

EMC CLARiiON Host Encryption. **brazil.emc.com**, 2010b. Disponível em: <brazil.emc.com/collateral/software/data-sheet/h7286-clariion-host-encryption-ds.pdf>. Acesso em: 02 Maio 2013.

ESSELINK, B. **A Practical Guide to Localization**. Amsterdam/Filadélfia: John Benjamins Publishing Company, v. 4, 2000.

ETTINGER, S. Formación de palabras y fraseología en la lexicografía. In: HAENSCH, G. **La lexicografía**: de la lingüística teórica a la lexicografía práctica. Madrid: Gredos, 1982. p. 233-258.

FABER, P.; MAIRAL USÓN, R. **Constructing a Lexicon of English Verbs**. Berlin, Nova York: Mouton de Gruyter, 1999.

FELBER, H.; PICHT, H. **Métodos de terminografía y principios de investigación terminológica**. Madrid: Instituto “Miguel de Cervantes” - CSIC, 1984.

FERNÁNDEZ POLO, F. J. Traducción y retórica contrastiva. A propósito de la traducción de textos de divulgación científica del inglés al español. **Revista Lucense de Lingüística y Literatura**, Santiago de Compostela, 1999.

FIALA, P. Pour une approche discursive de la phraséologie. Remarques en vrac sur la locutionnalité et quelques points de vue qui s'apparentent, sans doute. **Langage et société**, n. 42, p. 27-44, 1987.

FREIXA, J. Reconocimiento de unidades denominativas: incidencia de la variación en el reconocimiento de las unidades terminológicas. In: CABRÉ, M. **La Terminología científico-técnica: reconocimiento, análisis y extracción de información formal y semántica**. Barcelona: IULA, 2001. p. 57-65.

FREIXA, J. **La variació terminològica: anàlisi de la variació denominativa en textos de diferent grau d'especialització de l'àrea de medi ambient**. Barcelona: [s.n.], 2002. Tese de Doutorado. Universitat Pompeu Fabra.

FREIXA, J. Variación terminológica: ¿Por qué y para qué? **Translators' Journal**, 2005. Disponible em: <<http://id.erudit.org/iderudit/019917ar>>. Acesso em: 28 Abril 2013.

GAMBIER, Y. Présupposés de la terminologie: vers une remise en cause. **Cahiers de Linguistique Sociale**, v. 18, p. 31-58, 1991.

GOUADEC, D. **Nature et traitement des entités phraséologiques**. Terminologie et phraséologie. Acteurs et aménageurs: Actes de la deuxième. Paris: La Maison du Dictionnaire. 1994. p. 167-193.

GRÉCIANO, G. Vers une modélisation phraséologique: Acquis et projets d'EUROPHRAS. **Terminologies Nouvelles**, Bruxelles, p. 16-22, 1993.

GROSS, G. **Réflexions sur le traitement automatique des langues**. Actes de JADT. Paris: [s.n.]. 2004. p. 545-556.

GUENTHNER, F.; BLANCO, X. Multi-lexemic Expressions: An Overview. In: LECLÈRE, C., et al. **Lexique, Syntaxe et Lexique-Grammaire**. Amsterdam: John Benjamins Publishing Company, 2004. p. 239-252.

HATIM, B.; MASON, I. **Discourse and the Translator**. Londres: Longman, 1990.

HAUSMANN, F. Le Dictionnaire de Collocations. In: HAUSMANN, F.; REICHMANN, O.; WIEGAND, H. **Wörterbücher, Dictionaries, Dictionnaires. Ein internationales Handbuch zur Lexikographie**. Berlin, Nova York: Walter de Gruyter, 1989. p. 1010-1018.

HEID, U. Décrire les collocations: deux approches lexicographiques et leur application dans un outil informatisé. **Terminologie et traduction**, Bruxelles, v. 2, n. 3, p. 523-548, 1992.

HEID, U. **Towards a corpus-based dictionary of German noun-verb collocations**. Euralex'98 Proceedings. Liège: Université de Liège. 1998. p. 513-522.

HERMANS, T. Translational norms and correct translations. In: LEUVEN-ZWART, K. M. V.; NAAIJKENS, T. **Translation Studies: The State of the Art**. Amsterdam: Rodopi, 1991. p. 155-169.

HEWSON, L.; MARTIN, J. **Redefining Translation: The Variational Approach**. Londres: Routledge, 1991.

HOUSE, J. **A Model for Translation Quality Assessment**. Tübingen: Gunter Narr, 1977.

HOUSE, J. **Translation Quality Assessment. A Model Revisited**. Tübingen: Gunter Narr, 1997.

HURTADO ALBIR, A. **Traducción y traductología: introducción a la traductología**. 1. ed. Madrid: Cátedra, 2001.

ISO. ISO 9001:2000(E). **Quality management systems — Requirements**, Genebra, 2000.

ISO. Codes for the representation of names of languages - Part 1: Alpha-2 code. **iso.org**, 2013a. Disponível em:

<http://www.iso.org/iso/country_codes/iso_3166_code_lists/country_names_and_code_elements.htm><http://www.iso.org/iso/catalogue_detail?csnumber=22109>. Acesso em: 28 Abril 2013.

ISO. ISO 3166-1-alpha-2 code. **iso.org**, 2013b. Disponível em:

<http://www.loc.gov/standards/iso639-2/php/code_list.php>. Acesso em: 28 Abril 2013.

KELLY, N.; DEPALMA, D. A. The Language Services Market: 2012. **Common Sense Advisory**, 31 Maio 2012. Disponível em:

<http://www.commonsenseadvisory.com/Portals/0/downloads/120531_QT_Top_100_LSPs.pdf>. Acesso em: 28 Abril 2013.

KJAER, A. L. Phraseology research - State-of-art. Methods of describing word combinations in language for specific purposes. **Terminology science & research: Journal of International Institute for Terminology Research**, Viena, v. 2, n. 1-2, p. 3-20, 1990.

KRIEGER, M. D. G. Do Ensino da Terminologia para Tradutores: Diretrizes Básicas.

Cadernos de Tradução, Florianópolis, v. 1, n. 17, p. 189-206, 2006. Disponível em: <<https://periodicos.ufsc.br/index.php/traducao/article/view/6862/6413>>. Acesso em: 10 Abril 2013.

L'HOMME, M.-C. **Caractérisation des combinaisons lexicales spécialisées par rapport aux collocations de langue générale**. Euralex '98 Proceedings. Liège: Université de Liège. 1998.

L'HOMME, M.-C. Understanding specialized lexical combinations. **Terminology. International Journal of theoretical and applied issues in specialized communication**, Amsterdã, v. 6, n. 1, p. 89-110, 2000.

L'HOMME, M.-C.; BERTRAND, C. **Specialized Lexical Combinations: Should they be Described as Collocations or in Terms of Selectional Restrictions?** Proceedings. Ninth EURALEX International Congress. Stuttgart: Universität Stuttgart. 2000. p. 497-506.

LAINÉ, C. et al. La phraséologie - nouvelle dimension de la recherche terminologique. Travaux du module canadien du RINT. **L'Actualité terminologique**, Canada, v. 25, n. 3, p. 5-9, 1992.

LEVY, J. **Die literarische Übersetzung: Theorie einer Kunstgattung**. Frankfurt a.M.: Athenäum, 1969.

LISA – LOCALIZATION INDUSTRY STANDARDS ASSOCIATION. SRX 2.0 Specification – OSCAR Recommendation, 7 April 2008. **Globalization and Localization Association**, 2008. Disponível em: <<http://www.gala-global.org/oscarStandards/srx/srx20.html>>. Acesso em: 09 Outubro 2011.

LISA – LOCALIZATION INDUSTRY STANDARDS ASSOCIATION. TMX 1.4b Specification. **Globalization and Localization Association**, 2011. Disponível em: <<http://www.gala-global.org/oscarStandards/tmx/tmx14b.html>>. Acesso em: 09 Outubro 2011.

LISA. **The Localization Industry Primer**. 2. ed. Féchy: The Localization Industry Standards Association, 2003. Disponível em: <<http://www.ict.griffith.edu.au/~davidt/cit3611/LISAprimer.pdf>>. Acesso em: 15 Janeiro 2012.

LOMMEL, A. R.; RAY, R. **The Globalization Industry Primer**. Romainmôtier: The Localization Industry Standards Association, 2007.

LORENTE, M.; BEVILACQUA, C. R.; ESTOPÀ, R. **El análisis de la fraseología especializada mediante elementos de la lingüística actual**. Atas VI Simpósio da Rede Ibero-americana de Terminologia. Terminología, desenvolvimento e identidade nacional. Lisboa: Colibrí/ILTEC. 2002. p. 647-666.

MACMILLAN. string - definition. **Macmillan Dictionary**, 2013a. Disponível em: <www.macmillandictionary.com/dictionary/british/string>. Acesso em: 29 Abril 2013.

MACMILLAN. crowd-sourcing - definition. **Macmillan Dictionary**, 2013b. Disponível em: <www.macmillandictionary.com/dictionary/british/crowd-sourcing>. Acesso em: 29 Abril 2013.

MARTIN, W. Remarks on Collocations in sublanguages. **Terminologie et traduction**, Bruxelas, 2, n. 3, 1992. 157-164.

MARTON, G.; KATZ, B. Exploring the Role of Part of Speech in the Lexicon. **Massachusetts Institute of Technology**, 2002. Disponível em: <<http://www.ai.mit.edu/research/abstracts/abstracts2002/natural-language/08marton.pdf>>. Acesso em: 28 Abril 2013.

MICROSOFT. Microsoft Office. **Microsoft UK**, 2013a. Disponível em: <office.microsoft.com/en-gb>. Acesso em: 04 Maio 2013.

MICROSOFT. Microsoft Office. **Microsoft Brasil**, 2013b. Disponível em: <office.microsoft.com/pt-br>. Acesso em: 4 Maio 2013.

MICROSOFT. Microsoft Office. **Microsoft España**, 2013c. Disponível em: <office.microsoft.com/es-es>. Acesso em: 4 Maio 2013.

MICROSOFT. Microsoft Office. **Microsoft Deutschland**, 2013d. Disponível em: <office.microsoft.com/de-de>. Acesso em: 4 Maio 2013.

MICROSOFT. Microsoft Office. **Microsoft Italia**, 2013e. Disponível em: <office.microsoft.com/it-it/>. Acesso em: 4 Maio 2013.

MICROSOFT. Microsoft Office. **Microsoft France**, 2013f. Disponível em: <office.microsoft.com/fr-fr>. Acesso em: 4 Maio 2013.

MICROSOFT. Portal de Idiomas. **Microsoft**, 2013g. Disponível em: <www.microsoft.com/language/pt-br/default.aspx>. Acesso em: 5 Maio 2013.

MOLINA, L. **Análisis descriptivo de la traducción de los culturemas árabe-español**. Barcelona: Universitat Autònoma de Barcelona, 2001. Tese de Doutorado.

MOLINA, L.; HURTADO ALBIR, A. Translation techniques revisited: A dynamic and functionalist approach. **Translators' Journal**, v. 47, n. 4, p. 498-512, 2002.

MYERS, G. Lexical Cohesion and Specialized Knowledge in Science and Popular Texts. **Discourse Processes**, v. 14, p. 1-26, 1991.

MYKING, J. No fixed boundaries. In: ANTIA, B. E. **Indeterminacy in Terminology and LSP**. Amsterdam: John Benjamins Publishing Company, 2007. Cap. 6, p. 236.

NETAPP. Ask Dr Dedupe. **NetApp Community**, 2009. Disponível em: <<https://communities.netapp.com/community/netapp-blogs/drdedupe/blog/2009/09/16/you-lie-who-coined-the-term-data-deduplication>>. Acesso em: 29 Abril 2013.

NEWMARK, P. **Approaches to translation**. Oxford: Pergamon, 1981.

NIDA, E. A. **Toward a science of translating: with special reference to principles and procedures involved in Bible translating**. Leiden: Brill Archive, 1964.

NIDA, E. A. **Toward a science of translating: with special reference to principles and procedures involved in Bible translating**. Leiden: Brill Archive, 1964.

NIDA, E. A.; TABER, C. R. **The Theory and Practice of Translation**. Leiden: E.J. Brill, 1969.

NORD, C. **Textanalyse und Übersetzen. Theoretische Grundlagen, Methoden und didaktische Anwendung einer übersetzungsrelevanten Textanalyse**. Heidelberg: Julius Groos, 1988.

NORMAND, C. Le CLG: une théorie de la signification? In: NORMAND, C. **La quadrature du sens**. Paris: Presses Universitaires de France, 1990. p. 23-40.

OPENTAG.COM. \ Technology : Formats : TBX. **opentag.com**, 2013a. Disponível em: <www.opentag.com/tbx.htm>. Acesso em: 29 Abril 2013.

OPENTAG.COM. \ Technology : Formats : TMX. **opentag.com**, 2013b. Disponível em: <www.opentag.com/tmx.htm>. Acesso em: 29 Abril 2013.

OPENTAG.COM. \ Technology : Formats : XLIFF. **opentag.com**, 2013c. Disponível em: <www.opentag.com/xliff.htm>. Acesso em: 29 Abril 2013.

PAVEL, S. La phraséologie en langue de spécialité. Méthodologie de consignation dans les vocabulaires terminologiques. **Terminologies Nouvelles**, Bélgica, n. 10, p. 67-82, 1993.

- PESANT, G.; THIBAUT, E. Terminologie et cooccurrence en langue du droit. **Terminologies Nouvelles**, Bruxelles, v. 10, p. 23-35, 1993.
- PICHT, H. LSP Phraseology from the terminological point of view. **Journal of the International Institute for Terminology Research**, v. 1, n. 1-2, 1990.
- POTTIER, B. **Linguística geral**: teoria e descrição. Rio de Janeiro: Presença/Universidade Santa Úrsula, 1978. 320 p.
- PYM, A. Localization and Linguistics. **Anthony Pym**, 2001. Disponível em: <<http://usuaris.tinet.cat/apym/on-line/translation/loclinguistics.pdf>>. Acesso em: 30 Abril 2013.
- PYM, A. What Localization Models Can Learn From Translation Theory, Tarragona, 2003. Disponível em: <http://usuaris.tinet.cat/apym/on-line/translation/localization_translation_theory.pdf>. Acesso em: 15 Janeiro 2012.
- PYM, A. **The moving text**: localization, translation, and distribution. Amsterdam/Filadélfia: John Benjamins Publishing Company, 2004.
- PYM, A. Redefinindo Competência Tradutória em uma era Eletrônica. Em defesa de uma Abordagem Minimalista. **Cadernos de Tradução**, Florianópolis, I, n. 21, 2008. 9-40.
- PYM, A. Western translation theories as responses to equivalence. **Anthony Pym**, 2009. Disponível em: <http://usuaris.tinet.cat/apym/on-line/translation/2009_paradigms.pdf>. Acesso em: 29 Abril 2013.
- PYM, A. **Exploring Translation Theories**. Nova York: Routledge, 2010.
- PYM, A. Translation research terms: a tentative glossary for moments of perplexity and dispute. **Translation Research Projects**, Tarragona, p. 75-110, 2011. Disponível em: <http://isg.urv.es/publicity/isg/publications/trp_3_2011/index.htm>. Acesso em: 28 Abril 2013.
- REISS, K.; VERMEER, H. J. **Groundwork for a general theory of translation**. Tübingen: Niemeyer, 1984.
- REISS, K.; VERMEER, H. J. **Fundamentos para una teoría funcional de la traducción**. Madrid: Ediciones Akal, 1996.
- ROBERTS, R. Identifying the phraseology of languages for special purposes (LSPs). **Alfa: Actes de langue française et de linguistique**, Canadá, 7/8, 1994-1995. 61-69.
- ROCHE, C. **Should Terminology Principles be Re-examined?** Proceedings of the 10th Terminology and Knowledge Engineering Conference (TKE 2012). Madrid: Aguado de Cea et al. 2012. p. 17-32.
- SAG, I. A. et al. Multiword Expressions: A Pain in the Neck for NLP. **Computational Linguistics and Intelligent Text Processing**, Heidelberg, 2002. 1-15.
- SAP. SAPterm - SAP Terminology. **SAPterm**, 2013. Disponível em: <sapterm.com/>. Acesso em: 5 Maio 2013.

SDL. LISA QA Metric, 2013. Disponível em: <producthelp.sdl.com/SDL_TMS_2011/en/Creating_and_Maintaining_Organizations/Managing_QA_Models/LISA_QA_Model.htm>. Acesso em: 4 Maio 2013.

SDL. sdl.com. **SDL - Global Customer Experience Management**, 2013. Disponível em: <www.sdl.com>. Acesso em: 29 Abril 2013.

SELESKOVITCH, D.; LEDERER, M. **Intérpreter pour traduire**. 2. ed. Paris: Didier Érudition, 1984.

SNELL-HORNBY, M. **Translation Studies: An Integrated Approach**. Amsterdam: John Benjamins Publishing Company, 1988.

STEINER, T. R. **English Translation Theory 1650-180**. Assen: van Gorcum, 1975.

STRZALKOWSKI, T. et al. Evaluating Natural Language Processing Techniques IN Information Retrieval: A Trec Perspective. **Natural Language Information Retrieval**, Dordrecht, p. 113-145, 1999. Disponível em: <<http://www.calstatela.edu/faculty/jperezc/Publications/kluwer.chapter.pdf>>. Acesso em: 28 Abril 2013.

SUÁREZ, M. **Análisis Contrastivo de la variación denominativa en textos especializados: del texto original al texto meta**. Barcelona: IULA, 2004. Tese de Doutorado. Universitat Pompeu Fabra.

SYMANTEC. Norton Confidencial: A primeira solução ampla de segurança para transações online do mundo. **Symantec**, 2006. Disponível em: <www.symantec.com/pt/br/about/news/release/article.jsp?prid=20061019_01>. Acesso em: 5 Maio 2013.

SYMANTEC. Symantec lança versão do Norton AntiVirus 2009 especialmente para gamers. **Symantec**, 2008. Disponível em: <www.symantec.com/pt/br/about/news/release/article.jsp?prid=20081125_01>. Acesso em: 5 Maio 2013.

SYMANTEC. Erro: "A extração falhou. O arquivo está corrompido" ao instalar um produto Norton de download. **Norton**, 2010. Disponível em: <support.norton.com/sp/pt/br/home/current/solutions/20080702134123BR?entsrc=redirect_pu bweb&product=home&pvid=f-home&version=1>. Acesso em: 5 Maio 2013.

SYMANTEC. Norton License Agreement - Norton AntiVirus or Norton Internet Security. **Symantec**, 2011a. Disponível em: <www.symantec.com/content/en/us/about/media/NAV-NIS-2011-Subscription-EULA-USE.pdf>. Acesso em: 4 Maio 2013.

SYMANTEC. Onde encontrar o código do produto. **Norton**, 2011b. Disponível em: <support.norton.com/sp/pt/br/norton-management/current/solutions/v1027402_NIS_OEM_2012_pt_br?ispid=&product=Norton+Management&pvid=f-nm&version=current>. Acesso em: 5 Maio 2013.

SYMANTEC. Norton AntiVirus. **Norton**, 2013a. Disponível em: <br.norton.com/antivirus/>. Acesso em: 4 Maio 2013.

SYMANTEC. Symantec Language Portal. **Norton Together**, 2013b. Disponível em: <<http://together.norton.com/symtd/search>>. Acesso em: 4 Maio 2013.

SYMANTEC. Loja online da Symantec para usuários domésticos > Serviços Premium. **Norton**, 2013c. Disponível em: <buy.norton.com.gtm.symantec.com/pt-br/mf/premiumServices>. Acesso em: 5 Maio 2013.

SYMANTEC. My Norton product shows an incorrect subscription date. **Norton**, 2013d. Disponível em: <support.norton.com/sp/en/us/home/current/solutions/kb20100920013943EN_EndUserProfile_en_us>. Acesso em: 5 Maio 2013.

SYMANTEC. Meu produto Norton exibe uma data de assinatura incorreta. **Norton**, 2013e. Disponível em: <support.norton.com/sp/pt/br/home/current/solutions/kb20100920013943EN_EndUserProfile_pt_br>. Acesso em: 5 Maio 2013.

TOURY, G. **In Search of a Theory of Translation**. Tel Aviv: The Porter Institute for Poetics and Semiotics, 1980.

TRISTÁ, A. M. **Fraseología y contexto**. Havana: Editorial de Ciencias Sociales, 1988. 193 p.

VÁZQUEZ-AYORA, G. **Introducción a la traductología**. Washington: Georgetown University Press, 1977.

VINAY, J.-P.; DARBELNET, J. L. **Comparative Stylistics of French and English: A Methodology for Translation**. Amsterdam: John Benjamins Publishing Company, v. 11, 1995.

WÜSTER, E. **Introducción a la Teoría General de la Terminología y la Lexicografía Terminológica**. Barcelona: Documenta Universitaria, 1998.

ZHANG, Y.; PATRICK, J. Paraphrase Identification by Text Canonicalization. **Proceedings of the Australasian Language Technology Workshop 2005**, Sydney, p. 160-166, 2005. Disponível em: <<http://aclweb.org/anthology-new/U/U05/U05-1023.pdf>>. Acesso em: 28 Abril 2013. <http://aclweb.org/anthology-new/U/U05/U05-1023.pdf>.

ZULUÁGA, A. **La fijación fraseológica**. Bogotá: Instituto Caro y Cuervo, 1975. 24 p.