



Michel de Sousa Teixeira

O JOGO DA AVALIAÇÃO
Um estudo prático sobre tradução automática

Dissertação de Mestrado

Dissertação apresentada ao Programa de Pós-graduação em Estudos da Linguagem da PUC-Rio como requisito parcial para obtenção do grau de Mestre em Letras/Estudos da Linguagem.

Orientadora: Profa. Cláudia Freitas



Michel de Sousa Teixeira

O JOGO DA AVALIAÇÃO
Um estudo prático sobre tradução automática

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Estudos da Linguagem da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Profa. Cláudia Freitas

Orientadora
Departamento de Letras – PUC-Rio

Profa. Helena Franco Martins

Departamento de Letras – PUC-Rio

Profa. Teresa Dias Carneiro

UFRJ

Profa. Monah Winograd

Coordenadora Setorial do Centro de Teologia
e Ciências Humanas – PUC-Rio

Rio de Janeiro, 29 de novembro de 2018.

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e da orientadora.

Michel de Sousa Teixeira

Tradutor desde 2005 e sócio da empresa Sacra Domus Traduções. A partir de 2013, passou a se dedicar também à tradução editorial de ficção e não ficção, com livros traduzidos para a editora Objetiva e selo Suma de Letras. Especialista em Tradução de Espanhol (latu sensu) pela Universidade Gama Filho (2008). Bacharel em Comunicação Social — Publicidade e Propaganda pela Universidade Federal do Rio de Janeiro (1998). Mestre em Letras pelo Programa de Pós-graduação Estudos da Linguagem da PUC-Rio (2018).

Ficha Catalográfica

Teixeira, Michel de Sousa

O jogo da avaliação : um estudo prático sobre tradução automática / Michel de Sousa Teixeira ; orientadora: Cláudia Freitas. – 2018.

134 f. : il. color. ; 30 cm

Dissertação (mestrado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Letras, 2018.

Inclui bibliografia

1. Letras – Teses. 2. Tradução. 3. Tradução automática. 4. Avaliação de qualidade. 5. Avaliação de tradução. 6. Jogos de linguagem. I. Freitas, Cláudia. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Letras. III. Título.

CDD: 400

Para Flávia, a melhor pessoa que já conheci na vida
Para Aurora e Tarsila, pela continuidade da vida

Agradecimentos

A minha mãe, Olga, meu padrasto, Fernando, minha sogra, Iade, e meu cunhado, João Jacques, obrigado por todo o apoio que vocês sempre me deram.

À minha orientadora Cláudia Freitas, pela parceria, pelos caminhos apontados, pela diligência e pelas aulas incríveis que me estimularam a fazer este trabalho. E também pela paciência com esse orientando enrolado.

À professora Helena Martins, pelas aulas inspiradoras e por me apresentar à filosofia de Ludwig Wittgenstein, que simplesmente mudou minha maneira de pensar e ver o mundo.

A CAPES e FAPERJ, pelas bolsas que possibilitaram a realização deste estudo. O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

À PUC-Rio, pela excelente estrutura e pelo maravilhoso ambiente de estudos.

A todo o pessoal da secretaria do Departamento de Letras, em especial à Chiquinha, pela presteza e simpatia.

A todos os professores do mestrado em Letras da PUC-Rio, e principalmente aos professores do núcleo de tradução, Marcia Martins, Maria Paula Frota, Paulo Henriques Britto e à coordenadora do PPGEL, Erica Rodrigues.

Às professoras do Curso de Formação de Tradutores da PUC-Rio, que me abriram os caminhos da tradução ainda em 2005 e me ajudaram a abraçar esse ofício que eu tanto amo, com especial carinho a Adriana Rieche, Carolina Alfaro, Lourdes Sette e Teresa Dias Carneiro.

Aos queridos colegas Daniel Estill, pelo debate no CDL em 2017 e por todos os momentos em que dedicou um tempo para ajudar este ex-aluno, Jorge Davidson, pelas trocas de ideias sempre frutíferas, e Ricardo Souza, cuja defesa das tecnologias de tradução aguçou a minha curiosidade pela tradução automática lá no início deste mestrado.

À Flávia, minha mulher, minha companheira de tradução e minha melhor amiga, cada passo do caminho fica mais fácil sabendo que você está ao meu lado.

A Aurora e Tarsila, minhas filhas, por todas as pequenas e grandes alegrias.

Resumo

Teixeira, Michel de Sousa; Freitas, Cláudia. **O jogo da avaliação: um estudo prático sobre tradução automática**. Rio de Janeiro, 2018. 134 p. Dissertação de Mestrado - Departamento de Letras, Pontifícia Universidade Católica do Rio de Janeiro.

O presente trabalho discute práticas de avaliação de tradução, notadamente no contexto de sistemas de tradução automática (TA). É apresentada uma descrição de trabalhos relevantes de avaliação de TA, que abrangem a análise de sistemas e também a criação de taxonomias para avaliação de erros de TA. Em seguida, é realizada uma série de três jogos de avaliação de qualidade de tradução automática, baseados nos jogos de linguagem descritos nas *Investigações filosóficas* de Ludwig Wittgenstein. O jogo 1 abrangeu a pós-edição da tradução automática de pedidos de patentes por um tradutor profissional em dois sistemas distintos, com posterior avaliação de produtividade e análise das edições efetuadas. Os resultados quantitativos do jogo 1 revelam o impacto positivo do uso de TA na produtividade. Para uma avaliação qualitativa dos resultados, foi desenvolvida uma taxonomia das edições feitas durante a tarefa de pós-edição, e os resultados indicaram que o aumento de produtividade da TA não se deu às expensas da qualidade final do texto. O jogo 2 envolveu a comparação de duas traduções do mesmo texto, feitas pelo mesmo tradutor em um intervalo de vários anos, a primeira com memória de tradução, a segunda com pós-edição de TA, e serviu de base para uma discussão sobre a qualidade de tradução produzida em diferentes sistemas e condições de trabalhos. O jogo 3 compara a versão final dos dois textos-alvo editados no jogo 1 aos pedidos de patentes correspondentes, depositados do Instituto Nacional de Propriedade Industrial (INPI), para avaliar se o uso de TA teve algum impacto na qualidade final dos textos-alvo. Os resultados dos jogos 2 e 3 apontaram para um nível de qualidade superior das traduções produzidas por pós-edição de TA para este estudo, quando comparadas a traduções feitas em contextos profissionais.

Palavras-chave

Tradução; tradução automática; avaliação de qualidade; avaliação de tradução; jogos de linguagem

Abstract

Teixeira, Michel de Sousa; Freitas, Cláudia (Advisor). **The Assessment Game: A Practical Study on Machine Translation**. Rio de Janeiro, 2018. 134 p. Dissertação de Mestrado - Departamento de Letras, Pontifícia Universidade Católica do Rio de Janeiro.

This work discusses translation assessment practices, most importantly machine translation (MT) assessment. Relevant literature for MT assessment is discussed, from MT systems assessment to the design of error typologies for MT assessment. The study encompasses a series of three MT quality assessment experiments based on the language games as described by Wittgenstein in his work *Philosophical Investigations*. Game 1 encompassed post-editing two patent documents using different MT systems, and further evaluating translator's productivity and analyzing the edits made. The quantitative results showed the positive impact of using MT in productivity. The qualitative results included a typology of the edits made during the PE task in both MT systems, and the results achieved showed that increased productivity did not affect translation quality level. Game 2 compares two translations of the same text produced several years apart by the same translator. The former was produced using a CAT tool; the latter, a TA system. Both served as the basis for discussing the quality of translations produced in different systems and under different work conditions. Game 3 compares the post-edited documents produced in the PE task of Game 1 to the actual patent documents as filed in Brazil's Patent Office (INPI) in order to evaluate the impact of MT over the quality of post-edited documents. The results of both games 2 and 3 pointed to a superior quality of the post-edited translations of the present study over translations produced in professional contexts.

Keywords

Translation; machine translation; quality assessment; translation assessment; language games

Sumário

1	Introdução	14
1.1.	Estrutura da dissertação	18
2	A tradução e a linguística computacional	20
2.1.	Histórico	20
2.1.1.	Fase 1: a tradução automática como objetivo	21
2.1.2.	Fase 2: foco na inteligência artificial	22
2.1.3.	Fase 3: novas abordagens para a TA	24
2.1.4.	Fase 4: TA estatística, aprendizado de máquina, avaliação	25
2.1.5.	Além da fase 4: primeira década do novo milênio	27
2.2.	Tradução assistida por computador	28
2.2.1.	Memória de tradução	29
2.2.2.	Tradução automática	31
2.2.3.	Pós-edição de TA	37
3	Perspectiva teórica	39
3.1.	A filosofia da linguagem de Wittgenstein	39
3.2.	A tradução e seus jogos de linguagem	41
3.2.1.	O jogo da tradução humana	44
3.2.2.	O jogo da tradução automática	47
3.2.3.	O jogo da avaliação de tradução	50
3.3.	Resumo	55
4	Estudos relacionados	56
4.1.	Modelos de avaliação de tradução	56
4.2.	Estudos avaliativos	61
4.3.	Resumo	75
5	Três jogos de avaliação	79
5.1.	Jogo 1	80
5.1.1.	Sistemas de TA	81
5.1.2.	Piloto	83

5.1.3. Dinâmica	83
5.1.4. Análise de produtividade	89
5.1.5. Comparações de produtividade	91
5.1.6. Avaliação qualitativa	93
5.1.7. Resultados por sistema	94
5.2. Jogo 2	97
5.2.1. Análise quantitativa	98
5.2.2. Análise qualitativa	98
5.2.3. Análise comparativa	101
5.3. Jogo 3	105
5.3.1. Lilt x PD 1: análise qualitativa	105
5.3.2. Lilt X PD 1: análise comparativa	108
5.3.3. GNMT x PD 2: análise taxonômica	112
5.3.4. GNMT x PD 2: análise comparativa	114
5.4. Resumo	117
6 Considerações finais	121
7 Referências bibliográficas	125

Lista de figuras

Figura 1:	Classificação raiz da MQM (Lommel et al., 2014)	60
Figura 2:	Taxonomia de anotação de erros (Costa et al., 2014)	70
Figura 3:	Interface de usuário do Lilt	82
Figura 4:	Interface de usuário do memoQ com resultados do GNMT	82
Figura 5:	Etapas da dinâmica do jogo 1	84
Figura 6:	Produtividade por sessão (Lilt x GNMT)	90
Figura 7:	Comparação de produtividade entre estudos	92
Figura 8:	Resultados por categoria (Lilt x GNMT)	94
Figura 9:	Sugestões do banco terminológico no memoQ	96
Figura 10:	Interface do Lilt com recurso de tradução dinâmica	97
Figura 11:	Categorização das edições feitas em T1 e T2	99
Figura 12:	Comparação entre T1 e T2	101
Figura 13:	Categorização das edições feitas em PE-Lilt e PD 1	106
Figura 14:	Comparação entre PE-Lilt e PD 1	108
Figura 15:	Diferença entre títulos (PE-Lilt e PD 1)	109
Figura 16:	Categorização das edições feitas em PE-GNMT e PD 2	113
Figura 17:	Comparação entre PE-GNMT e PD 2	115
Figura 18:	Comparação do volume de edições nos jogos 2 e 3	119

Lista de quadros

Quadro 1:	Subcategorias de <i>linguagem</i> no modelo da LISA (Almeida, 2013)	60
Quadro 2:	Características dos estudos citados neste trabalho	77
Quadro 3:	Produtividade no Lilt	89
Quadro 4:	Produtividade no GNMT	89
Quadro 5:	Comparação terminológica (T1 e T2)	103
Quadro 6:	Comparação entre trechos (T1 e T2)	104
Quadro 7:	Exemplos de correções da subcategoria <i>estilo</i> (PD 1)	107
Quadro 8:	Extratos de traduções (PD 1 e PE-Lilt)	110
Quadro 9:	Comparação terminológica (PE-GNMT e PD 2)	116

Lista de tabelas

Tabela 1:	Produtividade média por categoria (pal./h) (Guerberof, 2014)	64
Tabela 2:	Porcentagem de erros por categoria (Guerberof, 2014)	64
Tabela 3:	Média de alterações por categoria (Almeida, 2013)	66
Tabela 4:	Erros cometidos por Moses e GT (Costa et al., 2014)	71
Tabela 5	Taxonomia para anotação de erros de TA (Martins e Caseli, 2013)	72
Tabela 6:	Descrição dos jogos de avaliação	79
Tabela 7:	Tarefas de pós-edição do jogo 1	81
Tabela 8:	Taxonomia de alterações	88
Tabela 9:	Produtividade no jogo 2	98

1 Introdução

O presente trabalho se propõe a realizar uma série de três exercícios de avaliação de tradução automática (TA). A inspiração para a descrição e a análise dos exercícios são os jogos de linguagem apresentados pelo filósofo Ludwig Wittgenstein em sua obra *Investigações filosóficas*. Adotaremos, assim, uma visão não-logocêntrica da avaliação de tradução, que considera o pesquisador como um jogador no jogo da avaliação, ou seja, como alguém cuja história pessoal, cujo conhecimento e cuja ideologia são parte indissociável do processo.

O primeiro jogo de avaliação engloba uma tarefa de pós-edição (PE) de tradução automática, efetuada por um tradutor profissional, com dois objetivos distintos: i) avaliar o desempenho de dois sistemas de TA disponíveis atualmente no mercado e ii) criar uma taxonomia de avaliação de TA baseada nas alterações feitas pelo tradutor. As ferramentas escolhidas para a pós-edição foram o Lilt, uma plataforma de TA interativa online oferecida por assinatura, e a versão profissional do Google Tradutor, disponibilizada através da Cloud Translation API mediante cobrança mensal baseada no volume de palavras traduzidas, vinculada ao memoQ, uma ferramenta de tradução assistida por computador.

A tarefa de avaliação de desempenho dos sistemas se baseou em critérios como o tempo dispendido na pós-edição, para avaliação de produtividade, e as alterações consideradas necessárias para que os textos traduzidos – pedidos de patentes pertencentes ao campo de alimentos para animais – alcançassem uma qualidade de publicação, para avaliação de qualidade. As alterações foram registradas em uma tabela de desempenho e categorizadas de acordo com a sua natureza, de modo a criar uma taxonomia das edições feitas e possibilitar um melhor entendimento das virtudes e fraquezas de cada sistema de TA.

O segundo jogo de avaliação se baseou na comparação de duas traduções do mesmo texto, efetuadas pelo mesmo tradutor com um considerável intervalo de tempo entre ambas. A primeira, feita em contexto profissional, foi executada em um sistema de memória de tradução; a segunda, voltada especificamente para este

trabalho, foi feita em uma tarefa de pós-edição de TA. O objetivo do jogo 2 foi avaliar o quanto o uso de TA influenciou nas escolhas e no estilo do tradutor e o impacto que pode ter causado na qualidade final do texto traduzido.

Por fim, o terceiro jogo de avaliação comparou as pós-edições realizadas no jogo 1 deste trabalho aos pedidos de patentes efetivamente depositados no Instituto Nacional de Propriedade Industrial (INPI), responsável pela análise de pedidos de patente brasileiros. O objetivo do jogo 3 era avaliar se o estilo, a clareza e as escolhas terminológicas, entre outros aspectos, das traduções produzidas no jogo 1 estavam de acordo com a expectativa de nível de qualidade das traduções produzidas em um contexto profissional e de uso real no mercado de patentes.

O interesse pela tradução automática como objeto de estudo se justifica, em primeiro lugar, pelo papel fundamental exercido pela tecnologia na profunda transformação sofrida pela tradução nas últimas décadas, ao trocar “suas práticas artesanais por procedimentos de alta tecnologia até se tornar uma indústria bilionária” (Estill, 2017, p. 9). De 2009 a 2016, a indústria de prestação de serviços linguísticos praticamente dobrou de tamanho, passando de US\$ 23,5 bilhões para US\$ 40,27 bilhões, segundo dados da consultoria Common Sense Advisory (CSA). Em segundo lugar, pelo rápido crescimento obtido pela TA dentro dessa indústria. Cadwell et al. (2016) citam dois estudos que corroboram o crescimento da TA no mercado de tradução, como segue. Uma pesquisa realizada em 2010 pela SDL, empresa líder do mercado de ferramentas de tradução, informou que 28% dos 228 gerentes entrevistados usavam ou planejavam usar TA.

A pós-edição (PE) surgiu como alternativa para atender às demandas de um mercado em franca expansão, que requer a tradução de grandes volumes de informação com maior agilidade, sem que isso gere um impacto maior nos custos. Almeida (2013) ensina que a PE se diferencia da tradução porque o pós-editor precisa, por exemplo, trabalhar com três textos em vez de dois (fonte, alvo e texto traduzido automaticamente) e desfazer a tradução de determinados termos, como nomes de produtos. Além disso, a oferta de serviços de PE também reflete o uso pretendido para o texto traduzido. Loffler-Laurian (1995 apud Almeida, 2013), propôs as seguintes definições: PE rápida, que realiza apenas correções essenciais para o entendimento do texto, e PE convencional, cujo objetivo é produzir um texto pronto para publicação. A remuneração e a expectativa de produtividade refletem o tipo de pós-edição escolhida.

Embora a pós-edição tenha se tornado uma tarefa comum na indústria de tradução nos últimos anos, ainda não há muitas informações disponíveis sobre as dificuldades enfrentadas por tradutores que se dediquem à PE ou se o nível de experiência do profissional afeta o resultado do processo (Almeida, 2013). Outra questão relevante é se a qualidade da TA produzida afeta a capacidade do tradutor de realizar alterações relevantes no conteúdo pós-editado.

Outro dado importante é a visão que os tradutores profissionais têm do processo de pós-edição. Um artigo de Cadwell et al. (2016) cita estudos (Guerberof, 2009; Plitt e Masselot, 2010) cujos resultados atestam que a “pós-edição de TA pode ser mais rápida que a tradução sem TA e não tem, necessariamente, um impacto negativo sobre a qualidade”. Ainda assim, continua o artigo, muitos tradutores “afirmam que produziriam mais rápido e com maior qualidade se traduzissem sem o uso de TA”¹ (Cadwell et al, 2016, p. 225).

As distintas visões de tradutores sobre a TA também são abordadas em pesquisa realizada por Fulford. Segundo ele, a visão predominante dos tradutores em relação à TA era de ceticismo:

Parecia que esse ceticismo se baseava não no temor de que os sistemas de TA representassem uma ameaça ao papel do tradutor humano, mas na crença de que a tarefa tradutória seja tão complexa que uma máquina jamais seria capaz de realizá-la com eficiência. Os que mostraram maior desdém à capacidade da TA foram, talvez de forma inevitável, aqueles que tiveram menor exposição a ela. (Fulford, 2002 apud Almeida, 2013, p. 32)

Segundo o estudo de Cadwell et al. (2016), os funcionários da Direção-Geral da Tradução (DGT) da Comissão Europeia veem a tradução automática como uma ferramenta útil. Entre as razões para o uso de TA, estão: 1) ganho de produtividade, 2) percepção de que a qualidade da TA é adequada para o uso a que se destina, 3) capacidade de oferecer sugestões em que os tradutores não pensariam antes e 4) redução no volume de digitação, prevenindo lesões por esforço repetitivo.

Como se vê, tradutores profissionais alocados em uma instituição se mostraram favoráveis ao uso de TA em sua rotina de trabalho. Nos últimos anos, o uso da tradução automática vem ocupando espaços em que a tradução humana propriamente dita é inviável em termos econômicos ou de agilidade de produção,

¹ Exceto quando indicado, todas as traduções de textos citados neste trabalho são minhas.

por exemplo. Um caso de estudo é a tradução automática de conteúdo gerado por usuários do eBay. Guha e Heger (2014) descrevem o sistema de TA da empresa de comércio virtual, mostrando como pesquisas realizadas em russo por usuários da Rússia são traduzidas automaticamente para encontrar no site os anúncios relacionados aos produtos buscados. Em texto publicado a convite em um blog sobre TA, Picinini (2016) explica que a demanda de tradução do conteúdo gerado por comentários de usuários sobre produtos anunciados no eBay com a agilidade necessária para a plataforma online só pode ser atendida por um sistema de TA.

A tradução automática e a publicação sem revisão de grandes volumes de texto, o uso de tradução automática em tempo real em sites ou dispositivos móveis e em dispositivos de reconhecimento de fala, entre outros usos possíveis, são necessidades do novo mundo hiperconectado. As maiores empresas de tecnologia do mundo investem maciçamente na pesquisa e no desenvolvimento de TA como forma de facilitar a comunicação entre mercados de distintos idiomas e a ampliação das fronteiras do comércio mundial, por exemplo. Estill (2017) afirma que, neste novo contexto, “a tradução humana tornou-se cara e lenta, uma barreira [...] a ser superada pela tecnologia, que pode tornar a tradução mais rápida e mais acessível, leia-se, mais barata.” (Estill, 2017, p. 106)

Diante da perspectiva de aprimoramento constante da TA com vistas a reduzir o volume de tradução humana no futuro e das limitações atuais da tradução por máquina, qual a posição do tradutor profissional autônomo? Vimos que o uso de TA sem intervenção humana, embora já atinja grande volume em termos de dados, ainda se circunscreve a situações específicas, tais como usos não comerciais, como traduções de conteúdo da internet pelo usuário, ou usos comerciais em que uma qualidade mínima (“*good enough*”) basta.

Quando a intervenção humana é necessária e as empresas prestadoras de serviços linguísticos recorrem à pós-edição por tradutores profissionais, qual o desempenho dos sistemas de TA atuais? Essa é uma das questões que esse trabalho pretende abordar. A outra, tão importante quanto, é discutir a o papel da avaliação de tradução e, especificamente, da avaliação de TA. Esperamos que, ao abordar ambos os assuntos, este trabalho possa dar uma contribuição relevante não só para os tradutores, mas também para a academia, ao estudar uma questão premente no contexto atual da tradução e que, por sua própria novidade, merece muita atenção de estudiosos e pesquisadores.

1.1. Estrutura da dissertação

O presente trabalho está organizado da seguinte maneira. O capítulo 2 apresenta o histórico da tradução automática no contexto da linguística computacional e do processamento de linguagem natural e discute o estado atual do uso de TA na indústria de tradução de hoje. Apresenta, também, o conceito de tradução assistida por computador e os diversos sistemas computadorizados disponíveis para uso por tradutores profissionais.

O capítulo 3 apresenta a perspectiva teórica que orienta o presente trabalho, baseada nos jogos de linguagem descritos por Ludwig Wittgenstein em suas *Investigações filosóficas*. Apresentamos, em seguida, a descrição da tradução sob a ótica wittgensteiniana, tratando dos jogos da tradução humana, da tradução automática e da avaliação de traduções. Discutimos, por fim, o conceito de categorização linguística no âmbito da avaliação.

O capítulo 4 apresenta uma resenha da literatura relevante para este trabalho, que abrange, por exemplo, modelos de categorizações linguísticas para avaliação de tradução e estudos que avaliam a qualidade de sistemas de TA e o uso de pós-edição de TA por tradutores. As taxonomias propostas e os resultados obtidos na literatura serão usados como base de comparação e discussão dos jogos de avaliação de tradução realizados para a presente dissertação e descritos no capítulo 5.

O jogo 1 inclui uma tarefa de pós-edição de pedidos de patentes na área de alimentos para animais e uma posterior categorização das alterações feitas na saída da TA durante a tarefa em uma taxonomia de pós-edições útil para a avaliação de qualidade de traduções. O jogo 2 traz a comparação entre duas traduções do mesmo texto efetuadas pelo mesmo tradutor em diferentes momentos, com um grande intervalo de tempo entre ambas, para avaliar a qualidade de traduções produzidas em diferentes sistemas e condições de trabalho. O jogo 3 traz a comparação entre as traduções produzidas na tarefa de pós-edição de TA e os pedidos de patente efetivamente depositados no INPI, para avaliação de qualidade.

O capítulo 6 apresenta as conclusões a que chegamos após os jogos de avaliação realizados e discute as novas abordagens para o conceito de qualidade na indústria de tradução de hoje. Por fim, discutimos as contribuições do presente trabalho para tradutores profissionais, pesquisadores e professores de tradução.

2

A tradução e a linguística computacional

Neste capítulo, abordamos a relação entre a tradução e a linguística computacional (LC), a começar pela presença, desde o início, da tradução automática como um dos objetivos do uso de computadores para processamento de linguagem natural. Apresentaremos as diversas fases de evolução da linguística computacional e o papel da TA em cada um desses períodos. Discutiremos o impacto da globalização no crescimento do mercado de tradução ao longo das últimas décadas, ancorado no desenvolvimento de sistemas de tradução assistida por computador, e traremos um panorama da pesquisa e do desenvolvimento de novas tecnologias de tradução automática nos últimos anos.

2.1. Histórico

A ciência conhecida como linguística computacional, que trata do estudo da linguagem por meio de processamento de dados, surgiu no final da década de 1940, a reboque dos primeiros computadores. Desde então, os estudos em linguística computacional ensejaram novas abordagens para o estudo da língua e impulsionaram o desenvolvimento de vários sistemas para processamento da linguagem natural (PLN). A seguir, traçamos um breve histórico da área, perpassando suas várias fases, tomando por base a divisão proposta por Sparck Jones (2001): fase 1, do fim da década de 1940 ao fim da década de 1960; fase 2, até o fim da década de 1970; fase 3, até o fim da década de 1980; fase 4, anos 1990. O foco de nossa abordagem será a tradução automática no contexto da linguística computacional e, especificamente, o estado atual da TA em vista do grande desenvolvimento tecnológico das últimas duas décadas.

2.1.1. Fase 1: a tradução automática como objetivo

Sparck Jones e outros autores consideram como “certidão de nascimento” da linguística computacional o memorando *Translation*, publicado pelo matemático Warren Weaver em 1949. Nesse texto, Weaver analisa diversos aspectos relacionados à possibilidade de fazer “tradução por meio do uso de computadores eletrônicos de grande capacidade, flexibilidade e velocidade” (Weaver, 1949, p. 1), discutindo estratégias linguísticas, estatísticas e computacionais para a tradução automática de textos.

Ao discorrer sobre o período inicial das pesquisas em processamento da linguagem natural, Kay (2005) reconhece duas grandes motivações dos estudiosos: uma teórica, que acreditava que os avanços em sistemas de computação teriam como frutos importantes avanços na teoria linguística, e outra, prática, cujo objetivo era criar uma tecnologia baseada em preceitos científicos que pudesse atender às necessidades específicas de processos de tradução, extração de informações, sumarização, verificação gramatical, entre outros. A principal linha de trabalho da primeira fase da linguística computacional começa com a tradução como pesquisa, com processamento palavra-por-palavra baseado em dicionários e tentativas de resolver ambiguidades semânticas e sintáticas, entre outras tarefas. (Sparck Jones, 2001).

Entretanto, havia um descompasso entre a magnitude dos desafios práticos enfrentados pelos pioneiros da TA, que usavam uma nova ferramenta, os computadores, com o objetivo de processar dados não-numéricos quando o próprio processamento de dados ainda não estava bem estabelecido (idem), e as expectativas geradas para o processamento de linguagem natural. O pressuposto geral era de que havia uma lógica inerente à linguagem humana, considerada um sistema formal passível de ser processado automaticamente. Faltava, apenas, na visão corrente à época, “traduzir” matematicamente o processo da linguagem.

Nos Estados Unidos, tanto o governo quanto empresas privadas investiram muitos recursos no desenvolvimento da tradução automática. Os cientistas envolvidos no projeto IBM-Georgetown, de 1954, que traduziu automaticamente mais de sessenta sentenças do russo para o inglês, acreditavam que o problema da tradução automática estaria solucionado em três ou cinco anos (Hutchins, 2005). A

impossibilidade de cumprir metas como essa acabou por criar uma sensação generalizada de que a tradução automática não estava à altura dos desafios apresentados pela tecnologia (Kay, 2005). Insatisfação refletida no relatório de 1966 da ALPAC – Comitê consultivo em processamento automático de linguagem do governo dos Estados Unidos –, que concluiu que a TA ainda estava muito distante de ser uma possibilidade real, embora recomendasse a manutenção do apoio à linguística computacional e a ferramentas computacionais de auxílio ao tradutor, como dicionários automáticos. Embora tenham havido muitas críticas à parcialidade e à falta de profundidade do relatório, o resultado prático foi a suspensão quase total dos investimentos em pesquisa de TA nos EUA ao longo dos dez anos seguintes.

2.1.2. Fase 2: foco na inteligência artificial

Com o corte de fundos, teve início uma nova fase nas pesquisas em linguística computacional, que se voltou para sistemas de inteligência artificial (IA) com “muito mais ênfase no conhecimento do mundo e seu papel na construção e manipulação da representação de sentido” (Sparck Jones, 2001, p. 4). Logo, o interesse maior era a construção e manipulação de grandes bases de conhecimento para compreender como a linguagem representa esse conhecimento. A ideia de um conhecimento puro ensejou a expectativa de se criar sistemas gerais de processamento da linguagem, que pudessem se adaptar com sucesso a distintos idiomas, campos de conhecimento ou demandas.

Podemos citar como exemplos deste posicionamento teórico as tentativas de programar as gramáticas transformacionais de Chomsky para análise de sentenças (Wilks, 2005). Esse tipo de sistema se baseava em ferramentas de conhecimento profundo, como gramáticas internas, analisadores linguísticos ou processadores lógicos para fazer o processamento da linguagem natural. Entretanto, apesar de todos os esforços de modelagem lógica da linguagem, os sistemas conseguiam analisar muito pouco além das sentenças para as quais haviam sido projetados (idem).

Aos sistemas de conhecimento profundo se contrapunham os chamados “toy systems”, que faziam a correspondência a partir de padrões superficiais, sem

qualquer tipo de análise linguística. O objetivo era modelar situações específicas, com base em padrões e em uma linguagem procedural que permitisse a interação entre o sistema e seres humanos. Muitos estudiosos se debruçaram sobre problemas relacionados à construção de sistemas interacionais de perguntas e respostas que não se baseavam em ferramentas de conhecimento profundo, como gramáticas internas, analisadores linguísticos ou processadores lógicos, mas em padrões e em uma linguagem procedural que permitia responder a questões.

Esses sistemas seguem um padrão semelhante ao apresentado por Turing (1950), em artigo que se inicia com a pergunta: “máquinas conseguem pensar?” Para responder a essa pergunta, Turing propõe realizar o que chamou de “jogo da imitação”, um jogo de perguntas e respostas que opõe um ser humano a uma máquina em um jogo de perguntas e respostas. A pessoa digitaria as perguntas — na visão de Turing, o meio para digitação e entrega das questões seria uma máquina de escrever cujo conteúdo pudesse ser “teleimpresso” em uma outra sala — e receberia as respostas por escrito, sem saber quem está do outro lado. Se, após algum tempo, a pessoa não percebesse que seu interlocutor era uma máquina, poder-se-ia dizer que a máquina é inteligente. Turing disse acreditar que, em cinquenta anos, a pessoa não teria mais de 70% de chance de identificar o interlocutor como uma máquina depois de cinco minutos de interação. Nem seria preciso esperar tanto. Menos de trinta anos depois da publicação do artigo já surgiriam exemplos bem-sucedidos de Testes de Turing (como esse tipo de sistema ficou conhecido).

Um sistema chamado PARRY (Colby, 1973 apud Wilks, 2005), por exemplo, emulava um paciente paranoico de um hospital para veteranos de guerra nos EUA em diálogos com usuários com acesso à então incipiente internet. O sistema se baseava em 6 mil padrões de perguntas e respostas. O sistema era “muito robusto, parecia se lembrar do que era dito e reagia violentamente quando parâmetros internos como FEAR (medo) e ANGER (raiva) ficavam altos”, além de demonstrar “interesse e capacidade de conversação”. Muitos dos que trocaram mensagens com o “veterano” não conseguiram acreditar que estavam falando com uma máquina. Wilks (2005) ainda destaca que esse tipo de sistema, que faz a correspondência a partir de padrões superficiais, sem qualquer tipo de análise linguística, viria a se tornar a tecnologia hoje conhecida como extração de informações, que extrai automaticamente informações estruturadas de documentos não-estruturados ou apenas parcialmente estruturados.

Embora os EUA tenham se afastado da pesquisa em TA e levado a União Soviética e alguns países da Europa a reduzir esforços no mesmo campo, outras nações, como Canadá, França e Alemanha, continuaram as pesquisas em tradução automática nos anos seguintes. O sistema de TA Systran foi adotado pela Força Aérea americana em 1970 e pela Comissão de Comunidades Europeias em 1976, devido ao grande volume de documentação produzido por esta entidade supranacional. Neste mesmo ano, a Universidade de Montreal, no Canadá, produziu um bem-sucedido sistema de tradução automática de relatórios meteorológicos denominado Météo.

Se, no início, a principal motivação para o desenvolvimento da TA era de natureza militar e técnico-científica — notadamente a tradução de documentos do russo para o inglês e vice-versa, como parte da disputa entre EUA e URSS pelo domínio mundial durante a Guerra Fria —, a partir de meados da década de 1970 a demanda por TA se ampliou para atender outros idiomas, com diferentes motivações. As necessidades administrativas de entidades como a Comissão de Comunidades Europeias e o governo do Canadá, um país oficialmente bilíngue, por um lado, e o incremento nas transações comerciais entre os países da Europa e o Japão e outras nações do Oriente, por outro, estimularam a pesquisa e o desenvolvimento de sistemas de tradução assistida por computador para suprir um volume de serviço que a oferta de tradutores humanos começava a não ser capaz de atender, seja em termos de tempo ou de custos.

2.1.3.

Fase 3: novas abordagens para a TA

No início da década de 1980, já havia ficado claro que era “muito mais difícil projetar sistemas robustos para processamento de linguagem natural [...] do que se acreditava” (Sparck Jones, 2001, p. 6). Uma nova tendência, de base gramatical e lógica, levou ao desenvolvimento de vários tipos de gramáticas funcionais e categoriais, orientadas pelo princípio de computabilidade, que seriam usadas como base em algoritmos de análise.

Sparck Jones identifica também uma renovação nos trabalhos dedicados à tradução automática no período, especialmente na Europa e no Japão, geralmente com ênfase em aplicações específicas. Hutchins (2005) destaca o aumento na oferta

de tipos de sistemas de TA. Além de sistemas de grande porte, como o Systran, que passou a ser oferecido em mais pares de idiomas, o Logos (alemão-inglês e inglês-francês), o Metal (alemão-inglês), desenvolvido pela Siemens, os sistemas internos da Organização Pan-Americana de Saúde (inglês-espanhol e vice-versa) e aqueles produzidos por empresas de informática japonesas para a combinação inglês-japonês e vice-versa, surgiram sistemas de menor porte, oferecidos na América do Norte, na Europa e no Japão por empresas como ALPS, Globalink, Sharp e Mitsubishi, e programas baseados em microcomputadores, produzidos na China, na Coreia e na União Soviética, por exemplo.

A década de 1980 também testemunhou pesquisas que buscavam produzir estratégias mais avançadas para produção de TA. A técnica dominante durante a maior parte do período foi o processo de tradução indireta, que usava uma representação intermediária, entre a decodificação do texto-fonte e a recodificação do texto-alvo. Em geral, essa representação intermediária tinha como base uma interlíngua, e se valia de análises sintáticas e morfológicas, bem como de bases de conhecimento não linguísticas, para se chegar ao resultado final. Hutchins (2005) lista os projetos mais importantes para o desenvolvimento da TA à época, tais como SUSY (Saarbrücken), Mu (Kyoto), Rosetta (Eindhoven), Carnegie-Mellon University (Pittsburgh) e dois projetos multilíngues internacionais: Eurotra, financiado pelas Comunidades Europeias, e o projeto japonês CICC, com participantes na China, na Indonésia e na Tailândia.

O incremento do volume de dados disponíveis e a capacidade de processamento dos computadores na segunda metade dos anos 1980 também permitiu a compilação de corpora anotados manualmente, como o Penn Tree Bank, e a criação de sistemas de anotação computadorizada. “Um curioso efeito de tudo isso foi trazer os programas, antes considerados como algo sem conexão direta com textos, para o espaço de objetos que são, em si, textos anotados” (Wilks, 2005, p. 8).

2.1.4.

Fase 4: TA estatística, aprendizado de máquina, avaliação

Para Sparck Jones (2001), o ramo mais impressionante da linguística computacional na década de 1990 foi o processamento estatístico da linguagem. A

união entre a análise de dados para classificação semântica e a aplicação direta de métodos estatísticos ao processamento de linguagem natural para análise probabilística permitiu o aprendizado de máquina a partir dados extraídos de corpora. Dessa forma, o computador não se baseia em uma estrutura modelada para resolver um problema específico, mas extrai informações de dados para lidar com situações imprevistas, aprendendo por meio de exemplos e tentativa e erro. Essa abordagem se tornou a base para o desenvolvimento de novos sistemas de tradução automática estatística (como, por exemplo, sistemas da TA baseados em exemplos) com vistas a solucionar o problema de traduzir “materiais ricos, variados e imprevisíveis, como textos de notícias”, para os quais “abordagens absolutistas são manifestamente inadequadas” (Sparck Jones, 2001, p. 8).

Hutchins (2005a) considera o período um momento decisivo para o desenvolvimento de sistemas de tradução automática e cita o sistema Candide, produzido por Jelinek e equipe na IBM, como um dos marcos da virada em direção ao desenvolvimento de sistemas estatísticos de TA, em detrimento dos antigos sistemas baseados em regras gramaticais e semânticas para a análise dos textos a serem traduzidos automaticamente. Outra inovação do Candide, descrita por Wilks (2005), foi o uso de aprendizado de máquina. A equipe usou corpora paralelos, como o Hansard canadense em inglês e francês, para emular a habilidade humana de tradução. Neste caso, o material de treinamento para a máquina não foi produzido especificamente para a tarefa; pelo contrário, são textos de linguagem natural. O objetivo era que o sistema tivesse “capacidade para aprender sem ser especificamente programado” para isso. A crescente capacidade de processamento de dados possibilitou uma ampla gama de abordagens para o aprendizado de máquina, tais como o aprendizado baseado em exemplos, encontrado no trabalho desenvolvido por diversos grupos de pesquisadores no Japão, que se tornaria a base para sistemas baseados em exemplos.

O processo básico de aprendizado baseado em estatística permitiu que se alimentasse o sistema com exemplos de uso para reconhecimento de padrões que o programa pudesse reconhecer ao realizar a operação de tradução. Esses exemplos são, basicamente, obtidos de corpora paralelos, que podem ser ou não traduções. No caso do aprendizado baseado em regras, o programa é alimentado com uma série de regras para dar conta de ambiguidades lexicais e estruturais, seja em uma língua

específica ou entre línguas (Hutchins, 2005a). Essas regras podem ser de natureza lexical, gramatical ou relativas à tradução.

A fase 4 se caracterizou também pelo interesse no desenvolvimento de recursos para processar grandes volumes de texto disponíveis na rede mundial de computadores e, também, pelo desenvolvimento da avaliação de sistemas em termos, por exemplo, de robustez e facilidade de personalização. O interesse em avaliação abrange a avaliação de tecnologias e sistemas, dos dados produzidos por esses sistemas, da interface com o usuário e, no caso de sistemas de TA, o desenvolvimento de critérios qualitativos e quantitativos de avaliação e métricas de análise.

2.1.5.

Além da fase 4: primeira década do novo milênio

A classificação histórica de Sparck Jones vai até o fim dos anos 1990. Os trabalhos de Hutchins (2005 e 2005a) e Wilks (2005), citados ao longo deste capítulo, também só dão conta do histórico da TA até o fim do milênio passado ou o início dos anos 2000. À época, já estavam em curso diversos esforços para o aprimoramento da tradução automática, que efetivamente se tornariam, no caso dos sistemas baseados em estatística, a abordagem padrão de pesquisa e desenvolvimento de TA. No entanto, nesta nova fase dos estudos de linguística computacional, a TA acabou relegada e uma posição de coadjuvante, cedendo espaço a pesquisas sobre inteligência artificial, processamento da linguagem natural, humanidades digitais e outros temas. Os sistemas disponíveis ainda dependiam de grande investimento para superar suas muitas limitações e a TA era oferecida rotineiramente apenas por motores de busca. Na indústria de tradução, a pesquisa e o desenvolvimento se voltaram para sistemas de tradução assistida por computador, ou, mais especificamente, sistemas de memória de tradução, que se consolidariam então como ferramentas de uso cotidiano por tradutores profissionais.

O conceito de memória de tradução surgiu na década de 1970, mas os primeiros sistemas só foram implementados de fato na década de 1980, geralmente como soluções criadas especificamente para atender às demandas de clientes de agências de tradução (Melby, 1995 apud Bowker, 2005). Assim, muitas ferramentas

que se estabeleceram no mercado surgiram para atender a empresas, não a tradutores individuais, um panorama que só mudaria em meados dos anos 1990, com o grande impulso da globalização.

A consolidação da rede mundial de computadores forneceu as bases para a criação de uma economia global e digital, baseada na informação, que abriu uma nova janela para a oferta de produtos e serviços das grandes corporações em todos os países do mundo e transformou a tradução em um produto de massa, notadamente no mundo dos negócios (Champollion, 2017). Segundo estimativa do Common Sense Advisory, a indústria de serviços terceirizados de linguagem cresceu mais de trinta por cento em um ano, passando de 9 bilhões de dólares em 2006 para 12 bilhões em 2007 (Beninato e De Palma, 2007 apud Drugan, 2013). O aumento significativo na demanda por serviços de tradução gerou uma maior pressão sobre os tradutores, com a exigência de que trabalhassem mais rápido, mas sem perder a qualidade (Bowker, 2002). Como consequência, o desenvolvimento de sistemas de tradução assistida por computador ganhou novo fôlego e passou a atender não só a empresas, mas também tradutores autônomos.

2.2. Tradução assistida por computador

Bowker (2002) faz a distinção entre duas abordagens para os sistemas computadorizados de tradução: a primeira, chamada de tradução de máquina assistida por humanos (TMAH), se refere a sistemas que efetuam a tradução automaticamente e podem sofrer intervenção humana antes (pré-edição) ou depois (pós-edição) do processo automatizado de tradução, mas não durante; a segunda, denominada tradução humana assistida por máquina (THAM), ou simplesmente tradução assistida por computador (TAC), trata de sistemas informáticos usados pelo tradutor em seu ofício, mas que não fazem a tradução propriamente dita. Freigang (2005), por sua vez, menciona a classificação proposta por Lehrberger e Bourbeau, que além de TMAH e THAM, ainda reconhece a tradução por máquina totalmente automática (TMTA), em que não há qualquer intervenção humana no processo.

A grande diferença entre tradução automática e tradução assistida por computador está em quem é o responsável pela tarefa de traduzir: no primeiro caso,

o responsável é o computador, no segundo, o tradutor humano. Enquanto os sistemas de TA tentam substituir os tradutores, as ferramentas de tradução assistida por computador os auxiliam, ajudando-os a trabalhar de forma mais eficiente (Bowker, 2002).

Assim, o termo tradução assistida por computador se refere a programas que “não foram projetados para realizar a análise sintática ou semântica de um texto-fonte, nem gerar um texto equivalente ao texto-fonte ou qualquer parte deste na língua-alvo” (Freigang, 2005). O autor também exclui da definição as ferramentas que não foram especificamente projetadas para uso por tradutores, tais como processadores de texto ou ferramentas de armazenamento e gerenciamento de arquivos. A tradução assistida por computador ocorre quando um texto “é processado em ferramentas computadorizadas para produzir uma tradução e o tradutor controla todos os estágios do processo e desempenha o processo intelectual da tradução”.

2.2.1. Memória de tradução

Via de regra, as ferramentas necessárias para este processo estão conjuntamente disponíveis em um sistema de memória de tradução (MT), que permite a realização das três tarefas do tradutor, conforme a classificação de Freigang (2005): edição, gestão terminológica e tradução propriamente dita. Para os usos deste trabalho, o tradutor realiza uma edição quando altera uma sugestão de tradução existente no sistema, seja proveniente de uma correspondência da memória de tradução ou de uma ferramenta de TA. A gestão terminológica é a inclusão novos termos ou a edição de termos existentes em um glossário, executada em uma ferramenta específica dentro do sistema de memória de tradução. A tradução propriamente dita ocorre quando não há correspondência (“*match*”) na memória (nem tradução automática) e o tradutor produz o texto traduzido “do zero”, ou usando apenas os termos sugeridos pelo glossário.

Para entender melhor o que é um sistema de memória de tradução, é preciso definir suas partes e analisar como funcionam. A memória de tradução é um corpus paralelo bilíngue que serve como banco de dados linguísticos. Os textos armazenados em uma memória são segmentados, ou seja, divididos em segmentos

bilíngues, que geralmente correspondem a frases, na maioria dos casos de uso, ou em parágrafos, e contêm o texto-alvo e o texto-fonte alinhados. À medida que são traduzidos no sistema, os segmentos bilíngues são armazenados no banco de dados. O sistema compara automaticamente os segmentos do texto-fonte com os segmentos armazenados na memória e, quando ocorre uma correspondência total ou parcial, o segmento da MT é oferecido como “sugestão” de tradução, que o tradutor pode aceitar ou não.

Existem distintas gradações de correspondência entre segmentos, como registrado por Bowker (2001):

- Correspondência exata: ocorre quando o conteúdo do segmento a ser traduzido é idêntico ao conteúdo de um segmento armazenado na MT. Neste caso, o sistema geralmente exibe a cor verde em algum ponto da interface de usuário, para facilitar a identificação.
- Correspondência completa: ocorre quando elementos invariáveis, como números, datas e *tags*, são a única diferença entre o novo segmento e o segmento da MT. Geralmente, também aparecem em verde. Muitos sistemas de MT disponíveis hoje alteram automaticamente os números contidos em segmentos com correspondência completa.
- Correspondência parcial: extrai da memória segmentos que são semelhantes, mas não idênticos. O sistema oferece a sugestão para que o tradutor a edite conforme a necessidade. Geralmente, aparecem destacadas na interface de usuário com a cor amarela. A porcentagem de correspondência parcial pode ser definida pelo tradutor conforme a necessidade: o padrão costuma ser de 75%. Uma porcentagem de correspondência abaixo de 50% vai gerar muitos resultados pouco aproveitáveis, e acima de 90% vai gerar poucos resultados e obrigar o tradutor a traduzir muitos segmentos do zero.
- Correspondência de termos: mesmo quando não há correspondência parcial, o sistema exibe para o tradutor as traduções dos termos contidos no glossário ou banco de termos.
- Correspondência de subsegmento: o sistema exibe trechos de correspondência parcial contidos dentro de um segmento quando este, como um todo, não exibe correspondência parcial.

Uma das principais vantagens de um sistema de memória de tradução é a economia de tempo, embora esta dependa do tamanho e da qualidade da memória de tradução. Uma MT nova, sem segmentos armazenados, vai demorar a produzir resultados de correspondência. Uma MT robusta, mas de baixa qualidade, produz correspondências que podem induzir o tradutor ao erro ou obrigá-lo a perder tempo reescrevendo segmentos ou realizando pesquisas para confirmar a tradução de determinados termos. Uma MT robusta e de qualidade, por sua vez, pode promover um ganho considerável de produtividade para o tradutor.

Outra vantagem é facilitar a manutenção da consistência da tradução com o uso de recursos como a inserção de termos do glossário no segmento durante a tradução ou a busca por determinado termo na memória de tradução. Em projetos divididos entre um grupo de tradutores, memórias e glossários podem ser compartilhados online, garantindo a consistência do projeto como um todo.

Bowker (2001), no entanto, alerta para os problemas causados pelo uso de sistemas de MT, como, por exemplo, a rigidez da segmentação, que pode levar o tradutor a tratar o texto frase a frase, em vez de considerá-lo como um todo, o que pode causar problemas de coesão e prejudicar a leitura. Outro fator listado pela autora é que sistemas de tradução assistida por computador são menos flexíveis do que processadores de textos, exigindo mais esforço para unir ou dividir segmentos, por exemplo, ou para editar segmentos já armazenados na MT, o que pode levar o tradutor a fazer menos revisões antes de entregar a tradução final. Outro problema, mencionado em Guerberof (2014), é o fato de que, ao trabalhar com memória de tradução, o tradutor tende a confiar mais na qualidade do texto traduzido e muitas vezes pode deixar passar erros de tradução por não fazer uma pesquisa mais profunda de terminologia, por exemplo.

2.2.2. Tradução automática

Os primeiros projetos de desenvolvimento de tradução automática, ainda na década de 1950, tinham como objetivo nada menos que “produzir uma tradução de alta qualidade, de modo completamente automático e para domínios irrestritos” (Hutchins, 2007 apud Caseli, 2015). Desde então, diversas estratégias foram adotadas para se desenvolver um sistema que atingisse esse objetivo.

Caseli (2015) explica que, em termos de estratégia, os sistemas iniciais efetuavam a operação de tradução automática de três formas distintas: direta, em que o mapeamento de unidades lexicais-fonte para unidades lexicais-alvo era feito diretamente, sem outros níveis de análise; por transferência, com base em regras de transferência sintática usadas para construir a representação equivalente na língua-alvo e, posteriormente, gerar as unidades lexicais-alvo, e por interlíngua, em que ocorre o mapeamento completo da língua-fonte e uma posterior representação abstrata do significado em uma etapa intermediária, que serve de base para a geração da tradução na língua-alvo. Em termos de arquitetura, os sistemas iniciais se baseavam em regras lexicais, sintáticas e semânticas, que eram usadas para realizar a transferência lexical e morfológica entre os idiomas. Conhecidos pela sigla RBMT (Rule-Based Machine Translation, ou tradução automática baseada em regras), esses sistemas foram o paradigma dominante da TA até meados da década de 1980, quando surgiram os sistemas empíricos ou baseados em corpus.

Nos sistemas baseados em corpus, o treinamento de máquina não é mais feito por regras, mas a partir da inclusão de um corpus paralelo cujo conteúdo bilíngue vai servir de exemplo para a geração do texto-alvo. Esses sistemas se dividem em duas abordagens distintas: EBMT (Example-Based Machine Translation, ou tradução automática baseada em exemplos) e SMT (Statistical-Based Machine Translation, ou tradução automática baseada em estatística).

O sistema EBMT busca reconhecer padrões recorrentes no corpus de treinamento, ou seja, em exemplos reais de sentenças-fonte, para produzir as sentenças-alvo por comparação e inferência. Nesse tipo de sistema, a sentença-fonte é decomposta em fragmentos, que são traduzidos separadamente e depois reagrupados para formar a sentença-alvo. Tais fragmentos se baseiam nos exemplos reais disponíveis no corpus de treinamento. Por trabalhar com dados, os sistemas EBMT não exigem o entendimento e a aplicação de regras linguísticas. O conhecimento linguístico é enriquecido por meio da adição de mais exemplos. Contudo, por se basear em exemplos, o sistema EBMT exige o processamento de um grande volume de dados de corpora para funcionar a contento, o que pode, por exemplo, impactar sua capacidade de resposta durante a realização de operações computacionais de armazenamento e de busca por exemplos similares.

Os sistemas SMT geram a tradução com base em probabilidades, que podem ser calculadas considerando-se apenas palavras ou sequências de palavras. Entre as

vantagens dos sistemas estatísticos está o baixo custo de geração e a facilidade de uso, considerando-se a disponibilidade de diversos kits de ferramentas de TA de código aberto, como o Moses (Koehn et al., 2007). O Moses oferece os componentes necessários para o pré-processamento de dados e o treinamento dos modelos de idiomas e de tradução. Diversos estudos acadêmicos desenvolveram sistemas de TA baseados em Moses para avaliação.

Outra vantagem dos sistemas SMT é que independem de língua, o que permite que padrões recorrentes sejam detectados para a construção dos modelos de língua e de tradução. Os sistemas SMT modelam o processo de tradução como um problema de aprendizado de máquina, no qual tentam reproduzir a capacidade humana de deduzir o significado aproximado de uma palavra por meio do contexto em que ela ocorre. A desvantagem dos sistemas SMT é a incapacidade de considerar aspectos estruturais e sintáticos da língua. No entanto, mesmo com essa limitação, a SMT se manteve como estado da arte durante mais de dez anos, até ser substituída recentemente pelo modelo NMT (Neural Machine Translation, ou tradução automática neural).

A NMT é uma nova abordagem para os sistemas de TA estatísticos, que modela o processo de tradução automática como um todo por meio de uma única rede neural artificial (Luong et al., 2016). Os sistemas NMT usam representações vetoriais para palavras e estados internos e um modelo sequencial único que prevê uma palavra de cada vez. Essa previsão está condicionada à sentença-fonte inteira e também à sentença-alvo inteira, o que permite que o sistema tenha acesso a um contexto mais amplo, que parece impactar positivamente a escolha de palavras, a morfologia e até trechos de sentenças (Koehn, sem data). Outras vantagens dos sistemas NMT são: o baixo uso de memória computacional, quando comparados aos sistemas estatísticos, e o fato de que todos os componentes do modelo de tradução neural são treinados em conjunto para maximizar o desempenho da tradução (Cho et al., 2014). A possibilidade de generalizar dados permite, por exemplo, “incrementar o conhecimento do comportamento de tradução para ‘carros’ a partir de exemplos que contenham ‘carro’ ou ‘automóveis’” (Koehn, sem data).

Por outro lado, alerta Koehn, os sistemas neurais têm vocabulário restrito devido a limitações computacionais, o que dificulta implementações de NMT com vocabulários especializados de grande volume e o funcionamento dos sistemas com

conjuntos de dados muito pequenos. Picinini (2018), por sua vez, afirma que a NMT muitas vezes produz sentenças-alvo de qualidade similar à humana que, contudo, não refletem o conteúdo da sentença-fonte, como, por exemplo, a omissão de partículas negativas ou a troca de um número ou apenas um termo da sentença-fonte por outro na sentença-alvo. Essa afirmação é corroborada por Koehn, que diz que o sistema NMT pode produzir palavras de saída que parecem bem no contexto, mas não tem relação com a sentença-alvo.

A NMT vem se mostrando muito competitiva quando comparada aos mais avançados modelos de SMT. Bentivogli et al. (2016) compararam o desempenho de um sistema de NMT a distintos sistemas SMT e o primeiro superou os últimos em todas as métricas de avaliação. Way (2018) afirma também que, do ponto de vista linguístico, a NMT produziu “menos erros morfológicos (-19%), lexicais (-17%) e consideravelmente menos erros de ordem de palavras (-50%)”. Por isso, apesar dos problemas enfrentados, a NMT é, hoje, o paradigma dominante na pesquisa e no desenvolvimento de sistemas de TA.

Os dois sistemas de tradução automática usados nos jogos de pós-edição deste trabalho, Lilt e Google Tradutor, usam tecnologia NMT. A plataforma do Lilt integra tradução automática, memória de tradução e bancos de dados terminológicos em uma interface que pode ser acessada na plataforma online ou por meio de uma API que pode, por exemplo, integrar o sistema a uma ferramenta de memória de tradução.

O Lilt é apresentado como uma inovação em relação ao processo tradicional de pós-edição de traduções automáticas por permitir edições em tempo real que alimentam o aprendizado de máquina. Essa tecnologia, chamada de tradução automática adaptativa, aprende e ajusta os resultados da saída de TA de acordo com as edições feitas por um tradutor humano e/ou por meio dos dados disponíveis na memória de tradução². O sistema também oferece sugestões de TA a nível de palavra, à medida que o tradutor vai traduzindo, ou seja, simultaneamente ao processo de tradução. Esse recurso está alinhado com a ideia de uso da TA como ferramenta para tradução humana assistida por computador, conforme apresentado em Luong et al. (2016).

² Disponível em: <<https://lilt.com/kb/memory/mt>>. Acesso em: 2 ago. 2018.

O sistema de tradução automática com tecnologia neural do Google (GNMT) foi divulgado ao público em setembro de 2016. Wu et al. (2016) explicam que o sistema não usa mais o inglês como etapa intermediária (língua-ponte) no processo de tradução, como ocorria antes. Segundo Johnson et al. (2017), a inclusão de um *token* artificial no início da sentença de entrada para especificar a língua-fonte permite o uso de um único modelo de TA neural multilíngue sem mudanças consideráveis de arquitetura, ou seja, sem exigir a inclusão de novos parâmetros, o que reduz o volume de processamento de dados.

O artigo apresenta uma avaliação automática realizada com a métrica BLEU (Papineni et al., 2002) no par espanhol-português para demonstrar que o novo sistema de TA neural é superior ao antigo sistema de TA baseada em frases e ao sistema neural anterior, que usava o inglês como ponte entre os dois idiomas. A versão do GNMT voltada para uso profissional está disponível via Google Cloud Translation API, que conecta o sistema online da empresa a outro software.

Em termos de uso de sistemas de TA, pode-se falar em sistemas totalmente automáticos, sistemas que permitem pré e pós-edição do conteúdo e sistemas interativos. O uso de TA sem manipulação humana é útil, por exemplo, quando o usuário deseja obter apenas uma ideia geral do conteúdo de um texto, sem grandes preocupações em relação ao estilo ou à correção gramatical, ou, como descreve Freigang (2005), quando o texto a ser traduzido se encaixa em uma sublíngua, isto é, em um domínio restrito de linguagem, permitindo maior controle sobre as configurações do sistema de TA. Exemplos do segundo caso são o já mencionado sistema Météo canadense, que traduz 45 mil palavras de conteúdo relacionado a boletins meteorológicos por dia do inglês para o francês, e o sistema suíço de alertas de avalanche.

Way (2013) descreve diversos outros casos de uso de TA não editada, e menciona que, em muitos deles, essa é a única possibilidade de tradução. Um exemplo desse tipo de caso de uso são as descrições de produtos do eBay e os comentários de usuários sobre produtos adquiridos no site de comércio online, mencionado anteriormente nesta dissertação. Comentários de usuários são o tipo de conteúdo cuja tradução não precisa ser perfeita, apenas minimamente compreensível para um comprador que não fale o idioma original no anúncio: o interesse aqui é saber o que outros compradores acharam do produto. Além disso, seria economicamente inviável para o site traduzir o volume de conteúdo gerado

por usuários todos os dias. Esse conteúdo é extremamente perecível, ou seja, vai perder relevância em pouco tempo, sendo substituído por outros comentários mais recentes, ou até mesmo caducar após a venda do produto.

Outros usos possíveis para a TA sem edição são, por exemplo, a tradução de e-mails e comunicados internos para os diversos escritórios internacionais de uma empresa de alcance global, de atualizações críticas em sites da internet, de relatórios do mercado financeiro (quando só for necessário ter uma ideia básica do conteúdo) e de respostas a questionários. Em todos esses casos de uso, a TA representa uma economia significativa de tempo, dinheiro ou ambos, e não há a necessidade de geração de conteúdo perfeito, mas apenas compreensível na língua-alvo.

Para usos que demandam maior qualidade do texto traduzido, a intervenção humana se faz necessária para melhorar a qualidade da saída de TA. A intervenção pode ser prévia, com o usuário alimentando o sistema com memórias de tradução, glossários, dicionários e outros recursos para ajustar o sistema, ou por meio da criação de regras para funcionamento do sistema para o tipo de domínio textual, por exemplo, e também com o uso de linguagem controlada, ou seja, de estratégias de redação que evitem ambiguidades ou expressões idiomáticas. Um exemplo desta última estratégia é o artigo de Barreiro e Mota (2017), que relata a criação de um sistema de parafraseamento, denominado eSPERTo, que pode ser usado no desenvolvimento de auxílios de escrita computadorizados para, entre outras possibilidades, ajudar na construção de textos que aprimorem os resultados apresentados por sistemas de tradução automática entre o português e o inglês.

Como visto anteriormente, a intervenção realizada após o texto ser traduzido recebe o nome de pós-edição (PE). Em resumo, o trabalho de pós-edição envolve revisar a saída bruta da TA para corrigir erros e adaptar o texto traduzido às expectativas de qualidade e custos do cliente da tradução. Nos últimos anos, por conta do potencial de redução de custos e tempo de produção, do aumento da qualidade de sistemas de tradução automática e da maior disponibilidade de programas gratuitos de qualidade para TA e PE, a pós-edição se tornou prática recorrente na indústria de tradução (O'Brien et al., 2014).

2.2.3. Pós-edição de TA

Em vista do aumento exponencial no conteúdo que precisa de tradução nos últimos anos e das novas exigências de produção cada vez mais rápida e barata de conteúdo traduzido, a tradução automática e a pós-edição da saída de TA passaram a ter muito mais importância (Dillinger, 2014). Esse novo contexto ensejou vários estudos sobre pós-edição, relacionados a tópicos como ganho de produtividade, esforço cognitivo, modelos de precificação, treinamento e qualidade, o que mostra uma importante mudança de abordagem. Se antes os estudos se concentravam em questões pedagógicas e conceituais, agora o foco são estudos práticos e empíricos, que abordam processos de tradução como realizados no mundo real.

A pós-edição passa a ser descrita como uma tarefa “*fit for purpose*”, ou seja, adequada às características e ao objetivo final do texto a ser traduzido, ocupando espaços em que a tradução humana não é viável por questões de tempo, de custo ou outras. Dessa forma, o objetivo, o volume, o público-alvo e a urgência de tradução do conteúdo (Penkale e Way, 2013) se tornam aspectos determinantes para se determinar quanto a saída de tradução automática precisa ser editada para se adequar ao objetivo esperado.

O’Brien et al. (2009) traz uma visão geral da pós-edição pela ótica de pesquisadores, tradutores e consumidores. O documento apresenta uma categorização da pós-edição segundo a natureza e o volume das alterações realizadas da saída da TA. As duas categorias propostas são: PE ligeira e PE convencional (ou PE completa). A primeira, como o próprio nome indica, privilegia a produtividade em detrimento da qualidade, considerada baixa, e por isso apenas alterações essenciais para o correto entendimento do texto são necessárias. Para a segunda, a expectativa é de qualidade média. Nas diretrizes, O’Brien destaca que a mensagem transmitida deve ser correta em ambos os casos, embora preconize que a PE ligeira não deve se concentrar em problemas gramaticais, a menos que afetem a mensagem, nem em padrões textuais, como coesão e coerência. A produtividade deve ser alta e a qualidade esperada é baixa. Na PE convencional, espera-se que a gramática esteja correta, mas continua a não haver grandes preocupações com estilo. Não se deve perder muito tempo com pesquisa de termos, o que é condizente com a expectativa de alta produtividade.

Way (2013), no entanto, entende de maneira distinta os graus de pós-edição. Para o autor, na PE ligeira, o tradutor edita a saída da TA para gerar um texto-fonte compreensível, mas sem se preocupar com o estilo da tradução. Ao texto, basta ser correto e adequado ao uso. A PE ligeira pode ser usada, por exemplo, na tradução de documentação de suporte para uso de um produto, de arquivos de ajuda online e de manuais, quando houver baixo risco de problemas de saúde e segurança.

A PE completa, no entanto, deve ter qualidade semelhante à de uma tradução humana. Neste caso, a pós-edição tem a vantagem de ser mais rápida e menos dispendiosa que a tradução humana, embora não tanto quanto a PE ligeira. Way lista possíveis casos de uso, tais como a tradução de manuais em que questões de saúde e segurança sejam relevantes, de informações sobre produtos a serem exibidos em sites de comércio online, de relatórios que necessitem de certificação, bem como de contratos e patentes. Way identifica ainda a transcrição³, ou seja, traduções criativas e adaptação cultural de conteúdos com fins de publicidade e marketing, como um caso de uso em que não é possível usar a TA. Podemos afirmar, em complemento, que todos os casos de uso de tradução em que a adaptação cultural e/ou o trabalho em grupo seja um fator chave, como a tradução de videogames, por exemplo, não são adequados ao uso de TA.

A influência do uso de pós-edição completa na qualidade final do texto traduzido é uma preocupação de autores como Almeida (2013) e Guerberof (2014) e será discutida no próximo capítulo, que tratará também da avaliação de qualidade da TA e da fundamentação teórica do presente trabalho.

³ Aqui, o termo transcrição é usado no sentido adotado pela indústria de tradução atual, qual seja, o processo de adaptar criativamente anúncios ou slogans publicitários para outra cultura. O conceito de transcrição proposto por Haroldo de Campos está fora do escopo deste trabalho.

3 Perspectiva teórica

Neste capítulo, vamos apresentar as perspectivas teóricas que orientam o presente trabalho, com destaque para a filosofia da linguagem de Wittgenstein, expressa principalmente nos jogos de linguagem descritos nas *Investigações filosóficas* e no slogan “o significado está no uso”. Nossa abordagem considera a tradução humana, a tradução automática e a avaliação de tradução como distintos jogos de linguagem, com aspectos que se superpõem e se entrecruzam, mas também com características próprias. Vamos elencar essas características e explicar como a visão dos jogos de linguagem se aplica ao nosso jogo de avaliação de sistemas de TA. Também vamos apresentar o conceito de categorização, apresentado detalhadamente em Ellis, e sua relação com a criação de taxonomias para avaliação de traduções. Por fim, discutiremos as noções de significado e sujeito no âmbito de teorias não-logocêntricas e sua relação com a filosofia de Wittgenstein e a abordagem da avaliação de tradução feita neste trabalho.

3.1. A filosofia da linguagem de Wittgenstein

Este trabalho está alinhado à visão pragmática da linguagem, conforme estabelecida pelos filósofos gregos sofistas, na qual a linguagem não é uma *representação* do real, como no pensamento de Platão, tampouco está subordinada à interpretação que a “alma” faz do real, conforme proposto por Aristóteles. Esse contraponto à concepção tradicional platônico-aristotélica pode ser expresso pela máxima de Protágoras, segundo a qual “o homem é a medida de todas as coisas”. Dito de outra forma, na visão pragmática, a linguagem não estabelece um nexo entre uma palavra e algo exterior; ao contrário, “é o objeto exterior que se manifesta no discurso” (Górgias, *Do não ser*, Fragmento B, III apud Martins, 2001).

Essa visão da linguagem encontra eco na obra *Investigações filosóficas*, de Ludwig Wittgenstein (doravante IF), publicada postumamente em 1953. Nela, Wittgenstein refuta a “distinção intrínseca entre sujeito e objeto” do pensamento

platônico-aristotélico, apresentando uma reflexão que vai em direção oposta e desloca o significado para o uso da linguagem. Assim, o significado de uma palavra se estabelece quando ela “é usada ou vivenciada no fluxo das práticas e costumes de uma comunidade linguística, histórica e culturalmente determinada” (Martins, 2001). Ou, nas palavras do próprio Wittgenstein:

Na *práxis* do uso da linguagem [...] podemos imaginar que todo o processo do uso das palavras [...] é um daqueles jogos por meio dos quais as crianças aprendem sua língua materna. Chamarei esses jogos de “jogos de linguagem”, e falarei muitas vezes de uma linguagem primitiva como de um jogo de linguagem. Chamarei também de “*jogos de linguagem*” o conjunto da linguagem e das atividades com as quais está interligada. (IF § 7)

Se o uso da linguagem é um jogo cuja *práxis* se dá na enunciação de palavras, essas são, na visão de Wittgenstein, como ferramentas dentro de uma caixa. Podemos nos valer da chave de fenda, do metro, do serrote, do martelo, de pregos e parafusos para montar, consertar, construir, modelar os mais variados objetos. E a linguagem se vale das palavras, suas ferramentas, para cumprir as mais variadas funções, os mais diversos usos, em uma inúmera variedade de jogos. O significado de cada palavra está no uso que é feito dela em uma determinada comunidade, em uma determinada situação ou em um determinado tempo. E esse uso nunca é uma limitação, mas um leque de possibilidades, pois não há qualquer restrição para o uso das ferramentas/palavras. Se tomarmos como exemplo as ferramentas propriamente ditas, veremos que nada nos impede que usemos, por exemplo, o cabo de uma chave de fenda para bater um prego, ou a cabeça de um martelo para coçar as costas. De fato, se pensarmos, por exemplo, no uso da ironia em um jogo de linguagem, uma palavra pode ter, em determinado uso, um significado oposto ao que teria em outro.

Quantas espécies de frases existem? Afirmação, pergunta e comando, talvez? — Há inúmeras de tais espécies: inúmeras espécies diferentes de emprego daquilo que chamamos de “signo”, “palavras”, “frases”. [...] O termo “jogo de linguagem” deve aqui salientar que o falar da linguagem é uma parte de uma atividade ou de uma forma de vida. (IF § 23)

Wittgenstein alerta que não é possível *explicar* ou *sistematizar* a linguagem, pois qualquer explicação ou sistematização também se dá dentro do âmbito da

linguagem. Não é possível *parar* o jogo da linguagem para se tentar inferir qualquer regra ou sistematização.

Nossos claros e simples jogos de linguagem não são estudos preparatórios para uma futura regulamentação da linguagem — como que primeiras aproximações, sem considerar o atrito e a resistência do ar. Os jogos de linguagem figuram muito mais como *objetos de comparação*, que, através de semelhanças e dessemelhanças, devem lançar luz sobre as relações da nossa linguagem. (IF § 130)

Logo, só podemos *descrever* como se dão os jogos de linguagem e *compará-los* a outros com relação à sua aplicação ou às suas funções, por exemplo. Este trabalho busca, então, descrever a tradução como um *jogo de linguagem*. Veremos quem são os *jogadores*, como é a *dinâmica*, que *lances* ocorrem e quais os *objetivos* de distintos jogos de tradução, com destaque para os jogos da tradução humana, da tradução automática e da avaliação de tradução.

3.2. A tradução e seus jogos de linguagem

Traduzir é um jogo de linguagem de múltiplas possibilidades. A tradução pode envolver distintas formas de expressão: i) de língua falada para língua falada, como na interpretação; ii) de língua escrita para língua escrita, como na tradução literária; iii) de língua falada para língua escrita, como na legendagem; iv) de língua falada para a língua visual, no caso de idiomas para cegos. Pode ser: i) de uma língua a outra; ii) de uma língua a outra variante da mesma língua, como no caso do português do Brasil e de Portugal. Podemos ter tradução até mesmo dentro de uma mesma língua, em formas de expressão distintas, se considerarmos sistemas computadorizados que traduzem línguas orais escritas para deficientes visuais. Dinâmicas tão distintas e, ainda assim, todas podem ser consideradas tradução, reproduzindo a “rede de semelhanças” que Wittgenstein enxergou nos jogos:

Considere, por exemplo, os processos que chamamos de “jogos”. Refiro-me a jogos de tabuleiro, de cartas, de bola, torneios esportivos etc. O que é comum a todos eles? Não diga: “Algo deve ser comum a eles, senão não se chamariam ‘jogos’”, — mas veja se algo é comum a eles todos. — Pois, se você os contempla, não verá na verdade algo que fosse comum a *todos*, mas verá semelhanças, parentescos, e até toda uma série deles. [...] E assim podemos percorrer muitos, muitos outros grupos de jogos e ver semelhanças surgirem e desaparecerem.

E tal é o resultado desta consideração: vemos uma rede complicada de semelhanças, que se envolvem e se cruzam mutuamente. Semelhanças de conjunto e de pormenor. (IF § 66)

Para os usos de nosso trabalho, vamos nos ater à tradução de textos escritos entre línguas distintas. Para falar desse jogo de tradução, nada mais justo que lançar mão de conceitos que remetam à ideia de jogo. Vamos falar dos *jogadores*, ou seja, de quem tem parte ativa no processo de tradução. Vamos falar do *objeto*, ou seja, o que se *usa* para jogar. Em um jogo seria a bola, as peças ou as cartas; na tradução, o texto-fonte, o texto-alvo e outros. Vamos falar do *ambiente* do jogo, onde ele ocorre, ou seja, que tipo de sistema, ferramenta ou plataforma o tradutor usa na *dinâmica* do traduzir. Vamos falar dos *lances* que os jogadores fazem: a escrita do texto-fonte pelo autor; a leitura, a tradução e a revisão de cada palavra e frase pelo tradutor, a leitura do texto-alvo pelo público. Vamos falar das *regras* e dos *objetivos* de cada jogo.

Assim, o jogo da tradução escrita tem como jogadores o autor, o tradutor e o leitor. Na dinâmica do traduzir, o tradutor expressa o texto-fonte, objeto ininteligível (caso desconheça o idioma) ou apenas inacessível para o leitor, de uma forma que seja usual para o tradutor e o leitor. O objetivo do jogo, para o tradutor, é tentar “transmitir aos leitores a mesma imagem, a mesma impressão que ele próprio teve” (Schleiermacher, 2011, p. 22), ou seja, causar no leitor o mesmo *efeito* que o texto original lhe causou.

O primeiro lance é do autor (ou autores), que produziu o texto-fonte. Toda tradução parte de um texto-fonte; sem ele, não há um traduzir possível. Logo, é justo que seja considerado o primeiro lance do jogo. O tradutor faz o segundo lance ao ler o texto-fonte, que então comparece como objeto, e o lance seguinte ao redigir sua tradução, conforme a sua interpretação do texto-fonte. Toda interpretação é única, e toda tradução dessa interpretação é única. Os lances de ler e redigir se repetem e compõem a *dinâmica* do traduzir, com o tradutor usando o seu domínio da técnica para chegar ao *objetivo*: produzir o texto-alvo conforme o melhor de suas habilidades. Entre os lances de leitura e tradução, outros são feitos, tais como pesquisas de termos para esclarecimento do significado de alguma palavra naquele uso. A dinâmica de revisão, por sua vez, geralmente comparece ao final da primeira parte do jogo da tradução, após o tradutor produzir a primeira versão do texto-alvo, ainda sujeita a lances para correção, aprimoramento e adequação à função ou

finalidade da tradução. Importante ressaltar a relevância da adequação da tradução ao objetivo do jogo. Caso isso não aconteça, os lances do tradutor podem perder eficácia, causando um efeito distinto do pretendido. Com lances equivocados, o objetivo não será alcançado.

Trataremos, agora, de diferentes tipos de tradução escrita como diferentes jogos de linguagem. Como se dão? Quais são as semelhanças e as dessemelhanças entre uma tradução literária de um livro de ficção e uma tradução especializada, como, por exemplo, de um pedido de patente de rações secas para animais? Certamente será possível encontrar semelhanças e parentescos entre ambos, e também o que há só em um deles e não no outro. Ao pensar no assunto, talvez seja possível descobrir relações inesperadas, que uma visão estanque do que seja cada domínio — literário *versus* científico — não nos permitiria vislumbrar.

Não posso caracterizar melhor essas semelhanças do que com a expressão “semelhanças de família”; pois assim se envolvem e se cruzam as diferentes semelhanças que existem entre os membros de uma família: estatura, traços fisionômicos, cor dos olhos, o andar, o temperamento etc., etc. — E digo: os “jogos” formam uma família. (IF § 67)

Sim, podemos dizer que os jogos de tradução também formam uma família, pois têm traços semelhantes. Não é necessário que todas as instâncias de um mesmo jogo de linguagem tenham pelo menos uma característica *comum a todas*. Se observarmos um número suficientemente grande de exemplos, pode ser que não encontremos tal característica comum, mas isso não invalida a semelhança entre essas instâncias.

Como explicaríamos a alguém o que é um jogo? Creio que lhe descreveríamos *jogos*, e poderíamos acrescentar à descrição: “isto e *outras coisas semelhantes* chamamos de ‘jogos’”. E nós próprios sabemos mais? Será que apenas a outrem não podemos dizer exatamente o que é um jogo? — Mas isto não é ignorância. Não conhecemos os limites, porque nenhum está traçado. Como disse, podemos — para uma finalidade particular, traçar um limite. É somente daí que tornamos o conceito útil? De forma alguma! A não ser para esta finalidade particular. (IF § 69)

Haveria algo de semelhante entre os jogos da tradução humana, da tradução automática e da avaliação de tradução? Certamente sim. Podemos encontrar características comuns se limitarmos nossa observação, análise e avaliação a uma determinada finalidade. Essa finalidade, deve-se ressaltar, é um limite tão arbitrário

quanto qualquer fronteira física, mas tem *utilidade*. Para os propósitos deste trabalho, essas características serão úteis para entendermos como se dá a relação entre o tradutor humano e a tradução automática, por exemplo, e para entender a necessidade da avaliação de tradução.

3.2.1. O jogo da tradução humana

Logo no começo da descrição do jogo da tradução humana, nos damos conta da dificuldade de *sistematizar* qualquer jogo de linguagem. Quem e quantos são os jogadores? Serão dois, o tradutor e o leitor, se quem escreveu o texto também o tiver traduzido. Um trio, se autor e tradutor forem duas pessoas diferentes, sendo que o autor comparece como o responsável pelo primeiro lance, o texto-fonte, também objeto do jogo, a quem o tradutor sempre retorna para lances de leitura durante a dinâmica do traduzir. E serão muitos, se pensarmos no jogo da interpretação em um evento: quem fala, quem traduz (geralmente, dois intérpretes em uma cabine) e o público no local. E serão incontáveis, se considerarmos um grande projeto de localização: múltiplos autores dos textos, múltiplos tradutores, múltiplos leitores.

Como é o ambiente? Pode ser estritamente físico, uma folha inserida em uma máquina de escrever na qual tradutor datilografa a tradução enquanto olha para o livro original aberto sobre um suporte, algo impraticável para qualquer jogo de tradução em que a tecnologia compareça. Outra possibilidade é o trabalho com uma ferramenta de tradução assistida por computador, o texto-fonte aberto em um arquivo PDF e várias abas de dicionários, glossários e afins abertas no navegador da internet. Nesse caso, o sistema local do tradutor seria o ambiente. Poderia ser um ambiente misto, físico e virtual, com um grupo de tradutores reunidos em uma sala, discutindo ideias enquanto trabalham em um sistema em rede, com memórias de tradução e glossários compartilhados. Ou totalmente virtual, com tradutores espalhados em diversos pontos do mundo editando o mesmo documento em uma plataforma online, segmento a segmento.

E como é a dinâmica dos lances da tradução? Talvez o tradutor leia todo o texto e depois volte para traduzir. Também pode traduzir frase a frase, palavra a palavra. É possível que escreva um grande rascunho tradutório e deixe para resolver

todas as dúvidas para o final. Ou resolva cada problema antes de continuar. Também pode resolver as dúvidas mais fáceis e deixar as mais difíceis para o fim. E a dinâmica de revisão? Pode revisar tudo após concluir a tradução ou simplesmente não revisar. Talvez revise duas vezes, uma cotejando o texto-fonte e outra lendo somente o texto-alvo. Por fim, que regras o tradutor adota ao dar seus lances no jogo, ou seja, ao fazer suas escolhas de tradução? Segue as orientações do cliente, ajusta o texto para agradar o público-alvo ou não faz concessões e traduz como quer?

No jogo da tradução humana, como dito anteriormente, o tradutor busca produzir no texto-alvo o mesmo efeito que percebeu no texto-fonte. Esse efeito, porém, varia de acordo com a finalidade da tradução. Logo, na tradução de textos literários, o efeito esperado é um, na transcrição de textos publicitários, outro, e na tradução de textos científicos, ainda outro. Podemos dizer que, no primeiro, o tradutor vai se preocupar mais com as *sensações e sentimentos* do leitor, no segundo, vai buscar cativar o leitor/consumidor da mesma forma como foi cativado pelo original, enquanto, no terceiro, vai privilegiar as informações contidas no texto-fonte. A finalidade da tradução orienta os lances dados para se chegar ao objetivo comum do jogo da tradução humana.

Uma visão logocêntrica, alinhada com a tradição platônico-aristotélica, dividiria os exemplos acima em categorias estanques de texto — literário, publicitário e científico, por exemplo — e adotaria uma forma *prescritiva* de defini-los. O texto literário precisaria conjugar fluência e estilo, enquanto a transcrição exigiria soluções criativas e capacidade de adaptação cultural. Textos científicos, por sua vez, precisariam mesclar clareza e precisão terminológica. Essa divisão parecer a pretensão de *apontar um caminho* para o traduzir. No entanto, cada tradutor faz seu caminho. Cada um dá os lances conforme uma dinâmica própria de tradução. Todos os tradutores tem à disposição as mesmas ferramentas — as palavras —, mas o uso que fazem delas é absolutamente individual. E esse uso reflete seu repertório linguístico, que não compartimentaliza os textos em diferentes categorias; ao contrário, o tradutor usa o conhecimento como um todo, o tempo todo, nos lances do jogo da tradução.

Uma visão relevante sobre a noção de categorias estanques na tradução é encontrada em Santos (2006), que apresenta uma problematização de diversas contraposições há muito cristalizadas no campo da tradução, tais como: tradução

literária e científica, tradução e criação, tradução humana e automática. No caso da dicotomia entre tradução literária e científica, Santos reconhece que existem técnicas e teorias diretamente aplicadas a cada um destes ramos da tradução, mas discorda da visão de que textos científicos são eminentemente objetivos e literais. Estes muitas vezes recorrem a metáforas e à criatividade “para transmitir novos conceitos e novas formas de ver ou novas teorias” (Santos, 2006, p. 5), e os conceitos científicos não são reflexos da verdade, mas dependem da visão e da subjetividade dos cientistas. Por fim, Santos argumenta que a língua “não tem compartimentos estanques para linguagem técnica e linguagem geral” (idem, p. 6).

Ao tratar da dicotomia entre tradução e criação, Santos questiona a ideia de que o tradutor apenas presta um serviço ao autor-criador, tentando, inutilmente, alcançar o valor do original.

Mas, se considerarmos que o tradutor é um criador, e que está a usar as ideias do criador inicial para uma nova criação (o que aliás – note-se – não é rigoroso, porque está a usar o seu conhecimento do autor, da língua do autor e da cultura do autor para deduzir uma interpretação das ideias do criador) pode imediatamente concluir-se que, como todo o acto de criação, pode ou não (se for possível e lícita a comparação) ser melhor do que o “original” ou primeira versão. (Santos, 2006, p. 8)

Santos frisa que essa possibilidade de aprimoramento do texto-fonte é considerada factível no caso de textos científicos (tradução especializada), mas não quando se trata de textos literários. Afirma, porém, que se considerarmos o tradutor como um criador que parte das ideias do autor e reinterpreta o texto-fonte para criar um novo texto em sua língua, é perfeitamente aceitável concluir que a tradução, como ato de criação, pode ou não ser melhor que o “original”. Em nossa visão, no entanto, ao entendermos a tradução como um jogo de linguagem, podemos muito bem aceitar que os lances do tradutor sejam mais efetivos na língua-alvo do que os do autor na língua-fonte em qualquer tipologia de texto, seja científica, literária, anedótica ou acadêmica, visto que o tradutor tem, também, mais elementos para jogar por conta de seu “conhecimento (linguístico e cultural) duplo” (idem, p. 8).

O texto-fonte é um fato móvel, logo, passível de aprimoramento a depender dos jogadores envolvidos e da dinâmica do jogo da tradução. Mesmo não sendo a *finalidade* primeira da tradução, o aprimoramento do texto-fonte comparece como elemento da *dinâmica do traduzir* de muitos tradutores. Jogadores diferentes fazem lances diferentes. O craque de um time de futebol vai fazer jogadas que um

companheiro mediano jamais faria. O objetivo dos dois é ganhar, mas os *lances* do jogo de cada um serão completamente distintos.

3.2.2. O jogo da tradução automática

Ao tratar de tradução humana e tradução automática, Santos pondera que a tradução tem por objetivo a comunicação entre seres humanos, logo, não é possível imaginar que “sejam substituídos por uma máquina sem que tenham um papel preponderante no processo” (idem, p. 4). Afirma também que o paradigma original da inteligência artificial, que via a tecnologia como substituta para o humano, deu lugar à ideia de apoio inteligente, “cujo objectivo é facilitar o trabalho intelectual de um (ou mais) seres humanos e não substituí-los, tentando todavia libertá-los de tarefas repetitivas e aproveitando aquilo que já foi feito” (ibidem). No entanto, o objetivo de produzir traduções automáticas prontas para uso continuou sendo o paradigma para a pesquisa e o desenvolvimento de sistemas de TA até recentemente.

Nos últimos anos, porém, emergiu um novo paradigma, alinhado com a visão de apoio inteligente, no qual a tradução automática deixou de ser vista como uma solução geral, que daria conta de traduções para qualquer finalidade, para se tornar uma ferramenta que pode fornecer sugestões úteis para tradutores humanos. As pesquisas se voltaram, então, para “formas de integrar melhor *a tradução humana e a tradução automática* para aprimorar a produtividade e a experiência de usuário do tradutor” (Bentivogli et al., 2016a, p. 388; grifo dos autores).

No mercado de tradução atual, pode-se perceber dois movimentos em relação ao uso de tradução automática: por um lado, a TA vem ocupando espaços que estão fora do escopo do tradutor profissional, nos quais, pelas mais diversas razões, a tradução humana é inviável; por outro, embora ainda a passos lentos, a TA vem deixando de ser vista como uma ferramenta de tradução de máquina assistida por humanos (TMAH), como antes proposto por Bowker (2002), e passando a ser considerada um sistema que auxilia um tradutor profissional em seu trabalho, ou seja, uma ferramenta de tradução assistida por computador (TAC). Assim, o paradigma da substituição perde duplamente o sentido: no primeiro caso, a tradução automática se torna uma ferramenta de apoio ao tradutor; no segundo, uma tecnologia disponível para usos que *não comportam* intervenção humana. Por isso,

para todos os usos de presente trabalho, vamos considerar os sistemas de tradução automática como ferramentas de tradução assistida por computador, ou seja, como ferramentas que fazem parte do processo e da *práxis* da tradução profissional.

A tradução automática é, então, um jogo de linguagem que pode ter um ou vários jogadores. No caso da TA sem edição, temos o autor ou autores do texto e o leitor da tradução, que efetivamente deu o lance de requerer a TA. Podemos considerar a máquina como o *meio* que permite as interações, ou seja, os lances da tradução entre autor e leitor, que neste caso pode ser um tradutor ou não, pode ser uma pessoa monolíngue ou não, pode falar o idioma do texto original ou não.

No caso da TA com pós-edição, temos como jogadores o autor ou autores do original, um tradutor profissional e o leitor ou leitores do texto. Neste caso, ao editar a saída da TA, o tradutor humano se torna o *mediador* entre o autor e o leitor. Assim, a pós-edição traz um novo *objeto* ao jogo da tradução. Se antes havia apenas o texto-fonte como objeto, agora a saída da TA também comparece nessa função, inclusive como objeto *de facto*, pois é o texto que será efetivamente manipulado durante a dinâmica do traduzir.

Em termos do ambiente do jogo da TA sem edição, os lances podem se dar em um sistema de TA de uso local, isto é, instalado no computador do leitor/receptor da tradução, em um tradutor online, como o Google Tradutor ou o Bing Translator, ou em um site que ofereça um serviço de tradução automática, como o Facebook, entre outras opções. O leitor dá o primeiro lance ao realizar a tradução automática da forma disponível na ferramenta. O lance seguinte é a leitura do texto traduzido. Nessa dinâmica, ao contrário do que ditava o antigo paradigma, o leitor não tem a expectativa de um texto perfeito, mas apenas minimamente compreensível.

Na dinâmica da pós-edição, seja completa, ligeira ou intermediária, o lance de realizar a tradução automática pode ou não ter sido dado antes que o tradutor tivesse acesso ao texto. Caso já tenha, a situação mais provável é que o tradutor profissional tenha recebido um pedido de tradução de uma agência ou de outro intermediário que recebeu o texto-fonte do autor. Neste caso, o autor pode ser entendido não como uma pessoa ou pessoas específicas, mas como a empresa que

produziu o texto⁴. Mais uma vez, o texto-fonte é o objeto de consulta, enquanto a saída da TA é o objeto *de facto* dos lances de leitura e pós-edição do tradutor.

A *expectativa de qualidade* do texto-alvo vai ditar a maneira como o tradutor dará seus lances no jogo da pós-edição: serão mais elaborados, no caso da pós-edição completa, ou mais rápidos, no caso da pós-edição ligeira. Dito de outra forma, as *regras* do jogo variam de acordo com a *finalidade* da tradução: produzir um texto gramaticalmente correto, mas sem grandes preocupações com estilo, caso em que o uso de pós-edição ligeira se aplica, ou um texto com qualidade de publicação, semelhante ao que se esperaria de uma tradução humana, caso em que se faria uma pós-edição completa. Assim, as regras do jogo da tradução são entendidas como uma prática, como em Wittgenstein: “Eis porque ‘seguir a regra’ é uma *práxis*. E *acreditar* seguir a regra não é seguir a regra” (IF § 202). Ou, como explicitou Glock em seu *Dicionário Wittgenstein*:

A relação entre uma regra e sua aplicação correta, assim como outras relações intencionais, é uma relação interna. Compreender a regra é saber como aplicá-la, saber o que pode ser considerado como agir em conformidade com ela ou transgredi-la. (Glock, 1998, p. 315)

Logo, o tradutor aplica a regra conforme compreende a finalidade da tradução. A aplicação da regra da qualidade estará sempre atrelada à finalidade do texto traduzido, que por sua vez, determina o tempo do jogo da tradução. Em termos práticos, o nível de expectativa de qualidade de um texto-alvo produzido por meio de pós-edição de TA é inversamente proporcional à expectativa de produtividade da tarefa. Para textos cuja finalidade é passar uma informação sem grandes preocupações estilísticas, maior produtividade, menor qualidade; para textos que exijam qualidade de publicação, maior qualidade, menor produtividade.

Por fim, outro aspecto importante a ser levado em conta no jogo da TA, com ou sem pós-edição, é a utilidade do texto-alvo. Se um texto traduzido automaticamente é útil sem qualquer edição por um tradutor humano, não há sentido, em termos econômicos e de tempo, em investir em uma tarefa de pós-edição. Em entrevista publicada online⁵, Jost Zetsche ilustra bem essa questão ao

⁴ Para uma discussão sobre o conceito atual de autoria, que não cabe nesta dissertação, vide Estill, 2017.

⁵ Disponível em: <<http://www.internationalwriters.com/BigWave/BigWaveJZ.htm>>. Acesso em: 15 ago. 2018.

afirmar que, com base em comentários de usuários da base de conhecimento da Microsoft, um texto traduzido automaticamente é tão útil quanto outro traduzido por um humano, embora o nível de qualidade do último seja muito maior. E para determinar a qualidade de um texto traduzido, seja por máquina ou por um humano, é preciso avaliar esse texto.

3.2.3. O jogo da avaliação de tradução

A avaliação de tradução é uma tarefa complexa, que envolve uma série de fatores linguísticos e extralinguísticos. As dificuldades inerentes a se tentar abordar uma tarefa subjetiva de forma objetiva mostram que não é possível pensar em uma solução única de avaliação. Uma avaliação de tradução é sempre feita em função de critérios que, por sua vez, são definidos em função da *finalidade* da tradução. Diferentes tipos de tradução demandam diferentes tipos de avaliação. Os aspectos e critérios que definem um jogo de avaliação mudam sensivelmente quando, por exemplo, tratamos da avaliação de tradução humana ou da avaliação de tradução automática. Em ambos os casos, “não há uma única abordagem ou métrica que baste para atender todos os objetivos e cenários de avaliação” (Hovy et al., 2002 apud Castilho et al., 2018).

O jogo da avaliação de tradução tem como objetivo determinar a adequação do conteúdo a determinada finalidade. No caso específico da avaliação de TA, o objetivo é analisar o desempenho de um sistema, ou seja, avaliar a qualidade da saída de TA enquanto resultado de um processo segundo determinados parâmetros pré-estabelecidos, os critérios de avaliação, para a finalidade proposta. No primeiro jogo, avalia-se a tradução *per se*; no segundo, avalia-se o sistema ou ferramenta de TA em função da tradução.

Em termos de jogadores envolvidos, temos, no jogo da avaliação de tradução humana: i) o autor, cujo original comparece como objeto central para a dinâmica de avaliação, essencial para análise da qualidade da tradução mediante cotejo (a não ser, situação rara, que a avaliação se concentre apenas no texto traduzido); ii) o tradutor, que comparece como *autor do texto-alvo* a ser avaliado; e iii) o avaliador, que pode ser, entre outros, tradutor, crítico, jornalista ou acadêmico, a depender do

tipo de texto e da finalidade da avaliação (fazer uma crítica, uma resenha, analisar um teste de tradução, entre outros).

Na avaliação de TA, mudam os jogadores e o objeto da avaliação. O autor do original, por exemplo, geralmente deixa de tomar parte ativa no jogo. Na maior parte dos casos, o autor do texto a ser avaliado é o sistema de TA ou o pós-editor. O texto-fonte deixa de ser o *objeto central* do jogo para se tornar um objeto de consultas eventuais, deixando o papel de *objeto da avaliação* para: i) a saída da TA, em avaliações de desempenho do sistema ou sistemas de TA; ou ii) o texto-alvo, em avaliações que abordem a tradução final produzida por um tradutor após um processo de pós-edição. Neste último caso, o avaliador provavelmente será um tradutor ou revisor humano, que fará uma avaliação qualitativa. Naquele, o avaliador pode ser um tradutor humano ou um sistema de avaliação automatizado.

Em relação à dinâmica, pode-se dizer que a avaliação de tradução depende de parâmetros (critérios) para ser levada a cabo a contento.

Buscando estabelecer tais parâmetros teóricos, tradutores e críticos têm discutido noções que afetam diretamente a prática da crítica de tradução, dentre elas: interpretação, fidelidade, equivalência, literalidade e fluência. Entretanto, como não há consenso sobre esses conceitos, tampouco há entendimento em torno do que seja qualidade em tradução ou uma metodologia amplamente satisfatória para a formulação da crítica de traduções. (Alves, 2009, p. 22)

Embora se concentre apenas na avaliação de tradução humana, o trabalho de Alves (2009) é relevante para discutir a dificuldade de estabelecimento de critérios para se avaliar um texto traduzido, a começar pela dificuldade de se superar a dicotomia entre tradução a literária e tradução técnica.

Mesmo ainda vigente e predominante, a distinção entre técnico e literário vem sendo problematizada. Em trabalhos acadêmicos que tratam especificamente dessa questão e em cursos de formação de tradutores, já se pode encontrar, mesmo que com frequência menor, diferentes distinções, como: tradução literária e tradução técnico-científica; tradução de ficção e tradução não ficcional (ou de não ficção); e ainda, tradução literária e não literária. Entretanto, ainda não se conseguiu fugir da polaridade de classificação textual. (Alves, 2009, p. 19)

Essa distinção se faz notar, inclusive, no nome que se dá ao ato de analisar a qualidade de uma tradução. De traduções literárias, seja de ficção ou não ficção, faz-se uma crítica; de traduções técnico-científicas, ou, como preferimos, traduções especializadas, faz-se uma avaliação. Ao analisar as teorias de autores como House, Reiss e Berman sobre a crítica de tradução, Alves (2009) afirma que a atividade “deve considerar de modo específico cada tipo de texto a ser avaliado, ou, mais ainda, deve considerar cada par original-tradução como um caso único”. Ou, em nossa visão, considerar cada avaliação como um jogo de linguagem único, com jogadores, dinâmicas e objetivos próprios.

As dinâmicas da avaliação de traduções humanas serão bem distintas, de acordo com a finalidade. No caso da crítica, Britto afirma:

Ainda que não haja consenso absoluto, [...] é possível utilizar o discurso racional para fazer avaliação e tecer considerações em torno de traduções, fazendo referência a certas propriedades dos textos traduzidos com relação às quais há um certo grau de acordo entre um bom número de pessoas envolvidas nas atividades de traduzir. (Britto, 2006 apud Alves, 2009, p. 23)

Em seus lances no jogo da avaliação da tradução de textos de ficção ou não-ficção, e de acordo com sua orientação filosófica em relação à linguagem, o crítico pode: i) se apoiar em conceitos como fidelidade e equivalência com base no cotejo com o texto-fonte, uma visão que admite uma estabilidade do significado das palavras no processo de traduzir, ou ii) estabelecer parâmetros baseados em propriedades razoavelmente consensuais a este ambiente da tradução, uma dinâmica mais próxima da nossa visão da linguagem como um jogo.

Na dinâmica de avaliação de sistemas de TA, métricas automatizadas de avaliação de TA são uma alternativa rápida e consideravelmente mais barata à avaliação humana. Na dinâmica da avaliação automatizada, por meio de métricas como BLEU e TER, por exemplo, temos um *processo* que gera resultados *rápidos*. Na maioria dos casos, o texto original não comparece neste jogo de linguagem, sendo substituído pela comparação da saída de TA com uma ou mais traduções humanas do mesmo texto original, chamadas de *golden*, ou padrão-ouro, que serão usadas como referência. Em suma, a função de métricas automatizadas “é medir a semelhança da saída de determinado sistema à(s) tradução(ões) de referência,

calculando um escore para quantificar essa semelhança” (Koehn, 2010 apud Castilho et al., 2018, p. 25).

A métrica TER (Translation Edit Rate, “taxa de edições na tradução”), por exemplo, se baseia na “quantidade de pós-edições (em termos de substituições, inserções e exclusões) necessárias para produzir uma sentença-alvo de boa qualidade a partir da saída da TA” (Way, 2018, p. 15). Quanto maior o número de alterações, menor o escore atribuído à tradução. Métrica automatizada mais usada pela indústria, o BLEU (Bilingual Evaluation Understudy, “estudo de avaliação bilíngue”) se baseia em duas premissas: uma métrica de semelhança de tradução e um corpus de traduções humanas de boa qualidade (Papineni et al., 2002). Dito de outra forma, quanto mais palavras e trechos a TA tiver em comum com a(s) tradução(ões) de referência, melhor será o escore.

Way, no entanto, questiona o paradigma da qualidade das traduções humanas: primeiro, ao afirmar que, ao usar métricas automáticas, “os desenvolvedores de TA se veem obrigados a (erroneamente) considerar traduções humanas como perfeitas” (Way, 2013, p. 2); segundo, ao concluir que “mesmo com revisores no ciclo controlado de tradução, edição e revisão, erros acontecem” (Way, 2018, p. 9). Por fim, lembra que, embora as métricas de avaliação funcionem melhor com várias referências, normalmente apenas um *golden* é fornecido. Além disso, por ignorar a sentença-fonte, as métricas não oferecem uma visão abrangente da qualidade da TA, tampouco fornecem uma indicação imediata dos tipos de problemas contidos no texto-alvo (Castilho et al., 2018), algo que só é oferecido pela avaliação humana.

A definição de categorias de erros de tradução (ou de alterações na tradução, dependendo da abordagem), por exemplo, é uma das estratégias mais comuns para a avaliação humana de traduções. Para Ellis (1993), a categorização “simplifica a complexidade da experiência ao reduzir uma variedade infinita [de fatos] a um conjunto finito de categorias”. Essa redução serve para agrupar como semelhantes coisas que não são semelhantes.

O conceito de agrupamento é um aspecto-chave na discussão sobre o processo de categorização. Para agrupar coisas não semelhantes em uma única categoria, podemos usar uma categorização funcional, ou seja, baseada em nossa experiência de mundo, e não em fatos de realidade. Uma categorização acontece para uma *finalidade*, serve a um propósito, e normalmente é uma *equivalência funcional*, ou seja, diferentes categorizações são possíveis de acordo com a função que cada uma

delas exerce. Um elemento que se encaixe em uma categoria para determinada função pode não se encaixar em outra.

Outro aspecto importante é a ideia de complexidade da categorização. “Quanto mais complexo é um sistema, menos informativo ele é” (Ellis, 1993, p. 33). Isso significa que quanto mais categorias um sistema possui, menos se pode extrair alguma utilidade dele. Uma categoria muito precisa significa que o número de elementos que nela se encaixam será muito reduzido, o que vai contra o propósito da categorização, que é *simplificar* a experiência do mundo.

Em termos práticos, categorizações menos complexas, com categorias mais simples de definir, facilitam o agrupamento dos elementos. No caso de categorizações linguísticas para avaliação de tradução, a definição de categorias mais amplas vai permitir, por exemplo, um maior nível de consenso entre os participantes da avaliação e facilitar a que se cheguem a conclusões úteis a partir das categorias criadas.

Na indústria de tradução, o uso de dinâmicas de categorização é uma estratégia comum para avaliar o trabalho de um tradutor. A TAUS (Translation Automation User Society; “sociedade de usuários de automatização de tradução”), por exemplo, mantém um repositório de artigos sobre avaliação de qualidade de tradução⁶ e um modelo de planilha de avaliação com uma taxonomia de erros⁷. Em suas diretrizes, a entidade recomenda às empresas de serviços linguísticos que forneçam uma categorização para seus revisores, que deve incluir uma política de pontuação de erros para aprovação ou rejeição de traduções, com pesos distintos para a gravidade de cada erro. Além de avaliar tradutores, o objetivo da avaliação de qualidade na indústria de tradução é “garantir que um nível especificado de qualidade seja identificado, medido e entregue ao cliente” (Castilho et al., 2018, p. 11).

A avaliação humana de traduções suscita, por fim, uma pergunta. Se, como disse Way (2018), a tradução humana não tem um nível único de qualidade, como fazer para garantir a qualidade da avaliação do avaliador? Estudos como Guerberof (2014) e Martins e Caseli (2013), por exemplo, usaram mais de um avaliador e

⁶ Disponível em: <https://www.taus.net/knowledgebase/index.php/Translation_quality_evaluation>. Acesso em: 24 ago. 2017.

⁷ Disponível em: <https://www.taus.net/knowledgebase/images/a/a6/Taus_dqf_errortypology_evaluation_template.xlsx>. Acesso em: 24 ago. 2017.

registraram a concordância entre os avaliadores. Já Bentivogli et al. (2016a) e, novamente, Guerberof (2014) adotaram a estratégia de criar estudos com parâmetros bem definidos e utilizar cálculos estatísticos que contemplem desvios ou efeitos de ordem, entre outras medidas.

Neste trabalho, não me alinho à pretensão logocêntrica de “buscar a ‘verdade objetiva’ a respeito da realidade” (Coracini, 2003, p. 19). Isso significa, por exemplo, tentar não transpor para o texto “a visão de imparcialidade imputada à ciência” e evitar “estratégias formais capazes de causar [uma] ilusão de objetividade” (Coracini, 2003, p. 21). Ao contrário, a intenção é apresentar *minha versão* do jogo da avaliação de traduções automáticas, (como, por exemplo, o uso da primeira pessoa ao descrever os exercícios de avaliação no capítulo 5) comparecendo como jogador que tem *autoridade* para jogar o jogo. Autoridade essa que não deriva da racionalidade ou objetividade ao analisar, mas da aplicação de um conhecimento adquirido como tradutor profissional — e, agora, pesquisador — que é útil para a tarefa. Dessa forma, posso produzir avaliações que, embora sem o verniz da objetividade científica, tenham *utilidade*.

3.3. Resumo

Neste capítulo vimos as bases teóricas que norteiam o presente trabalho como um todo e, mais especificamente, os jogos de avaliação de tradução descritos no capítulo 5. Tratamos aqui da filosofia linguística de Wittgenstein e descrevemos como a tradução dá margem a diversos jogos de linguagem, inclusive o jogo da avaliação, em que a tradução deixa de ser o *objetivo* e se torna *objeto* do jogo. Apresentamos também a importância do conceito de categorização para a linguagem e, conseqüentemente, para jogos de avaliação de tradução. Por fim, apresentamos nossa visão não-logocêntrica do pesquisador como um jogador *relevante* para o jogo da avaliação.

4 Estudos relacionados

O presente capítulo apresenta uma resenha da literatura sobre avaliação de tradução computadorizada. Procuramos nos concentrar nos estudos mais recentes, embora também sejam abordados trabalhos mais antigos que tratam da avaliação de tradução automática para o português. Vamos começar descrevendo alguns modelos de avaliação de tradução utilizados na indústria de tradução, seus objetivos e suas formas de estruturação. Em seguida, apresentaremos avaliações de desempenho de tradutores, de ferramentas e sistemas que apresentem dados relevantes para o presente estudo. Procuramos descrever os estudos como jogos de linguagem, identificando os jogadores, o ambiente, a dinâmica, a finalidade e os objetivos, de forma a ter subsídios para comparação entre esses trabalhos e nossos jogos de avaliação, descritos no capítulo 5.

4.1. Modelos de avaliação de tradução

No contexto atual, em que a TA e a pós-edição de TA estão cada vez mais presentes, notadamente em domínios técnicos e especializadas e na localização de softwares e conteúdo para a internet, os novos processos de tradução exigem novas abordagens de avaliação que contemplem os avanços tecnológicos.

Nesse cenário tão complexo e fascinante, parece legítimo argumentar que, embora até agora as noções de qualidade sejam diferentes para TH [tradução humana] e TA, é provável que, em algum momento, ocorra uma convergência a até uma unificação das metodologias e abordagens usadas para a avaliação de qualidade de tradução, como resultado da crescente integração entre as diferentes formas de traduzir (Castilho et al., 2018).

O estudo realizado por O'Brien (2012) em colaboração com a TAUS é uma das primeiras tentativas de se propor um novo modelo de avaliação que levasse em

conta as novas características de uma indústria de tradução em franca expansão, deixando para trás os antigos modelos, considerados estáticos, normativos e lentos. O modelo deveria contemplar os novos tipos de conteúdo que vinham sendo produzidos, tais como tuítes, postagens em blogs e conteúdo gerado por usuários, as novas tecnologias de tradução, notadamente a tradução automática, e as novas demandas de produtividade advindas do desenvolvimento dessas novas tecnologias, entre outros aspectos.

Para chegar à nova proposta, O'Brien (2012) fez um estudo comparativo de onze modelos de avaliação então em uso no mercado de tradução. Oito eram de uso particular de empresas consumidoras de tradução (seis da área de TI) e três estavam disponíveis para o público em geral: o modelo da LISA (Associação de Padrões da Indústria de Localização) e os padrões J2450, produzido pela SAE (Sociedade de Engenheiros Automotivos), e EN15038, do Comitê Europeu de Normalização).

A maioria dos modelos de avaliação analisados se baseava na identificação de tipos de erros de tradução, com aplicação de pontos de penalização com pesos distintos, dependendo da natureza dos erros. A prática mais comum era atribuir três níveis de gravidade: erros menores, erros graves e erros críticos. As traduções avaliadas seriam aprovadas ou reprovadas segundo um limite máximo de pontuação por erros.

O modelo de avaliação por taxonomia de erros da TAUS, citado anteriormente, prevê que os parâmetros de avaliação sejam definidos pelo coordenador de avaliação, que deve definir o valor da penalização, em pontos, por tipo de erro e estabelecer um limite máximo de pontuação para o número de palavras do extrato de texto a ser avaliado.

No modelo da TAUS, um erro menor é aquele que não causa perda de sentido, mas é percebido pelo usuário. Um erro grave pode confundir o usuário ou impedir o uso correto de um produto ou serviço. Um erro crítico pode prejudicar o funcionamento de um produto ou serviço, causar danos à imagem da empresa, afetar a saúde ou a segurança do usuário ou ter implicações jurídicas, entre outros.

Um exemplo de modelo de avaliação de qualidade encontrado em minha experiência no mercado de tradução foi o seguinte: uma planilha de avaliação, apresentada em formato Excel, dividida em abas, tais como: i) instruções de preenchimento; ii) explicação das categorias e dos níveis de gravidade dos erros; iii) detalhamento dos erros encontrados, que traz, entre outras informações, nome

do arquivo, categoria e gravidade do erro, texto-fonte, texto-alvo e descrição do erro, além de um espaço para indicar repetição de erro; e iv) relatório de avaliação, com espaço para inserção do par de idiomas, da quantidade de palavras do texto avaliado e, de resto, preenchido automaticamente de acordo com os dados inseridos na aba de detalhamento.

O modelo conta com cinco categorias de erros — *precisão, terminologia, linguagem, estilo e erro funcional* — e três níveis de gravidade — *menor, maior e crítico*, cujas definições seguem a descrição do modelo da TAUS. Há também a categoria *preferencial*, que não penaliza o tradutor e indica sugestões para melhoria do texto. Se o mesmo erro acontece mais de uma vez, o avaliador marca a caixa de repetição na aba de detalhamento, mas só a primeira ocorrência é penalizada.

Para extratos de texto de mil palavras, o máximo de pontos admitidos por categoria variava entre um e dois pontos de erro. Acima disso, a tradução era reprovada. A pontuação máxima permitida para o extrato como um todo era de quatro pontos de erro. Um erro crítico significava reprovação imediata.

Em sua análise comparativa de modelos de avaliação, O'Brien (2012) identificou as categorias de erros mais comuns nos modelos de avaliação analisados: *linguagem, terminologia, precisão e estilo*. Linguagem abrange principalmente erros de gramática e sintaxe. Terminologia se refere à não observação dos termos do glossário do cliente ou da terminologia específica da área em questão, além de falta de consistência. Precisão trata de omissões e adições desnecessárias e erros de referência. Estilo se refere, principalmente, à falta de aderência ao guia de estilo do cliente.

Após a fase de análise comparativa dos modelos de avaliação, a autora estabeleceu contatos com representantes de empresas filiadas à TAUS para que definissem os tipos de conteúdo que eram mais importantes para cada uma e depois atribuíssem a importância das variáveis *utilidade, tempo e sentimento*, consideradas aspectos chave para um novo modelo de avaliação. Utilidade se refere à importância da função do conteúdo traduzido; tempo, à velocidade em que o conteúdo precisa chegar ao mercado; sentimento, ao impacto que a tradução tem na imagem de uma marca ou produto. As informações obtidas na comparação e nessa etapa de contatos com empresas serviram de base para a proposta de uma nova abordagem, dinâmica, para a tarefa de avaliação de qualidade.

Essa abordagem dinâmica se baseava em três critérios para definição dos modelos de avaliação mais recomendados:

1. O canal de comunicação em que o conteúdo estaria disponível, tais como: comunicação interna, conteúdo da empresa para o consumidor ou da empresa para outra empresa, entre outros;
2. O perfil do conteúdo, tais como: textos para interface de usuário, material de treinamento, ou conteúdo para um website;
3. O ranking de *utilidade*, *tempo* e *sentimento*.

A análise desses três critérios, em conjunto com as variáveis utilidade, tempo e sentimento, é a dinâmica para a escolha os três principais modelos a serem usados na avaliação do conteúdo traduzido. No total, oito modelos foram propostos: aderência a instrumentos regulatórios, avaliação de usabilidade, taxonomia de erros, adequação/fluência, avaliação da comunidade, avaliação de facilidade de leitura, ranking de sentimentos em relação ao conteúdo (positivo ou negativo, por exemplo) e opinião dos consumidores.

Em minha experiência como tradutor profissional que trabalha com empresas internacionais prestadoras de serviços linguísticos e nas pesquisas posteriores que realizei para este trabalho, no entanto, não encontrei registro da adoção do modelo dinâmico da TAUS no mercado de tradução. A prática comum na indústria hoje é a realização da avaliação de qualidade em paralelo ao trabalho de tradução, com o uso de programas de verificação de erros autônomos ou nativos de ferramentas de tradução assistida por computador (Castilho et al., 2018), ou, após a etapa de tradução, com o uso de um modelo tradicional de avaliação baseada em erros, aplicado em amostras de texto traduzido.

Um dos modelos ainda adotado pela indústria é o da hoje extinta LISA, citado em Almeida (2013), criado na década de 1990. A categorização da LISA foi pensada para a avaliação de traduções para o mercado de localização e tinha como objetivo, por exemplo, definir se determinada tradução poderia seguir para outra etapa do ciclo de tradução ou avaliar se um tradutor estava apto a participar de determinado projeto de localização. A *dinâmica* do modelo previa o preenchimento de uma tabela que incluía penalizações pela gravidade do erro cometido (menor, maior ou crítico). Além da categoria *linguagem*, cujos detalhes apresentamos na tabela abaixo, o modelo de avaliação da LISA ainda incluía categorias específicas para a indústria de localização, tais como, *formatação de software*, *formatação de*

ajuda e teste de funcionalidade de software, pensadas para a avaliação da tradução de conteúdos para publicação na internet.

Categoria <i>linguagem</i>	
Subcategoria	Descrição
Erro de tradução	Tradução incorreta do texto-fonte
Precisão	Informações ausentes ou adicionadas na tradução, que não reflete o texto-fonte (mas não são um erro de tradução do sentido original)
Terminologia	Inadequação de terminologia/léxico ao contexto
Linguagem	Problemas relacionados a gramática, semântica, ortografia e pontuação
Estilo	Não-aderência ao guia de estilo do projeto
País	Padrões de país incorretos, tais como moeda e separadores decimais
Consistência	Falta de padronização da terminologia usada no texto

Quadro 1. Subcategorias de *linguagem* no modelo da LISA (Almeida, 2013)

Um outro modelo de categorização relevante é a estrutura MQM (*Multidimensional Quality Metrics*; “métricas de qualidade multidimensionais”) (Lommel et al., 2014). A MQM tem como característica principal a granularidade, ou seja, é um modelo que conta com um grande número de categorias divididas em diversos níveis. A estrutura parte de onze categorias principais, chamadas de categorias raiz, que se subdividem em diversos ramos, como se pode ver na lista disponível online⁸.

⁸ Disponível em: <<http://www.qt21.eu/mqm-definition/issues-list-2015-12-30.html>>. Acesso em: 10 ago. 2017.

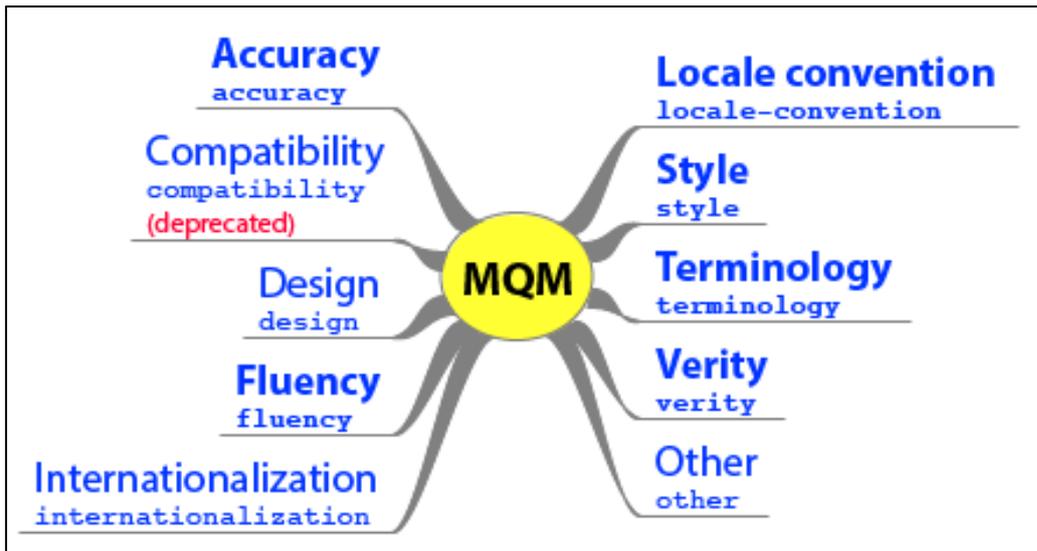


Figura 1. Classificação raiz da MQM (Lommel et al., 2014)⁹

No total, a lista abrange 114 categorias de erros de tradução, e todas contam com uma descrição detalhada a partir de links disponíveis na página da MQM. Lommel et al. (2014) aconselham que o nível de complexidade das métricas para avaliação humana esteja de acordo com as exigências da tarefa proposta e afirmam, corroborando a teoria de categorização de Ellis (1993), que quanto maior o número de categorias, maior a tendência de desacordo entre avaliadores e anotadores. A finalidade principal da estrutura MQM é servir como ponto de partida para a criação de modelos de avaliação de qualidade de tradução personalizados, ou seja, adaptados aos objetivos de diferentes projetos de avaliação de tradução, tais como, por exemplo, a avaliação de TA.

4.2. Estudos avaliativos

O primeiro estudo com resultados relevantes para este trabalho é Bowker (2005). Embora não tenha como objeto de estudo sistemas de tradução automática, mas de memória de tradução, o experimento mostra como o uso de tecnologia de tradução pode impactar o desempenho de tradutores e a qualidade da tradução. O objetivo do jogo era traduzir um texto de 387 palavras do francês para o inglês em até 40 minutos. O objeto era um anúncio de emprego na área de tradução. Os

⁹ Nomes das categorias em português (sentido horário, a partir do alto, à dir.): precisão, convenção local, estilo, terminologia, veracidade, outros, internacionalização, fluência, design, compatibilidade (abandonada).

jogadores, nove alunos de tradução da Escola de Tradução e Interpretação da Universidade de Ottawa, no Canadá, foram divididos em três grupos de três tradutores.

A dinâmica foi a seguinte. O grupo A fez a tradução a partir do zero, sem qualquer correspondência de memória de tradução (MT), o grupo B usou uma MT com conteúdo de traduções profissionais e o grupo C, uma MT semelhante à do grupo B, porém com segmentos deliberadamente alterados para conter erros de tradução.

Dois dos participantes do grupo A não conseguiram concluir a tarefa no prazo estipulado e o terceiro produziu uma versão não revisada, em que restavam várias dúvidas, notadamente terminológicas, marcadas no texto. Em termos de tempo, apenas um participante estava dentro da média de 580 palavras por hora prevista para a tarefa. Nenhum dos estudantes do grupo A cometeu erros graves e não houve omissões ou adições, embora em todas as traduções houvesse frases com problemas de construção (gramaticalmente corretas, mas pouco fluentes).

Todos os participantes do grupo B conseguiram concluir a tarefa. Dois deles terminaram antes do tempo, obtendo altos índices de produtividade (1.581 e 1.338 palavras por hora, respectivamente). Em termos de qualidade, no entanto, foram encontrados problemas de paralelismo e de inconsistência na terminologia, devido a diferenças entre os segmentos de correspondência parcial e correspondência total contidos na memória de tradução, como, por exemplo, o mesmo termo traduzido de formas distintas em segmentos diferentes e uma lista de qualificações que misturava verbos e substantivos.

Os participantes do grupo C trabalharam em uma memória deliberadamente alterada para conter erros de tradução. Cinco erros foram incluídos em segmentos de correspondência total e cinco em segmentos correspondência parcial. Todos os participantes terminaram a tarefa no prazo estipulado: dois se consideraram satisfeitos com o resultado, enquanto o terceiro gostaria de ter tido mais tempo para revisar a tradução. Em termos de qualidade, apenas um dos três participantes conseguiu identificar erros em segmentos com correspondência total; mesmo assim, só corrigiu dois dos cinco segmentos incorretos. A quantidade de erros foi menor em segmentos com correspondência parcial: dois dos participantes deixaram de corrigir apenas um erro, enquanto o terceiro deixou de corrigir os erros contidos em dois segmentos.

Apesar de envolver poucos participantes, o estudo revela algumas tendências interessantes. Em termos de produtividade, fica claro que o uso de memória de tradução ajuda os tradutores a trabalhar mais rapidamente. A produtividade média dos seis participantes dos grupos que usaram memória de tradução foi 873 palavras por hora, muito acima da produção média de 335 palavras por hora relatada em pesquisa de 2012 da Common Sense Advisory¹⁰. Porém, o custo desse aumento de produtividade foi uma queda considerável na qualidade ao comparar as traduções desses grupos às realizadas pelos participantes do grupo A.

Bowker (2005) pondera que os estudantes dos grupos A e B tenderam a aceitar sem questionar a qualidade das sugestões da memória de tradução, uma tendência que se mostrou ainda mais nítida quando os resultados eram de correspondência total. Ao editar segmentos com correspondências parciais, os participantes confiaram menos nas sugestões e acabaram deixando passar menos erros. Autores como Way (2018) extrapolam a ideia da menor confiança de tradutores nas sugestões de correspondência parcial de MT e apontam para a ocorrência do mesmo fenômeno em relação a sugestões provenientes de sistemas de tradução automática.

Os resultados do jogo de avaliação de Guerberof (2014), no entanto, mostram que tradutores profissionais foram mais produtivos e cometeram menos erros ao editar segmentos provenientes de correspondências parciais (CP) de memória de tradução (85-94%) e de um sistema de TA baseado em Moses do que ao traduzir segmentos sem correspondência (SC), ou seja, sem sugestões de ferramentas de tradução assistida por computador. Guerberof (2014) avaliou de forma qualitativa e quantitativa o desempenho de 24 tradutores profissionais ao traduzir textos da área de TI na combinação inglês para espanhol. O grupo mesclava profissionais com e sem experiência em pós-edição.

A avaliação quantitativa se baseou na produtividade de cada tradutor, expressa em palavras por minuto, e foi medida por uma ferramenta que registrava o tempo gasto pelos participantes na tradução de cada um dos segmentos. A avaliação quantitativa foi feita por três revisores profissionais com pelo menos três anos de experiência em localização, familiaridade com ferramentas de tradução assistida por computador e pelo menos seis meses de experiência com pós-edição.

¹⁰ Disponível em: <<https://www.common senseadvisory.com/default.aspx?Contenttype=ArticleDet&tabID=64&moduleId=392&Aid=2874&PR=PR>>. Acesso em: 24 set 2018.

A qualidade das traduções foi medida pelo total de erros encontrados no texto-alvo pela equipe de revisão.

O sistema de TA do experimento foi treinado com uma memória de tradução e três glossários de um fornecedor de serviços de gestão de cadeia de produção. A mesma memória foi usada para a geração das correspondências parciais. A tradução foi realizada em uma plataforma online, que não revelava aos participantes a origem do texto pré-traduzido para evitar qualquer possibilidade de preconceito dos participantes em relação à tradução automática.

Guerberof (2014) afirmou que o experimento foi projetado para se aproximar ao máximo das condições de trabalho normais de um profissional da tradução. No entanto, os participantes não podiam revisar os segmentos traduzidos após apertar o botão que levava ao segmento seguinte, o que impede uma discussão sobre o efeito de uma segunda revisão de uma saída de TA na qualidade final do documento. Além disso, o texto foi apresentado aleatoriamente, de forma não linear. Embora o estudo afirme que não é incomum que tradutores recebam segmentos sem continuidade em situações de trabalho reais, a contextualização certamente causaria um impacto positivo na qualidade do texto final.

Na avaliação de produtividade, os segmentos produzidos por TA tiveram o melhor tempo médio de processamento, seguidos pelas correspondências parciais e, por fim, pelos segmentos sem correspondência. Todos os 24 tradutores testados tiveram um ganho de produtividade ao usar TAC. Destes, 18 tiveram maior ganho médio de produtividade com os segmentos produzidos por tradução automática, e seis com as correspondências parciais. As médias de produtividade por categoria, convertidas para palavras por hora, são exibidas na tabela abaixo.

Categoria	Média estimada	Mais baixa	Mais alta
CP	957	827,4	1.107
TA	1.032,6	892,8	1.194,6
SP	647,4	559,8	749,4

Tabela 1. Produtividade média por categoria (pal./h) (Guerberof, 2014)

A avaliação de qualidade mostrou que os tradutores cometeram mais erros ao traduzir do zero do que quando tiveram alguma sugestão de sistemas de auxílio ao tradutor (27,28% dos segmentos com erros no primeiro caso, 18,17% e 18,11 nos últimos). Além disso, a proporção de erros foi muito semelhante ao editar o texto

proveniente da saída de TA ou de correspondências parciais, um resultado que aponta para uma relação positiva entre qualidade e produtividade em ambas as categorias.

<i>Categoria</i>	<i>Sem erros</i>		<i>Com erros</i>	
	<i>Segmentos</i>	<i>%</i>	<i>Segmentos</i>	<i>%</i>
<i>CP</i>	2.946	81,83	654	18,17
<i>TA</i>	2.889	81,89	639	18,11
<i>SC</i>	2.618	72,72	982	27,28
<i>Todas</i>	8.453	78,79	2.275	21,21

Tabela 2. Porcentagem de erros por categoria (Guerberof, 2014)

Os resultados do estudo indicam que o uso de tradução automática e memória de tradução ajuda os tradutores a aumentar a produtividade. A média de economia de tempo foi da ordem de 32% para a categoria correspondências parciais e de 37% para tradução automática. Quanto à qualidade, os tradutores cometeram mais erros nos segmentos sem correspondência, em que tiveram que fazer a tradução do zero. Os segmentos com a saída da TA e a correspondência parcial de memória de tradução mostraram maior qualidade que os segmentos sem correspondência e tiveram um percentual de erros semelhante entre ambos.

Por fim, Guerberof (2014) sugere, por um lado, que a menor qualidade das traduções da categoria sem correspondência talvez se deva ao fato de não haver possibilidade de revisão da tradução após o envio do segmento no sistema. Por outro, afirma que o bom desempenho dos tradutores nos segmentos das categorias TA e CP pode estar atrelado ao bom nível das sugestões exibidas pelo sistema.

A se comparar os resultados das categorias TA e CP com os números obtidos nas categorias com memória de tradução de Bowker (2005), a hipótese da influência da qualidade das sugestões na qualidade do texto-alvo faz sentido. Uma memória de tradução de alto nível causou um impacto positivo na tradução, ao passo que uma MT que gerava resultados com problemas de consistência ou deliberadamente incorretos causou um impacto negativo. É preciso, no entanto, levar em conta que os participantes do estudo de Bowker (2005) eram estudantes com pouca experiência de tradução, ao passo que o grupo de Guerberof (2014) era composto por tradutores profissionais.

Almeida (2013), por sua vez, mesclou profissionais com muita e pouca experiência no mercado em um estudo voltado para a análise de desempenho de

tradutores. O estudo buscou traçar um amplo perfil dos participantes, correlacionando as alterações feitas no processo de pós-edição a nível de experiência, produtividade, qualidade do trabalho e também à opinião de cada tradutor sobre tradução automática e pós-edição, colhida em questionário.

O grupo de participantes era composto por oito tradutores do par inglês-francês e dez do par inglês-português. A dinâmica se baseou na pós-edição de 74 segmentos traduzidos automaticamente nos dois pares acima citados, que somavam 1.008 palavras no total. Os textos eram da área de TI e o tipo de pós-edição escolhido foi a PE intermediária, significando que o texto não precisava ter qualidade similar à de uma tradução humana, como na PE completa, mas exigia algum nível de ajuste em termos de fluência, em vez de apenas a correção de erros, como na PE ligeira. Os segmentos foram apresentados de forma aleatória e a tarefa devia ser cumprida em até duas horas.

Em termos quantitativos, o estudo revelou que a produtividade média dos tradutores do par inglês-português, calculada em função do tempo gasto para a conclusão da tarefa, foi de 827 palavras por hora. Os resultados comparativos do estudo encontraram uma correlação fraca entre o nível de experiência em tradução ou em pós-edição e o tempo despendido para concluir a tarefa, indicando apenas uma leve tendência de que tradutores mais experientes concluíssem a tarefa em menos tempo do que tradutores menos experientes.

A seção qualitativa de Almeida (2013) se baseou na categorização das alterações realizadas pelos participantes. A definição das categorias em uma taxonomia de alterações foi feita pela autora do estudo, que também foi responsável pela avaliação das alterações e classificação dentro das categorias taxonômicas após a etapa de pós-edição. A taxonomia do estudo se divide em três níveis: categorias mestras, categorias principais e subcategorias vinculadas às categorias principais. As quatro categorias mestras se baseiam da natureza das alterações feitas na pós-edição e foram definidas como segue (Almeida, 2013, p. 100-1):

- *Alterações essenciais*: quando não são feitas, a frase (ou parte dela) fica:
a) gramaticalmente incorreta; ou b) gramaticalmente correta, mas imprecisa em relação ao conteúdo do texto-fonte.
- *Alterações preferenciais*: ocorrem quando a frase produzida pela TA já era gramaticalmente correta, inteligível e precisa em relação ao conteúdo do texto-fonte antes de qualquer alteração.

- *Alterações essenciais não implementadas*: por não terem sido feitas, a frase permaneceu gramaticalmente incorreta ou imprecisa.

- *Erros introduzidos*: ocorrem quando o tradutor insere na frase produzida pela TA um erro que não havia antes e que tornou a frase gramaticalmente incorreta ou imprecisa.

As categorias mestras revelam, assim, o volume de correções necessárias e desnecessárias feitas pelos tradutores. A tabela abaixo mostra o número médio de correções realizadas por categoria mestra nas duas combinações de idiomas.

Alterações essenciais	Alterações preferenciais	Alterações essenciais não implementadas	Erros introduzidos
147,22	45,16	31,66	11,5

Tabela 3. Média de alterações por categoria (Almeida, 2013)

Os resultados do estudo mostram que o número de alterações essenciais, ou seja, necessárias para tornar a frase gramaticalmente correta ou inteligível, respondem por pouco mais da metade do total de todas as categorias (62,6%). É preciso, no entanto, relativizar os números para se ter uma visão mais precisa. Se considerarmos as alterações feitas de forma ativa pelos participantes (excluindo a categoria *Alterações essenciais não implementadas*), temos que as alterações desnecessárias equivalem a mais de um quarto do total (27,9%). Esses números corroboram duas afirmações de Way: a primeira, de que revisores humanos muitas vezes se sentem “compelidos a fazer alterações desnecessárias” (Way, 2018, p. 8) e que não existe um único nível de qualidade possível para traduções, qual seja, “traduções humanas ‘perfeitas’” (idem, p. 9). Embora sejam consideradas como tal quando usadas como referência para a avaliação automatizada de TA, traduções humanas contêm erros, e a revisão de traduções também.

Além das quatro categorias mestras, a taxonomia tem oito categorias principais e 21 subcategorias. Embora a autora tenha expressado a preocupação de encontrar um equilíbrio para que não houvesse categorias demais que “pudessem tornar a análise menos clara” (Almeida, 2013, p. 79), seguindo os princípios subjacentes ao processo de categorização de Ellis, nos parece que a taxonomia tem alta granularidade (33 categorizações), o que poderia dificultar a categorização de outros estudos que usassem outros avaliadores além da autora.

Em termos dos perfis dos participantes, a autora avalia que os resultados do estudo indicaram que os participantes com maior nível de experiência tendem a realizar mais *Alterações essenciais* e menos *Alterações preferenciais*. Comenta também que os dois participantes do par inglês-português brasileiro que efetuaram mais alterações preferenciais não estavam no grupo dos tradutores mais experientes, mas expressaram opiniões negativas ou neutras sobre TA e PE, o que parece indicar uma correlação entre essas opiniões e o número de alterações.

Em geral, o estudo não registrou nenhuma forte correlação entre tempo despendido, qualidade na pós-edição, nível de experiência ou opinião sobre TA dos tradutores, o que parece demonstrar, em nossa opinião, que os resultados são mais um retrato das características de cada tradutor do que de qualquer impacto da pós-edição sobre a qualidade final do texto traduzido.

O estudo de Bentivogli et al. (2016a) comparou o desempenho quali-quantitativo de tradutores em uma tarefa de pós-edição de TA nos pares inglês-francês e inglês-italiano, realizada em dois sistemas distintos, como segue: i) sistema Moses baseado em frases, chamado de sistema estático; ii) sistema adaptativo, que se adapta logo após a edição de um segmento pelo usuário. “Informações monolíngues e bilíngues são extraídas de cada segmento pós-editado e injetadas no sistema adaptativo” (Bentivogli et al., 2016a, p. 390). A ferramenta de tradução assistida escolhida foi o MateCat, que abrigou tanto o sistema estático quanto a TA adaptativa.

Os participantes do experimento foram 16 tradutores profissionais, contratados através de uma agência de tradução para realizar a pós-edição de textos dos domínios jurídico e TI. Em termos de ambiente, os participantes trabalharam remotamente e usaram uma grande memória de tradução pública em modo somente leitura e uma memória particular, na qual as pós-edições eram gravadas. Cada tradutor realizou duas sessões de pós-edição do mesmo texto (uma no sistema estático, outra no sistema adaptativo).

Duas variáveis de avaliação foram usadas: i) uma quantitativa, o esforço de pós-edição, calculado segundo a métrica HTER (Snover et al., 2006)¹¹, que usa a distância de edição entre a saída de TA original e o segmento pós-editado para avaliar quantitativamente o desempenho dos sistemas; e ii) uma qualitativa,

¹¹ HTER é uma variante da métrica TER que se usa anotadores humanos para determinar a menor taxa de distância de edição possível.

realizada por 32 tradutores humanos. A hipótese do estudo era que, com o passar do tempo, o sistema de TA adaptativo geraria sugestões de tradução de maior qualidade que o sistema estático, reduzindo o esforço de pós-edição.

Na determinação do esforço de pós-edição, o estudo fez um cálculo estatístico para descontar os efeitos de ordem e de tradução automática nos resultados. O efeito de ordem está relacionado ao fato de que cada tradutor executou a mesma tarefa de edição duas vezes (editou o mesmo texto), a primeira no sistema estático, a segunda no adaptativo. O efeito de TA é o fator de contraste relacionado ao uso de diferentes sistemas, e significa o cálculo estatístico da redução de esforço quando se compara os resultados em HTER dos sistemas estático e adaptativo. Considerando o efeito de TA, o ganho do sistema adaptativo sobre o estático no domínio de TI foi de 10,57%, um valor estatisticamente relevante, enquanto o ganho de 5,66% em JU foi considerado irrelevante. O resultado geral, que considera os dois domínios, apesar da diferença de relevância entre ambos, apontou para uma redução da ordem de 70% no esforço de pós-edição.

Na etapa seguinte do estudo, uma avaliação qualitativa foi realizada para confirmar se o ganho em esforço de PE se devia à melhor qualidade dos resultados apresentados pelo sistema de TA adaptativa. A análise foi feita em duas etapas: na primeira, avaliou-se a saída da TA sem edição, na segunda, a pós-edição feita pelos tradutores. Os segmentos traduzidos foram avaliados por tradutores especializados em pelo menos um dos domínios do estudo, que atribuíram notas em uma escala de 1 a 5 a cada segmento traduzido: 1 significava que a TA era “imprestável” e 5, que tinha “qualidade humana”. Cada tradutor avaliou cada segmento do texto-fonte uma vez, com o texto-alvo variando entre a saída da TA e pós-edições de todos os participantes do domínio em questão. O estudo revela, porém, que houve grande inconsistência nas notas 1 e 2 atribuídas a segmentos jurídicos, por isso, decidiram unir essas notas em um único nível 2, para evitar distorções.

A avaliação de qualidade da saída da TA indicou uma qualidade média de 3,01 para a TA estática e de 3,07 para a TA adaptativa na junção dos dois domínios, o que significou um ganho de qualidade de 1,99% para a segunda, considerado estatisticamente relevante. Esse ganho de qualidade, por si só, gerou uma redução de 9,17% no esforço de PE. O estudo afirma que há grande variação da pontuação entre segmentos e entre avaliadores, provavelmente devido às dificuldades intrínsecas das sentenças e do gosto pessoal dos avaliadores, mas conclui que a

“redução observada claramente indica que a TA adaptativa pode ajudar a melhorar a produtividade e a experiência dos usuários” (Bentivogli et al., 2016a, p. 397).

Costa et al. (2014) avaliaram o desempenho de dois sistemas, Google Tradutor e Moses, na tradução de três tipos de textos, como segue: cinquenta frases extraídas de um TED Talk de Barry Schwartz, chamado *On our loss of vision*; cinquenta extraídas da revista UP Magazine, da companhia aérea TAP; cinquenta de um corpus disponibilizado por Li e Roth, da coleção TREC, chamado de *Questions* no artigo. A avaliação é baseada na quantificação dos erros de tradução cometidos pelos sistemas. A classificação dos erros foi feita por um falante nativo de português de acordo com a taxonomia abaixo.

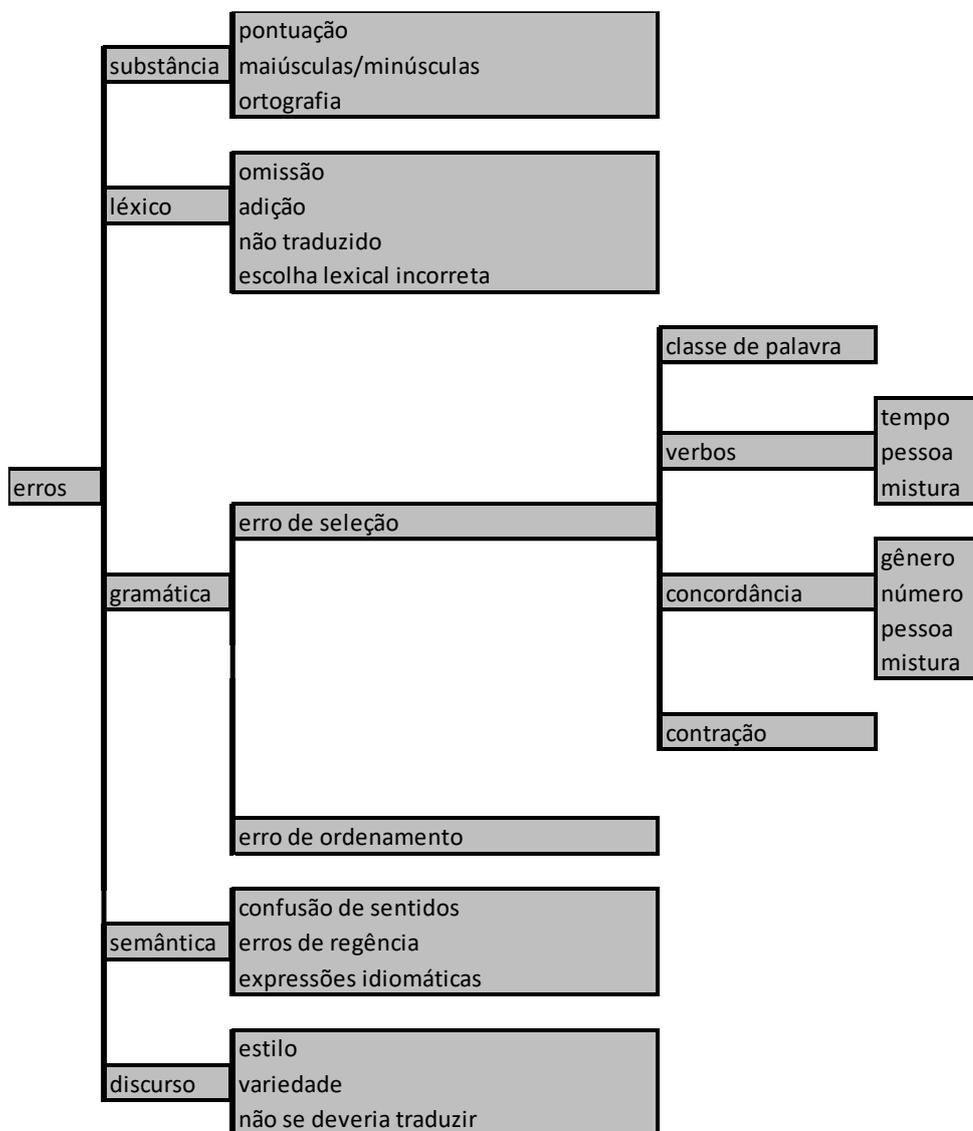


Figura 2. Taxonomia de anotação de erros (Costa et al., 2014)

A categorização dos erros foi dividida em níveis, como se vê na figura acima. O nível *substância* abrange erros de pontuação, maiúsculas e minúsculas e ortografia. O nível *léxico* se refere à forma como uma palavra é traduzida. O nível *gramática* trata de erros morfossintáticos. O nível *semântica* abrange erros de sentido. Por fim, o nível *discurso* trata de escolhas estilísticas e palavras/termos que não deveriam ter sido traduzidos. Neste nível, o sentido é preservado, mas há inadequação de variedade ou registro do texto-alvo.

	<i>Moses/GT</i>			
	TAP	Questions	TED	Total
<i>Substância</i>	4/1	0/0	4/2	8/3
<i>Léxico</i>	256/50	73/26	109/34	438/110
<i>Gramática</i>	96/28	26/16	87/43	209/87
<i>Semântica</i>	35/26	11/16	40/20	86/62
<i>Discurso</i>	8/17	0/4	2/33	10/54

Tabela 4. Erros cometidos por Moses e GT (Costa et al., 2014)

O estudo revelou que o Google Tradutor (GT) obteve melhor desempenho geral, superando o sistema baseado em Moses na maioria dos níveis. O GT foi superior em *gramática* e *léxico*, com melhor desempenho em todos os tipos de texto. Em *substância*, o GT foi superior nos corpora TAP e TED, e ambos os sistemas não cometeram erros no corpus Questions. Em *semântica*, o Moses foi superior nos corpora TAP e Questions. O maior ponto fraco do Moses foi no léxico, o sistema não sabia como traduzir muitas palavras. O GT teve problemas no nível *semântico*. Por fim, no nível *discurso*, o Moses foi superior ao GT em todos os corpora. Os autores apontam que isso se deve, em grande parte, ao fato de o sistema usar vocabulário e formas gramaticais em português brasileiro, considerados como erros no estudo. Esse tipo de distorção demonstra a importância de se avaliar todos os aspectos dos resultados obtidos para se obter um resultado mais consistente.

Martins e Caseli (2013) definiram uma taxonomia de erros e regras de anotação para categorização de erros de TA em frases traduzidas do inglês para o português brasileiro. As cerca de 1.300 sentenças anotadas formaram um corpus de treinamento para uso na construção de um sistema de pós-edição automática de erros de TA com o objetivo de: “(i) identificar automaticamente o que é um erro de TA e (ii) realizar a correção automática do erro previamente identificado”.

Categoria	Subcategoria	Descrição
A. Erros morfosintáticos	A.1. Concordância em número	Erros de concordância de singular e plural
	A.2. Concordância em gênero	Erros de concordância para masculino e feminino
	A.3. Flexão verbal	Verbos que apresentam conjugação ou forma verbal incorreta
	A.4. PoS (Part of Speech)	Mudança de categoria. Ex.: de verbo para substantivo
	A.5. Outros	Erros em traços morfológicos que não se enquadram em nenhuma das subcategorias acima
B. Erros lexicais	B.1. Palavra extra	Uma tradução na saída da TA que não tem nenhuma correspondência na sentença original
	B.2. Palavra ausente	Uma palavra presente na sentença original (Fonte) para a qual a tradução não está presente na saída da TA
	B.3. Palavra não traduzida	Uma palavra presente na sentença original (Fonte) que é mantida igual na saída da TA, ou seja, a palavra não foi traduzida
	B.4. Palavra incorretamente traduzida	Uma tradução correspondente ocorreu, mas está incorreta
	B.5. Palavra com erro de grafia	Uma tradução correspondente ocorreu na TA e a escolha da palavra está correta, mas sua grafia não
C. N-grama errado	C.1. expressão extra	Uma tradução na saída da TA que não tem nenhuma correspondência na sentença original
	C.2. expressão ausente	Uma expressão presente na sentença original para a qual não há tradução na saída da TA
	C.3. expressão não traduzida	Uma expressão presente na sentença original que é mantida igual na saída da TA
	C.4. expressão incorretamente traduzida	Uma tradução correspondente ocorreu, mas está incorreta
D. Ordem errada		
E. Erros fonte-alvo		

Tabela 5. Taxonomia para anotação de erros de TA (Martins e Caseli, 2013)

O processo de anotação¹², realizado por dois anotadores brasileiros com bons conhecimentos do inglês, foi descrito em Aziz e Specia (2011 apud Martins e Caseli, 2013). Em um teste realizado com 126 sentenças anotadas em paralelo por ambos, chegou-se a um nível de concordância de 63%, calculado segundo os erros marcados exatamente da mesma forma, “ou seja: mesma categoria e mesmas palavras nas sentenças fonte (src), referência (ref) e alvo (sys)” (Martins e Caseli, 2013, p. 17).

As autoras consideraram o nível satisfatório para a complexidade da tarefa e o rigor do cálculo de concordância, mas, como o processo de anotação prosseguiu com cada anotador anotando sozinho metade das sentenças restantes, talvez fosse mais recomendável rever a taxonomia em busca de uma forma de obter maior concordância. A chave da questão pode estar no nível de especificidade esperado pelas autoras para a tarefa de anotação — mesma categoria, mesmas palavras nos textos fonte, alvo e de referência —, que dificulta a concordância. Afinal, o ponto fundamental das categorias linguísticas é “reduzir a complexidade do mundo para que [...] os falantes possam se orientar nelas, ou seja, conhecê-las” (Ellis, 1993, p. 33).

Ao aumentar a complexidade da tarefa de anotação, as autoras reduziram o grau de consenso entre os anotadores. Seria recomendável pensar em uma nova categorização, menos rígida, para produzir maior consenso e gerar uma taxonomia com mais categorias abrangentes e, conseqüentemente, mais úteis para o objetivo de se produzir um sistema de pós-edição automática de erros de TA.

Segundo o estudo, das 1.314 sentenças anotadas no total, 435 (33,1%) estavam corretas. As 879 restantes (66,9%) continham pelo menos um erro. Os resultados mostram um predomínio de *erros lexicais* (44,52% do total), principalmente nas subcategorias *palavra ausente* (14,9%, maior porcentagem geral), *palavra incorretamente traduzida* (14,17%, segunda posição) e *palavra não traduzida* (9,92%, quinta posição). A primeira subcategoria trata de termos que estavam presentes no texto-fonte e não apareceram no texto-alvo, a segunda, de erros de tradução propriamente ditos, a terceira, de palavras mantidas no texto-alvo exatamente como estavam no texto-fonte.

¹² Anotar significa classificar os erros encontrados em uma frase segundo a taxonomia dada.

A segunda categoria em que mais foram encontrados erros foi a de *erros morfossintáticos*, que respondeu por 38,39% das ocorrências. As duas subcategorias que mais se destacaram foram as relacionadas a *concordância de número* e *concordância de gênero*, com 13,94% e 12,48% dos erros, ocupando a terceira e a quarta posições no ranking, respectivamente. Os problemas, acreditamos, espelham diretamente a dificuldade de tratar a concordância (e a *flexão verbal*, outra subcategoria com alta incidência de erros) na tradução entre uma língua, o inglês, que não marca gênero e número, e outra em que ambos são marcados, o português, como de resto ocorre com todas as línguas latinas. Por fim, vale mencionar que a categoria *n-grama errado* pode ser considerada irmã da categoria *erros lexicais*, pois tem as mesmas subcategorias que esta, porém, com marcação de mais de uma palavra na anotação. Se somássemos ambas, veríamos que os erros relacionados a traduções incorretas ou ausência e presença de palavras/expressões no texto-alvo responderiam por mais da metade do total (52,43%).

Por fim, acreditamos que o corpus escolhido pode aumentar as dificuldades de treinamento do sistema de TA. Consideramos que o corpus a ser usado para o estudo deveria ter cunho generalista, para dar conta de todas as categorias contempladas. Porém, o corpus utilizado, chamado Corpus FAPESP, foi compilado a partir de “cerca de 2700 documentos paralelos, totalizando mais de 150 mil sentenças alinhadas” (Aziz e Specia, 2012), provenientes de artigos jornalísticos de conteúdo científico da revista eletrônica FAPESP Online¹³. O treinamento com domínios muito específicos pode dificultar o tratamento de linguagem mais geral pelo sistema. “Se pretendemos ser realmente científicos sobre o estudo da linguagem, melhor não usar termos científicos como exemplos” (Ellis, 1993, p. 36). Categorizações complexas demais oferecem pouco conteúdo informativo, categorizações menos complexas permitem categorias mais abrangentes, com níveis de informação *gerenciáveis* e *úteis* para processos como o treinamento de sistemas de TA.

¹³ Disponível em: <<http://revistapesquisa.fapesp.br/>>. Acesso em: 6 out. 2018.

4.3. Resumo

Vimos, neste capítulo, alguns exemplos de modelos de avaliação de tradução e de tradução automática baseados em taxonomias de erros. O'Brien (2012) passou em revista vários modelos de avaliação usados na indústria de tradução para propor uma avaliação dinâmica, que levasse em conta o tipo de conteúdo, aspectos como a utilidade de texto, o uso esperado e o tipo de sentimento que deveria suscitar no público-alvo, bem como os canais de comunicação utilizados, para ditar que tipo de modelo seria mais adequado para avaliar cada situação específica.

O modelo de avaliação da LISA, que tem sido um dos principais padrões do mercado de tradução, conforme descrito em Almeida (2013), foi apresentado neste capítulo. Criado na década de 1990, foi pensado para o mercado de localização, com categorias voltadas para a publicação de conteúdo na internet, por exemplo, e previa o preenchimento de uma tabela de erros que incluía também uma pontuação baseada na gravidade dos erros. Por fim, A estrutura MQM (Lommel et al., 2014) representou uma tentativa de oferecer uma categorização de erros de tradução abrangente, com diversos níveis de categorias, que pudesse ser adaptada para o tipo de uso buscado por uma empresa ou um avaliador.

Também foram apresentados diversos exemplos de jogos de avaliação: de tradução em si, de tradução automática e de pós-edição de TA. O quadro nas páginas seguintes apresenta um panorama dos principais aspectos abordados nos trabalhos aqui citados.

	Guerberof (2014)	Bentivogli et al. (2016a)	Costa et al. (2014)	Martins e Caseli (2013)	Almeida (2013)
Objetivo	Comparar desempenho de TA, correspondências parciais (85-94%) e TH	Comparar sistemas de TA estática e adaptativa	Avaliar desempenho de dois sistemas de TA	Anotação de erros de TA; criação de corpus de treinamento p/ ferramenta de PE automática	Avaliação de desempenho de pós-editores
Tipo de texto	TI	Jurídico e TI	Transcrição de fala, jornalístico, perguntas	Jornalístico	TI
Corpus	MT e 3 glossários de gestão de cadeia de produção	MTs: pública (somente leitura) e particular (para gravação da PE)	MTs: Europarl, TED Talk, Up Magazine, coleção TREC	Artigos jornalísticos de conteúdo científico: revista eletrônica FAPESP Online	MT: documentação de software da Autodesk
Objeto	Pós-edição de TA e MT, tradução	Saída de TA estática e TA adaptativa	Saída de TA dos dois sistemas	TA sem edição	TA sem edição
Perfil dos participantes	24 tradutores (PE); 3 revisores (avaliação)	16 tradutores (PE); 32 avaliadores	1 linguista (avaliação)	2 anotadores	20 participantes (10 por par de idiomas)
Idiomas	EN→ES	EN→FR, EN→IT	EN↔PTPT	EN→PTBR	EN→PTBR; EN→FR
Sistemas	Moses (sugestões de TA), interface própria (edição)	MateCat (TA estática e TA adaptativa)	Moses, GT	Moses	Post-Editing Workbench (Autodesk)
Volume	Total de palavras: 2.124 (TA: 757, FM: 618, NM: 749)	Total (pal.): 13.924 3.284 (piloto/TI) 3.874 (teste/TI) 3.125 (piloto/JU) 3.641 (teste/JU)	Em tokens do texto-fonte: TAP, 984; TED, 844; Questions, 377	1.314 sentenças anotadas	2.119 palavras

	Guerberof (2014)	Bentivogli et al. (2016a)	Costa et al. (2014)	Martins e Caseli (2013)	Almeida (2013)
Avaliação humana	Anotação de erros segundo taxonomia da LISA	Atribuição de notas por segmento (escore 1 a 5)	Anotação de erros segundo taxonomia própria	Anotação de erros segundo taxonomia própria	Registro de pós-edições segundo taxonomia própria; questionário para tradutores
Taxonomia: principais categorias	Modelo da LISA	N/A	Substância, léxico, gramática, semântica, discurso	Erros morfosintáticos, erros lexicais, N-grama errado, ordem errada, erros fonte-alvo	Alterações essenciais ou preferenciais, alterações essenciais não implementadas, erros introduzidos
Resultado: AH	Qualidade: correspondências parciais e TA tiveram qualidade significativamente maior do que TH	TA adaptativa: ganho de 1,34% em qualidade sobre TA estática (estatisticamente irrelevante)	GT superior nos níveis substância, léxico e gramática, Moses superior no nível discurso	EM: 38,39%, EL: 44,52%, NG: 7,91%, OE: 9,19%, FA: 0%	Comparações entre desempenho dos tradutores e opinião dos participantes em questionário
Avaliação automatizada	Tempo de edição: calculado pela ferramenta de PE criada para a tarefa	Esforço de PE: métrica HTER	N/A	N/A	Contagem de tempo, uso de mouse/teclado, contagem de pesquisa de termos online
Resultado: AA	Ganho médio de produtividade sobre TH: correspondências parciais, 32%; TA, 37%	TA adaptativa: redução de 70% no esforço de PE em relação à TA estática	N/A	Classificação dos erros	Contagem e classificação dos erros; questionário para os participantes

Quadro 2. Características dos estudos citados neste trabalho.

Considerando cada estudo como um jogo de avaliação, podemos fazer o seguinte resumo:

Guerberof (2014) teve como jogadores 24 tradutores da combinação inglês para espanhol fazendo lances: i) de pós-edição, após receber segmentos gerados por TA ou memória de tradução, e ii) de tradução de trechos sem sugestão. A dinâmica tinha como finalidade que os tradutores não soubessem de onde vinha o lance que recebiam. O objetivo foi confirmar se a tradução automática gerava economia de tempo (produtividade) sem perda de qualidade do texto pós-editado.

Bentivogli et al. (2016a) fizeram um jogo de pós-edição cujos lances eram dados por tradutores dos pares inglês-francês e inglês-italiano e cuja dinâmica ocorreu em uma ferramenta de tradução assistida por computador. O objetivo foi avaliar o esforço de pós-edição dos tradutores e a qualidade da tradução realizada em dois sistemas de TA, um estático e outro adaptativo.

Martins e Caseli (2013) e Costa et al. (2014) descrevem jogos de avaliação com criação de taxonomias de erros em traduções automáticas do inglês para o português (do Brasil e de Portugal, respectivamente). O primeiro teve como objetivo apontar caminhos para o aprimoramento da TA; o último, comparar o desempenho de dois sistemas de TA: Moses e Google Tradutor.

Almeida (2013) teve como jogadores tradutores nos pares inglês-francês e inglês-português brasileiro. Os lances foram dados pela pós-edição de TA. Na dinâmica de Almeida (2013), importaram mais os resultados dos lances de PE do que a saída de TA inicial, e o objetivo foi entender a relação de tradutores profissionais com a pós-edição de TA e com a TA em si.

No próximo capítulo, vamos propor novos jogos de avaliação de tradução automática, centrados na figura do tradutor profissional.

5 Três jogos de avaliação

Este capítulo descreve três exercícios de avaliação — ou jogos — de tradução automática de pedidos de patente do inglês para o português. Sou o principal jogador nas três situações de avaliação: um jogo de pós-edição, um jogo de comparação de traduções e um jogo de comparação de tradutores. Em vista dessa dinâmica e, conforme discutido anteriormente, da abordagem não-logocêntrica do meu trabalho, adotei a narração em primeira pessoa em diversos trechos da descrição dos jogos, de modo a transmitir de maneira mais clara a minha visão, enquanto tradutor profissional, sobre distintos aspectos da avaliação de TA. A tabela abaixo resume as características dos jogos realizados.

Jogo	Dinâmica	Sistema	Finalidade
1	Pós-edição de TA; avaliação	Lilt Google	Avaliação de produtividade; avaliação de qualidade
2	Pós-edição <i>versus</i> memória de tradução	Lilt TP	Avaliação de qualidade; impacto no estilo
3	Pós-edição <i>versus</i> tradução comercial	Lilt Google TP	Avaliação de qualidade

Tabela 6. Descrição dos jogos de avaliação

O objetivo do jogo 1 é avaliar como o uso de TA em uma tarefa de pós-edição pode impactar a produtividade, comparando os resultados obtidos à tradução humana sem uso de ferramentas computadorizadas e aos resultados obtidos em estudos relacionados, conforme descritos no capítulo anterior. Para oferecer uma visão qualitativa, mais detalhada, do desempenho dos sistemas de TA, organizei as alterações feitas ao longo do processo de PE em uma taxonomia de edições¹⁴.

O jogo 2 compara duas traduções de um mesmo pedido de patente, feitas em dois momentos distintos: a primeira, realizada por volta de 2011, foi uma tradução

¹⁴ Decidi adotar, nesses jogos, o conceito de taxonomia de edições para se referir às mudanças feitas no texto para adequação do conteúdo ao uso esperado. Como a TA é considerada uma ferramenta de tradução assistida por computador, que oferece *sugestões* ao tradutor, não há sentido em falar em erro.

profissional (TP) produzida por mim com um sistema de memória de tradução; a segunda, produzida em 2018, foi feita para este trabalho em uma tarefa de pós-edição de TA. O objetivo aqui é avaliar o quanto a saída de TA pode influenciar o estilo e as escolhas terminológicas de um tradutor e o que essa influência significa em termos de qualidade final da tradução.

O jogo 3 compara a tradução final dos dois documentos pós-editados no jogo 1 à tradução profissional dos pedidos de patentes correspondentes, conforme encontrados no banco de dados do INPI. O objetivo foi avaliar se o estilo de pós-edição adotado nesse jogo, qual seja, a PE completa, está de acordo com as expectativas de qualidade para a finalidade de depósito de um pedido de patente.

5.1. Jogo 1

O primeiro jogo é um exercício de pós-edição da tradução automática de pedidos de patente do inglês para o português brasileiro. O objetivo é produzir um documento com qualidade de publicação, o que, neste caso, significa um pedido de patente que atenda às condições para depósito no INPI. A escolha de textos da área de patentes como objeto do jogo 1 se deve a vários fatores. Primeiro, por minha experiência de trabalho nessa área, que inclui três anos como tradutor de documentos do domínio de patentes em um escritório de patentes no Rio de Janeiro, principalmente versões para o inglês de pareceres do INPI sobre a aceitabilidade ou não de um pedido de patente, e, depois, como autônomo, na tradução de pedidos de patentes e conteúdos assemelhados para o português. Por estar habituado ao conteúdo, pude avaliar com pertinência as edições consideradas necessárias na pós-edição da TA, notadamente a estrutura e o jargão utilizados na formulação de pedidos de patentes, independentemente da natureza do produto para o qual se busca proteção patentária.

A escolha também teve por finalidade minimizar o impacto de fatores capazes de influenciar minha produtividade, como, por exemplo, a necessidade de realizar muitas e extensivas pesquisas sobre termos específicos. Tais fatores estão presentes e são importantes na dinâmica do jogo da tradução, mas não estão dentro do escopo de interesse *deste* jogo. Assim sendo, foram selecionados pedidos de patente de produtos do segmento de alimentos/rações secas para animais, com cuja

terminologia eu tenho familiaridade. Todos os pedidos foram originalmente redigidos em inglês, traduzidos para o português brasileiro e depositados no Instituto Nacional da Propriedade Industrial — INPI. Logo, são traduções feitas por um tradutor profissional, com qualidade de publicação.

Três documentos foram utilizados na dinâmica de pós-edição do jogo 1: Piloto, Doc. 1 e Doc. 2. Todos foram divididos em trechos curtos, de cerca de quinhentas palavras cada. A escolha de trechos curtos se deveu a uma tentativa de manter um desempenho uniforme em todas as sessões, evitando oscilações de atenção e produtividade normais na tradução de trechos muito longos. O processo de pós-edição de cada trecho foi chamado de *sessão de tradução*, e o conjunto das sessões de tradução de um documento específico compõe uma *tarefa de tradução*. A tabela abaixo apresenta informações básicas sobre as tarefas de tradução e os documentos utilizados.

Documento	Software	Total de sessões	Total de palavras
Piloto	Lilt + GNMT	8	4.207
Doc. 1	Lilt	8	4.214
Doc. 2	GNMT	6	3.109

Tabela 7. Tarefas de pós-edição do jogo 1

5.1.1. Sistemas de TA

Seguindo a intenção de oferecer a visão de um tradutor profissional sobre o desempenho atual da tradução automática a partir do uso de sistemas de combinações de tecnologia neural e recursos de ferramentas de auxílio ao tradutor, escolhi para o jogo os sistemas Lilt e Google Tradutor, já na versão mais recente, doravante chamado pela sigla GNMT, do inglês *Google Neural Machine Translation*.



Figura 3. Interface de usuário do Lilt

Neste trabalho, usei a versão online do Lilt (vide acima). A interface de usuário é semelhante à de uma ferramenta de memória de tradução, que mostra as possíveis correspondências existentes na memória de tradução, mas tem como diferencial a exibição dinâmica das sugestões de tradução automática (palavras, trechos de frase ou frases inteiras), que o tradutor pode escolher à medida que trabalha no segmento a ser traduzido, e a tecnologia de TA adaptativa. Na linha 4 da interface de tradução, vemos o original na caixa de texto superior, o espaço para a inserção da tradução na caixa de texto do meio e as sugestões da TA dinâmica (que são exibidas palavra a palavra) na caixa de texto inferior. Sempre que concorda com a sugestão, o tradutor aperta a tecla Enter para que a palavra sugerida seja inserida na tradução.

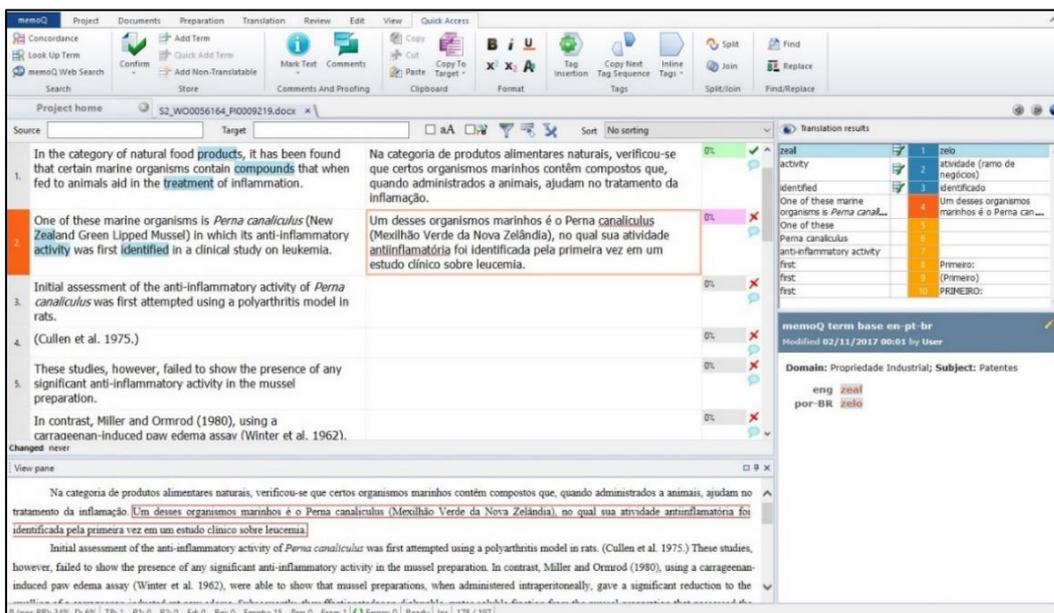


Figura 4. Interface de usuário do memoQ com resultados do GNMT

A versão do GNMT voltada para uso profissional, por sua vez, está disponível via Google Cloud Translation API, que conecta o sistema online da empresa a outro software. Para este jogo, conectamos o GNMT ao memoQ, um sistema de memória de tradução, através da interface de tradução automática deste último. O sistema foi configurado para oferecer resultados de concordância parcial acima de 75%. Quando não houvesse resultados na memória de tradução, as sugestões de TA do GNMT seriam exibidas automaticamente no segmento. O painel superior direito mostra as sugestões da TA ao lado da caixa de cor laranja (vide fig. 4, acima).

5.1.2. Piloto

Para familiarização com a interface de usuário e as características técnicas dos dois sistemas, realizei uma tarefa piloto de pós-edição, em que o arquivo foi editado de forma alternada nos dois sistemas: as sessões 1, 3, 6 e 7 foram feitas no GNMT e as sessões 2, 4, 5 e 8, no Lilt. Escolhi fazer duas sessões seguidas na mesma ferramenta (4 e 5 no Lilt, 6 e 7 do GNMT) para ver se causaria alguma mudança nas sugestões exibidas, tais como uma percepção mais clara do funcionamento da TA adaptativa nas sugestões do Lilt ou o aparecimento de sugestões da MT do memoQ, mas não houve qualquer alteração no comportamento dos sistemas. De todo modo, a tarefa de pós-edição correu sem problemas técnicos, e por isso me considerei preparado para iniciar as tarefas de avaliação propriamente ditas. Por conta de suas características, o piloto não foi considerado nas etapas de avaliação quantitativa e qualitativa do jogo.

5.1.3. Dinâmica

O objetivo do jogo 1 era produzir um texto-alvo com qualidade de publicação, por isso, a modalidade de pós-edição escolhida foi a PE completa, cuja expectativa de qualidade é semelhante à qualidade de um tradutor humano. Cada um dos dois documentos principais foi inteiramente pós-editado em uma única ferramenta, de forma a se ter um panorama do desempenho do sistema específico na tarefa. As etapas da dinâmica da pós-edição são mostradas na figura abaixo.

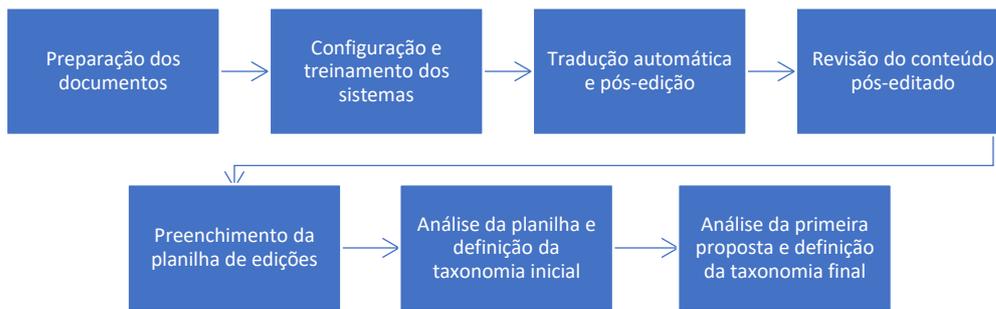


Figura 5. Etapas da dinâmica do jogo 1

5.1.3.1. Preparação dos documentos

Nenhuma edição foi feita na ordenação do conteúdo, os segmentos traduzidos aparecem na ordem em que estavam no texto fonte. Apenas tabelas e legendas foram retiradas, de forma a evitar interrupções no fluxo normal do texto.

5.1.3.2. Configuração e treinamento dos sistemas

Ambas as ferramentas foram alimentadas com uma memória de tradução e um glossário. A memória é um corpus paralelo bilíngue inglês-português brasileiro e originalmente continha mais de 310 mil segmentos de textos do domínio de patentes, notadamente análises do mérito de pedidos de patente feitas pelo INPI e publicadas na Revista de Propriedade Industrial (RPI) do instituto. Para aumentar a relevância dos resultados a serem exibidos nos dois sistemas, foram adicionados à memória dez pedidos de patentes relacionados ao campo de alimentos para animais. O glossário, carregado em opção somente leitura para uso como referência, contém terminologia relacionada a patentes e foi compilado por mim ao longo da minha experiência profissional no ramo.

5.1.3.3. Tradução automática e pós-edição

As sessões de tradução do GNMT ocorreram da seguinte forma: cada documento foi traduzido automaticamente no início, usando o recurso de pré-tradução do memoQ. A versão pré-traduzida foi salva para uso posterior na tarefa de comparação de traduções do jogo 3. Após a pré-tradução, fiz a pós-edição segmento a segmento.

No Lilt, a pré-tradução precisou ser feita segmento a segmento. Ao final desse processo, o documento foi baixado da plataforma online e salvo para uso posterior no jogo 3. Em seguida, fiz a pós-edição segmento a segmento. Com essa dinâmica, necessária para se obter o arquivo com a TA sem edição, acabei não utilizando o recurso de tradução dinâmica do Lilt, que oferece sugestões de tradução palavra a palavra. Embora essa dinâmica não tenha interferido no recurso de aprendizado de máquina do Lilt, pois ele ocorreu durante a pós-edição dos segmentos pré-traduzidos, decidi usar o sistema na tarefa de comparação entre as traduções do mesmo texto com uso de tradução automática e memória de tradução do jogo 2. Foi a maneira que encontrei para ter a chance de usar (e avaliar) todos os recursos do Lilt.

O tempo de cada sessão de tradução para avaliação de produtividade foi registrado logo após o término da pós-edição. Para tanto, usei as ferramentas de registro de tempo de edição de documento disponíveis nos dois sistemas. Os números de produtividade do jogo 1 se referem exclusivamente ao tempo da sessão, sem incluir a tarefa de revisão posterior. Essa decisão foi tomada por uma razão prática: a maioria dos documentos citados neste trabalho fez o registro do tempo gasto nos experimentos sem contar com a tarefa de revisão. Dessa forma, os meus resultados de produtividade foram medidos da mesma forma que a maioria das minhas referências.

5.1.3.4. Revisão do conteúdo pós-editado

Esta etapa de revisão consistiu na revisão do texto pós-editado para aprimoramento do estilo e da fluência da tradução, bem como para validação da terminologia. Não houve registro do tempo despendido na tarefa.

5.1.3.5. Preenchimento da planilha de edições

A planilha de edições foi preenchida após a conclusão da revisão do texto pós-editado e foi dividida em quatro colunas: texto original, saída da TA, texto pós-editado e comentários. Os comentários trazem anotações sobre a natureza das

edições realizadas com o objetivo de elaborar a primeira proposta de categorização das edições feitas.

5.1.3.6.

Análise da planilha e definição de taxonomia inicial

Os comentários sobre as edições serviram de ponto de partida para a definição da proposta inicial de taxonomia de edições. A ideia de criar uma categorização composta de níveis distintos surgiu logo no início do processo. Nesse primeiro momento, também era minha intenção que a planilha de comentários abrangesse tantas categorias quanto eu julgasse necessárias para o processo de análise das edições de pós-edição segundo minha visão de tradutor profissional. Uma segunda análise tornaria possível identificar quais categorias eram específicas demais para, em seguida, começar a juntá-las em categorias mais gerais, de modo a facilitar o processo de categorização. Esse processo culminou na primeira proposta de taxonomia, composta de seis categorias e 32 subcategorias.

5.1.3.7.

Análise da primeira proposta e definição da taxonomia final

Após a conclusão da primeira versão da taxonomia, realizei uma segunda revisão das edições realizadas e uma análise extensiva de toda a categorização proposta inicialmente. Essa segunda revisão e análise teve como objetivo contrastar minha proposta inicial com as propostas encontradas na literatura. Considerei importante validar minhas categorias e avaliar em que medida refletiam as propostas de outros autores e como poderiam trazer uma visão diferente e relevante para tradutores e pesquisadores.

Assim, ao fim do processo de revisão, a taxonomia foi reduzida a cinco categorias principais e vinte subcategorias. Três das categorias principais se basearam em subcategorias de *linguagem* do modelo da LISA: *tradução*, *terminologia* e *linguagem*. Outra categoria, *revisão*, foi inspirada nas categorias mestras de Almeida (2013). Por fim, tem-se a categoria *estilo*, definida por mim. A seguir, as definições de cada uma:

- Tradução: conjuga características das subcategorias *precisão*, *erro de tradução* e *país* do modelo da LISA e engloba traduções incorretas, informações ausentes ou adicionadas na tradução e padrões de país incorretos;
- Terminologia: mescla as subcategorias *terminologia* e *consistência* da LISA, englobando a inadequação e a falta de padronização terminológica;
- Linguagem: semelhante à subcategoria de mesmo nome da LISA, abrange problemas relacionados a gramática, semântica, ortografia e pontuação;
- Revisão: trata da natureza das edições que realizei na segunda revisão do conteúdo, quando percebi os equívocos que havia cometido na primeira revisão;
- Estilo: trata das edições que fiz no sentido de melhorar o estilo e a fluência do texto final, seguindo a lógica da pós-edição completa que visa obter uma tradução com qualidade humana.

A tabela a seguir mostra divisão completa da taxonomia de edições realizadas no jogo 1, com todas as categorias e seus respectivos subníveis. A descrição destes segue a orientação de basear a taxonomia na *natureza das edições*, não em *erros* da saída da TA.

Categorias	Subcategorias	Descrição
Tradução	Adição	Edição para incluir no texto-alvo uma informação que estava ausente da TA, mas presente no texto-fonte
	Erro de tradução	Edição de tradução inadequada do texto-fonte; em casos de trechos muito truncados, o trecho inteiro foi incluído como erro de tradução
	Erro do original	Edição para corrigir no texto-alvo um erro causado por incorreção no texto-fonte
	Exclusão	Edição para excluir do texto-alvo uma informação que estava presente na TA, mas ausente no texto-fonte
	Local	Edição para adequar a saída da TA aos padrões do país, tais como separadores decimais, uso de aspas e citações de títulos de obras
	Não traduzido	Edição para corrigir palavras/expressões que não foram traduzidas na saída da TA

	Ordem errada	Edição para corrigir uma ordem incorreta ou imprecisa das palavras no texto-fonte
Linguagem	Concordância	Edição para corrigir erros de concordância verbal e nominal na saída da TA
	Determinantes	Edição para corrigir o uso de determinantes (artigos, pronomes possessivos e demonstrativos) na saída da TA
	Ortografia	Edição para corrigir palavras com erro de grafia na saída da TA
	Paralelismo	Edição para corrigir a falta de correspondência e/ou simetria entre estruturas presentes na saída da TA
	Pontuação	Edição para corrigir pontuação incorreta ou imprecisa no texto-fonte
	Preposições	Edição para corrigir uso incorreto de preposições na saída da TA
	Registro	Edição para adaptar o uso de termos/expressões na saída da TA à finalidade do texto-alvo
Terminologia	Adequação	Edição para adequar termos e expressões usados da saída da TA à finalidade do texto-alvo
	Consistência	Edição para corrigir no texto-alvo a falta de consistência no uso de termos/expressões na saída da TA
Preferencial	<i>sem subnível</i>	Edição em termos, expressões ou estruturas não incorretas para aprimorar ou adequar a saída da TA à finalidade do texto-alvo
Revisão	Erro de correção	Revisão de pós-edição de TA em que uma tradução incorreta foi substituída por outra igualmente incorreta
	Correção desnecessária	Revisão de pós-edição de TA em que uma tradução correta foi substituída por outra igualmente correta
	Revisão incorreta	Revisão de pós-edição de TA em que uma tradução correta foi substituída por uma incorreta
	Falta de correção	Revisão de pós-edição de TA em que uma tradução incorreta não foi corrigida

Tabela 8. Taxonomia de alterações

A finalidade principal da criação dessa taxonomia foi fornecer subsídios para uma análise mais rica e embasada das saídas de TA do Lilt e do GNMT. Com ela, podemos comparar os sistemas em diversas categorias e traçar um perfil mais completo do desempenho de ambos na seção de análise qualitativa do presente trabalho.

5.1.4. Análise de produtividade

A avaliação quantitativa do jogo 1 teve como objetivo medir a produtividade do tradutor ao usar cada uma das ferramentas de TA. Os resultados foram expressos em palavras traduzidas por hora, uma medida de produtividade rotineira para tradutores. Essa unidade contempla tanto as medições por palavras, que reflete a forma como a tradução humana de pedidos de patentes é cobrada, quanto por tempo, que reflete a forma de cobrança por uma tarefa de pós-edição. Os quadros a seguir mostram a produtividade do tradutor nas sessões de tradução no Lilt e no GNMT e na tarefa de tradução como um todo.

Doc. 1 (sessão)	Software	Tempo total	Total palavras	Palavras/hora
1	Lilt	0h50	505	605
2	Lilt	0h42	558	793
3	Lilt	0h32	548	1.009
4	Lilt	0h38	582	913
5	Lilt	0h38	553	867
6	Lilt	0h48	497	614
7	Lilt	0h49	505	613
8	Lilt	0h38	466	722
Total geral		5h35	4.214	755

Quadro 3. Produtividade no Lilt

Doc. 2 (sessão)	Software	Tempo total	Total palavras	Palavras/hora
1	GNMT	0h53	512	627
2	GNMT	0h57	561	631
3	GNMT	0h32	490	912
4	GNMT	0h36	525	952
5	GNMT	1h11	494	440
6	GNMT	0h20	527	1.573
Total geral		4h29	3.109	693

Quadro 4. Produtividade no GNMT

É possível notar que a variação de produtividade foi menor nas sessões de tradução realizadas no Lilt: a sessão 3 (1.009 palavras/hora) foi 166% mais produtiva que a sessão 7 (613 palavras/hora). A média de produtividade geral foi de 755 palavras/hora. No GNMT, a sessão 6 (1.573 palavras/hora) foi 357% mais produtiva que a sessão 5 (440 palavras/hora). A média de produtividade geral foi

de 693 palavras/hora. A figura abaixo mostra a variação de produtividade sessão a sessão (oito no Lilt, seis no GNMT).

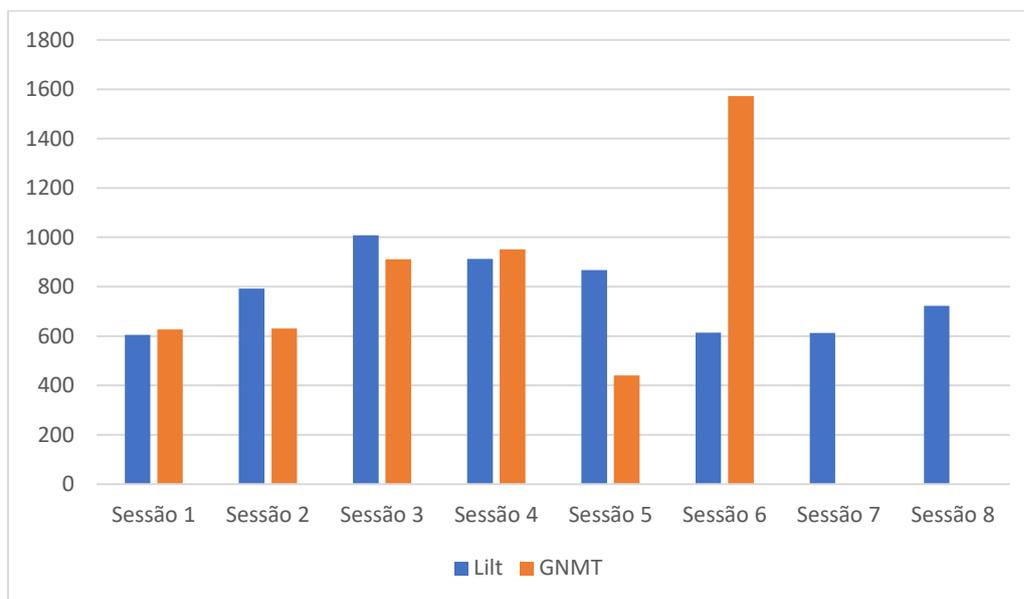


Figura 6. Produtividade por sessão (Lilt x GNMT)

A variação exibida no GNMT se deve, possivelmente, às características específicas do conteúdo das sessões de tradução 5 e 6 do Doc. 2. Embora já tivesse familiaridade com a terminologia de patentes do campo de rações para animais, precisei, ao editar a sessão 5, fazer uma pesquisa extensiva para validar a tradução de uma expressão recorrente no setor de criação de porcos, a saber: “nursery pigs”, cuja tradução mais comum em português brasileiro é “leitões de creche”. Durante a pesquisa, várias outras traduções possíveis foram encontradas, tais como: “porcos de berçário”, “porcos em fase de desmame”, “porcos de viveiro” etc. A expressão “leitões de creche” foi, por fim, validada em sites especializados em criação de porcos¹⁵, mas o tempo despendido na pesquisa impactou muito a produtividade dessa sessão.

A sessão 6 do GNMT, por outro lado, continha segmentos muito semelhantes entre si, que precisaram apenas de pequenas edições. Esta é uma característica de muitos pedidos de patentes, que apresentam pequenas variações no objeto do pedido para aumentar as chances de concessão de patente. Por conta disso, a pós-edição dessa sessão foi consideravelmente mais rápida que as outras. Se as sessões mais e menos produtivas fossem desconsideradas, os resultados de produtividade

¹⁵ Exemplos disponíveis em: <<https://www.deheus.com.br/artigos/creche-uma-fase-que-requer-atencao-536>>; <<https://pt.engormix.com/suinoicultura/artigos/nutricao-leitoes-creche-t37446.htm>>; <<http://www.sossuinos.com.br/tecnicos/info142.pdf>>. Acesso em: 20-21 set. 2017.

média seriam consideravelmente distintos: 754 palavras/hora no Lilt e 780 palavras/hora no GNMT.

Essa disparidade de valores ilustra bem a dificuldade de se emular a realidade em experimentos. Basta uma simples mudança de paradigma na observação para se obter resultados diferentes. E como fazer para se isolar as tantas variáveis envolvidas em estudos com humanos e todas as suas subjetividades? Experimentos como Guerberof (2014) e Bentivogli (2016a), por exemplo, usaram cálculos estatísticos para gerar resultados estimativos ou compensar fatores de ordem. São formas de se tentar restringir o incontrolável da realidade.

Também é possível comparar os resultados dos meus jogos a outros experimentos semelhantes, como, por exemplo, ao medir a produtividade contando apenas o tempo de pós-edição, mas não o de revisão. Grosso modo, a tarefa de revisão não deveria mostrar grandes variações de tempo entre sistemas, pois independe delas: as ferramentas de tradução assistida por computador oferecem recursos para facilitar o processo de tradução, mas são de pouca oferecer ajuda para, por exemplo, validar a tradução de algum termo.

Comparar é preciso. Se, ao categorizar, “tornamos igual o que não é igual”, como mostrou Ellis (1993), ao comparar resultados segundo critérios definidos de categorização, damos utilidade à comparação. A próxima seção compara os resultados de produtividade obtidos neste estudo e na literatura, atentando para as distintas abordagens de cada um como forma de relativizar o absolutismo dos números.

5.1.5. Comparações de produtividade

A pesquisa de 2012 da Common Sense Advisory, citada no capítulo 4, divulgou que a produção média de um tradutor profissional é de 2.684 palavras por dia (ou 335 palavras por hora, para oito horas de trabalho). Se compararmos essa cifra à produtividade média das tarefas de pós-edição deste estudo, veremos que a produtividade obtida no Lilt é 125% maior (755 palavras por hora), e no GNMT, 107% maior (693 palavras por hora). Para se ter quadro mais abrangente, no entanto, é preciso comparar os resultados do presente estudo a outros obtidos na literatura (vide fig. 9, abaixo).

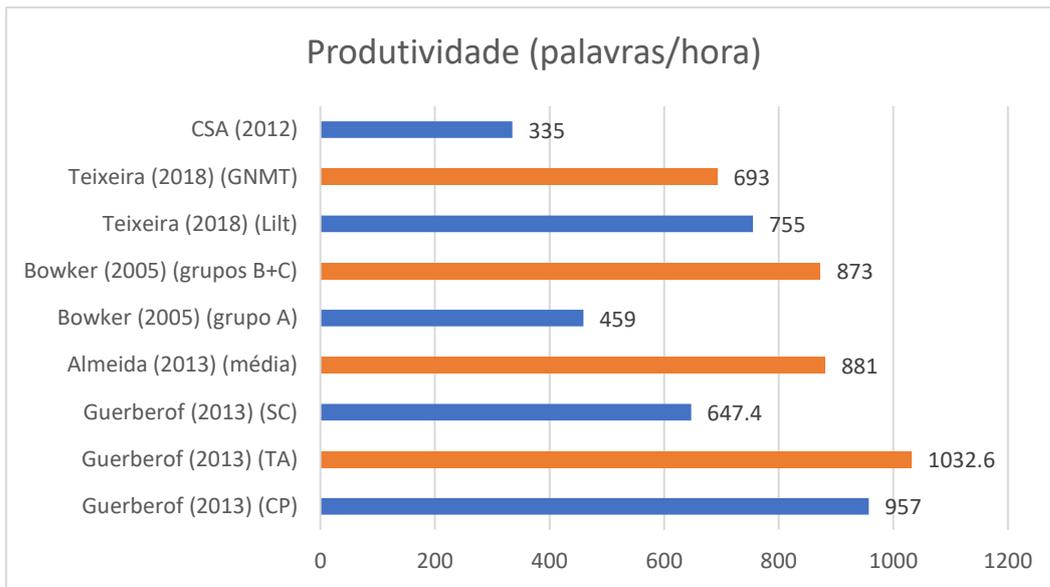


Figura 7. Comparação de produtividade entre estudos

CSA (2012) se refere aos resultados do Common Sense Advisory (2012), citados no parágrafo acima. Teixeira (2018) são resultados relatados na seção anterior deste trabalho. Bowker (2005) se refere aos grupos de estudantes de tradução citados no estudo, como segue: o grupo A engloba os três estudantes que fizeram a tarefa de tradução sem o uso de memória de tradução, o grupo B, os que traduziram com o auxílio de sugestões da memória, o grupo C, os que receberam sugestões de memória de tradução deliberadamente erradas. Almeida (2013) se refere à média de produtividade dos dez tradutores que participaram do estudo de pós-edição de textos de TI traduzidos automaticamente do inglês para o português brasileiro. Por fim, Guerberof (2014) engloba as três categorias analisadas no estudo: CP se refere a segmentos com sugestões de correspondência parcial de memória de tradução (85%-84%), TA, a segmentos com sugestões de tradução automática, e SC (sem correspondência) a segmentos traduzidos do zero, sem sugestões de ferramentas de auxílio à tradução.

Como se vê, os resultados obtidos no presente estudo estão acima dos números de produtividade relacionados à tradução sem o auxílio de ferramentas, porém abaixo dos resultados obtidos tanto com memória de tradução quanto com tradução automática em outros estudos. É preciso, porém, analisar as características de cada um deles para entender melhor os resultados.

Almeida (2013), por exemplo, obteve uma média de 881 palavras/hora, mas em uma tarefa de pós-edição intermediária, ou seja, que não exigia qualidade de

tradução humana, e que apresentava os segmentos em ordem aleatória, dispensando os tradutores de se preocuparem, por exemplo, com coesão e consistência textual, tarefas que certamente demandaram tempo de tradução no presente trabalho, por conta dos textos lineares.

Guerberof (2014) também apresentou os segmentos para tradução de forma aleatória; além disso, nos segmentos provenientes de memória de tradução, usou apenas correspondências entre 85% e 94%, que exigem pouca edição e, conseqüentemente, possibilitaram que os participantes atingissem a alta produtividade de 957 palavras/hora.

Bowker (2005), por sua vez, relata que os participantes dos grupos B e C, estudantes de tradução, cometeram muitos erros em suas traduções, notadamente os que receberam sugestões deliberadamente erradas. Além disso, um dos seis participantes relatou que gostaria de ter mais tempo para revisar o texto traduzido. Apesar da alta produtividade de 873 palavras/hora, parece, por um lado, que falta experiência aos estudantes para revisar e avaliar a qualidade do próprio trabalho, por outro, que sofreram menos pressão para produzir um trabalho de qualidade por não se tratar de tarefa profissional, mas acadêmica, e por saber que suas competências tradutórias não estavam em avaliação. Discutiremos adiante, nos jogos 2 e 3, como o fato de realizar uma tradução em ambiente profissional ou acadêmico pode influenciar a qualidade final.

Em resumo, por conta de todas as variáveis citadas anteriormente, e também por minha postura de tentar produzir o melhor texto possível a partir da pós-edição de TA, sem uma preocupação maior com prazos, acredito que os resultados de produtividade obtidos no jogo 1 estão à altura dos números divulgados na literatura.

5.1.6. Avaliação qualitativa

A avaliação qualitativa apresentada a seguir teve como objetivo a criação de uma taxonomia de edições feitas durante a pós-edição da saída da TA. A partir dela, foi possível visualizar as características da tradução automática gerada pelo Lilt e pelo GNMT para o par inglês-português no domínio de patentes, verificar quais aspectos da saída de TA de ambos os sistemas demandam mais atenção de um pós-editor profissional e avaliar a relação entre produtividade e qualidade do texto-

fonte. A figura a seguir apresenta os resultados de cada sistema por categoria para ajudar na análise de ambos e dar início à discussão sobre a relação custo-benefício do uso de TA como ferramenta de tradução por um tradutor autônomo.

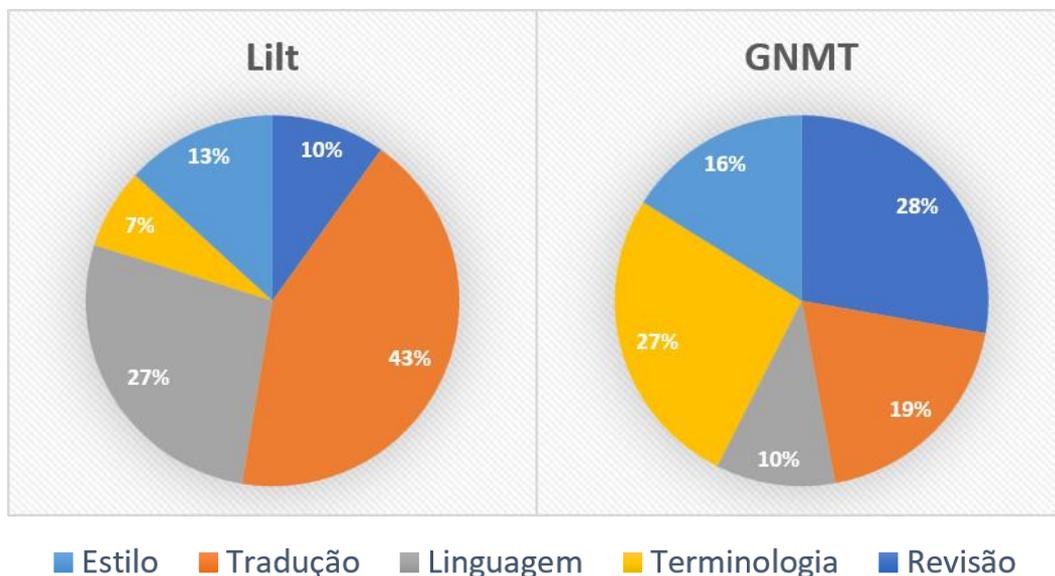


Figura 8. Resultados por categoria (Lilt x GNMT)

5.1.7. Resultados por sistema

Dos 209 segmentos do Doc. 1 editados no Lilt, 199 foram alterados (95% dos segmentos). No total, 760 edições foram feitas durante as oito sessões de tradução, o que significa uma média de 95 correções por sessão de tradução.

O número de edições na categoria *tradução* realizadas no Lilt é consideravelmente maior do que no GNMT. O Lilt teve diversos problemas no subnível *erro de tradução* (60% das edições feitas na categoria *tradução*; 25,5% de todas as edições feitas). A categoria *erro de tradução* abrange desde erros na tradução de uma pequena sequência de palavras até longas passagens sem sentido, em foi preciso verificar o original para compreender o que estava sendo dito, o que dá ainda mais peso às porcentagens acima. Talvez com a continuação das traduções e o consequente aprendizado de máquina a cada edição realizada pelo tradutor no sistema, o Lilt possa mostrar um salto qualitativo de desempenho no futuro.

Na categoria *linguagem*, a segunda mais editada no Lilt, um problema recorrente foi o uso de partículas: 25,6% do total de 53 edições recaíram sobre as categorias *preposições* e *determinantes*. *Concordância* foi a categoria mais

problemática em *linguagem*, o que parece demonstrar que o Lilt enfrenta um problema comum na tradução do inglês para as línguas românicas: adequar sentenças de gênero neutro a uma linguagem que exige determinação de gênero.

Por fim, é importante mencionar que categoria Revisão representou 13,1% das edições no Lilt, o que está abaixo do valor médio encontrado em Almeida (2013) para a combinação das categorias *mudanças essenciais não implementadas* e *erros introduzidos* para os tradutores do português brasileiro: 15,13% e 5,28%, respectivamente, ou 20,41% de todas as edições feitas naquele estudo. Esses números apontam para um maior nível de qualidade do texto-fonte do presente trabalho após a primeira pós-edição, que sofreu muito menos edições dessa natureza do que os textos de Almeida (2013). A comparação entre as traduções que produzi para este jogo e os documentos originais depositados no INPI, que vai ser discutida adiante no relato sobre o jogo 3, podem lançar nova luz sobre a discussão do impacto da pós-edição de TA na qualidade da tradução.

Dos 158 segmentos do Doc. 2 editados no GNMT, 135 foram alterados (85%). Foram feitas 435 edições durante as seis sessões de tradução, o que significa uma média de 72,5 alterações por sessão. As categorias de edição foram distribuídas de maneira mais uniforme no Lilt do que no GNMT. A maior parte das edições se encaixou na categoria *terminologia*, com 27% das ocorrências. A alta incidência de edições terminológicas pode ser vista de duas formas. Uma, positiva, é a indicação que a saída da TA do GNMT é consideravelmente mais fluente do que a do Lilt, um aspecto que parece ser corroborado pela maior proporção das categorias *tradução* e *linguagem* na divisão taxonômica (respectivamente 19% e 10% no Google, e 43% e 27% no Lilt). Outra, negativa, é a impossibilidade de se treinar o GNMT diretamente para a tradução de patentes, já que o sistema não oferece essa opção nem na versão para uso profissional.

Nesse caso, a solução seria a inclusão de novos termos no banco terminológico do memoQ na etapa de configuração do sistema. Essa solução, no entanto, é paliativa, já que o memoQ sugere os termos em seu painel, mas não inclui a sugestão automaticamente na saída da TA. Como se pode ver na figura abaixo, termos incluídos no banco terminológico do memoQ (destacados nos quadrados vermelhos) não aparecem como sugestões na tradução automática. Exemplos: o termo “atividade”, que deveria aparecer como sugestão de tradução para “activity”, aparece como “actividade” (grafia do português de Portugal) na saída da TA.

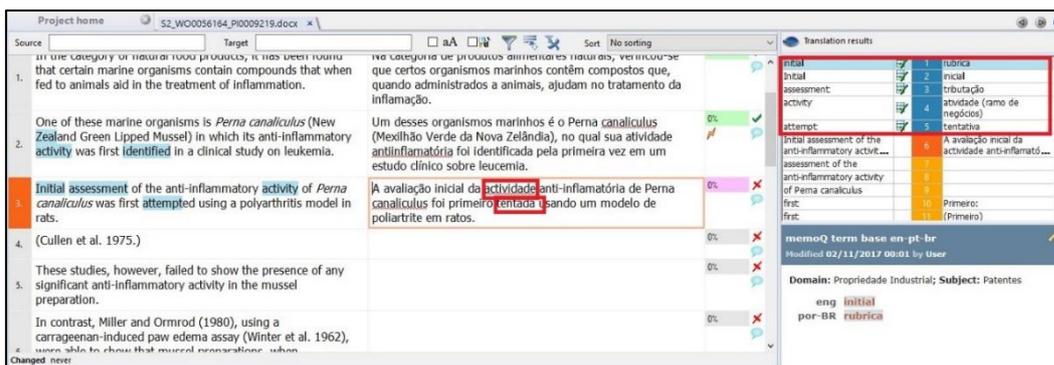


Figura 9. Sugestões do banco terminológico no memoQ

Outra observação relevante diz respeito à categoria *revisão*. Embora a proporção dessa categoria na divisão taxonômica do GNMT seja semelhante à obtida pelo Lilt (16% e 13%, respectivamente), a divisão das subcategorias parece trazer um aspecto distinto. Enquanto no Lilt a subcategoria *falta de correção*, que representa os erros não percebidos pelo tradutor na primeira revisão da pós-edição, responde por 29% das edições na categoria *revisão*, no GNMT, essa proporção sobe para 45,7%. Os dois números parecem confirmar uma tendência mencionada em Way (2018): quando a qualidade de uma tradução é alta, como na saída de tradução automática neural, os revisores (ou o tradutor, em nosso jogo) podem acabar ignorando possíveis erros por causa da fluência do texto traduzido. Outro dado relevante é a proporção do subnível *correção desnecessária* na categoria *revisão* no GNMT (8,5%), o que parece corroborar uma maior confiança na saída da TA, gerando menos edições no texto-alvo.

Na categoria *linguagem*, os subníveis *preposições* e *determinantes*, tomados em conjunto, representam apenas 8,8% de todas as edições no GNMT (em comparação com 25,6% no Lilt). Se compararmos a porcentagem da categoria *terminologia* no GNMT e no Lilt (27% contra apenas 7%), temos uma diferença que parece refletir o menor número de mudanças na categoria *tradução* (19% de todas as mudanças, contra 43% no Lilt), corroborando, em tese, a melhor qualidade geral do GNMT sobre o Lilt, já que, neste último, as edições terminológicas parecem ter ficado mascaradas dentro das edições da categoria *tradução*.

O jogo 1 mostrou que a pós-edição de TA melhora claramente a produtividade em relação às traduções não assistidas por computador. De acordo com nossa taxonomia para o português brasileiro e a análise feita na segunda etapa da avaliação, o GNMT mostrou melhor qualidade do que o Lilt, embora ambos devam ser considerados como ferramentas úteis para a tradução assistida por computador.

5.2. Jogo 2

Após a conclusão do jogo 1, surgiu uma outra questão pertinente em relação à avaliação de TA: o que aconteceria se comparássemos duas traduções de um mesmo texto, feitas pelo mesmo tradutor em momentos diferentes e distantes no tempo, uma com o uso de memória de tradução, outra com o uso de TA? O resultado poderia dizer algo sobre a influência da TA no estilo do tradutor? Haveria algum impacto negativo na qualidade, ao contrário do que ocorrera no jogo 1?

O jogo 2 é um jogo de avaliação que se propõe a comparar duas traduções do mesmo texto feitas por mim. A primeira foi produzida em uma ferramenta de memória de tradução, por volta de 2011; a segunda, na plataforma online do Lilt, especificamente para este estudo. A avaliação de ambas aborda os *lances* feitos por mim durante os dois jogos de tradução.

Embora tenha tido um resultado inferior ao Google Tradutor no jogo 1, o Lilt foi escolhido, em primeiro lugar, por conta da promessa de aprimoramento da qualidade dos resultados com o uso recorrente da ferramenta, graças à tecnologia de tradução adaptativa. Em segundo lugar, para ter a chance de usar os todos os recursos disponíveis. No jogo 1, aceitei a tradução automática do segmento inteiro e depois fiz a pós-edição do conteúdo traduzido; no jogo 2, utilizei o recurso de tradução dinâmica da ferramenta, que oferecia sugestões de tradução para cada termo que ia surgindo (vide fig. 10, abaixo).

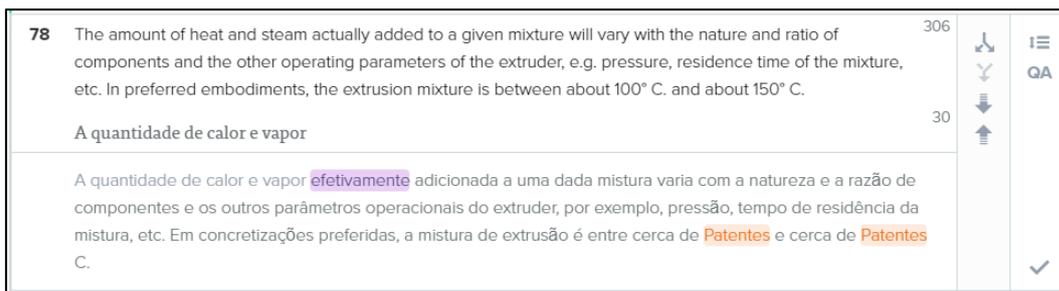


Figura 10. Interface do Lilt com recurso de tradução dinâmica.

5.2.1. Análise quantitativa

A análise da tradução atual incluiu uma parte quantitativa, cujos resultados foram expressos em termos de palavras traduzidas por hora. A finalidade dessa avaliação é comparar a produtividade da tradução feita com o Lilt aos resultados divulgados na literatura citada neste trabalho. Os resultados quantitativos foram colhidos em dois momentos distintos: primeiro, ao terminar a pós-edição, ou seja, antes da etapa de revisão final; o segundo, após a revisão, quando dei a tradução como pronta.

	Total de palavras	Tempo	Produtiv. (pal./h)
Sem revisão	2.466	2h39min	907,1
Com revisão	2.466	3h12min	754,5

Tabela 9. Produtividade no jogo 2

A produtividade sem revisão no jogo 2 no Lilt ocuparia a terceira posição na comparação de produtividade entre estudos exibida anteriormente neste capítulo (vide fig. 7), ficando atrás apenas dos resultados de traduções efetuadas com TA (1.032,6 pal./h) e correspondências parciais (957 pal./h) em Guerberof (2014). Já a produtividade com revisão, que ocuparia uma posição intermediária na comparação, foi praticamente idêntica à obtida por mim com o Lilt no jogo 1 (754,5 contra 755 pal./h, respectivamente). Esse resultado parece corroborar a minha produtividade média enquanto pós-editor, mostrando que o desempenho da dinâmica de pós-edição se repetiu entre jogos realizados no mesmo sistema de TA.

5.2.2. Análise qualitativa

Passando à fase qualitativa, realizei a análise dos documentos. O modelo de avaliação adotado segue os padrões da indústria, com a análise de um extrato de cerca de mil palavras de cada documento para identificação e categorização das alterações. Não houve penalização pelas alterações, já que não se trata de uma análise para aprovação ou reprovação de traduções, mas da comparação de dois

documentos. Caso uma mesma alteração ocorra mais de uma vez, apenas a primeira será registrada.

A taxonomia das alterações realizadas em ambos os documentos pode fornecer mais dados para a avaliação da qualidade da tradução. Os gráficos abaixo mostram a categorização das alterações feitas e logo se nota que a categoria de maior percentagem em T1 e em T2 é *estilo* (neste, junto com *tradução* e *linguagem*), o que, inicialmente, aponta para uma boa qualidade em ambos os documentos, já que a maioria das alterações teve caráter preferencial.

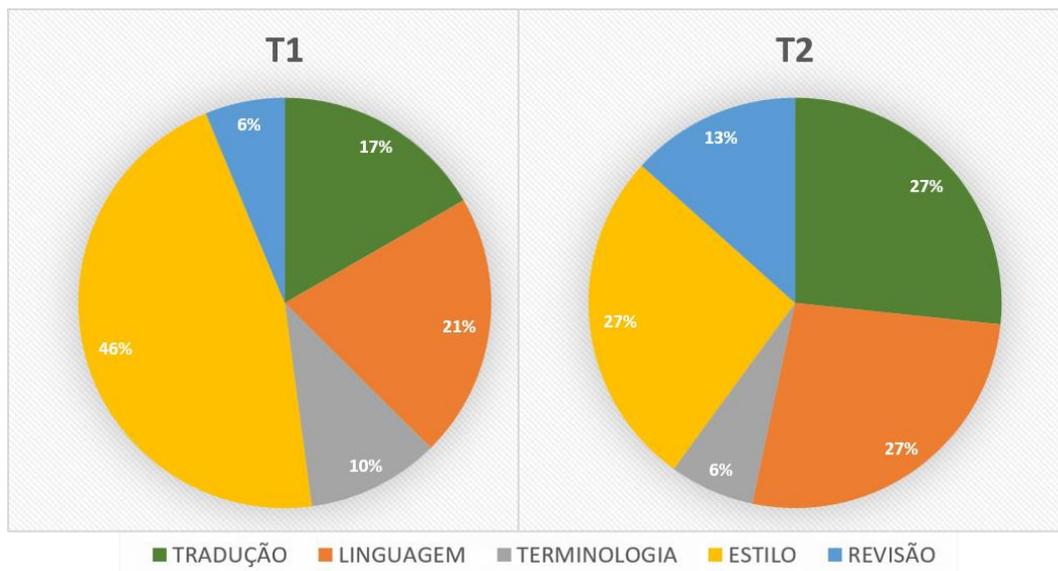


Figura 11. Categorização das edições feitas em T1 e T2

Uma informação importante, no entanto, é fornecida pelos números absolutos: T1 sofreu 48 alterações no processo de revisão para este jogo, enquanto T2 sofreu apenas 15. Considerando-se esse aspecto, T1 teve 5,5 vezes mais alterações que T2, mostrando que o texto do primeiro documento precisava de muito mais ajustes que o do segundo. Além disso, a alta proporção de alterações em *estilo* contribuiu para diminuir o percentual de outras categorias.

Tomemos como exemplo as categorias *tradução* e *linguagem* de T2: cada uma responde por 27% das alterações. Esse percentual poderia indicar problemas de qualidade do documento, mas significa apenas quatro alterações em cada categoria. É importante frisar também que a versão do documento T2 usada neste jogo foi produzida logo após a etapa de pós-edição, ainda sem a segunda revisão do conteúdo, realizada durante a fase de análise das categorias para a primeira proposta de taxonomia de edições. Em uma tradução profissional, essa segunda revisão

corresponderia a uma leitura apenas do texto-alvo, o que certamente ajudaria na identificação de possíveis melhorias em tradução, linguagem, terminologia e estilo.

No caso de T1, as proporções de *tradução* e *linguagem* foram de 17% e 21%, significando um total de oito e dez alterações, respectivamente. São, obviamente, números mais problemáticos que os de T2, e poderiam significar uma reprovação para o tradutor no caso de aplicação de um modelo de avaliação de qualidade razoavelmente rigoroso, como, por exemplo, o descrito no capítulo 4 desta dissertação. É preciso, no entanto, relativizar a análise. O modelo descrito no capítulo 4 era voltado para trabalhos de localização de textos de tecnologia da informação — conteúdos para websites e marketing online, principalmente —, que geralmente trazem muitos segmentos repetidos ou semelhantes. Em projetos nessa área, é comum encontrar memórias de tradução robustas, com alto nível de especificidade para cada cliente de tradução, o que aumenta não só a chance de o tradutor receber sugestões de segmentos e de termos, mas também de encontrar frases semelhantes ao usar o recurso de busca por contexto na memória de tradução.

No caso de patentes, embora haja uma recorrência de termos do jargão da área, ou seja, do texto que estrutura o pedido de patente, o tipo de conteúdo é menos específico, pois se refere a distintos ramos de conhecimento, tais como indústria alimentícia ou farmacêutica, bioquímica, medicina, eletrônica, tecnologia da informação, engenharia mecânica, engenharia civil, indústria automobilística etc. Há menos recorrência de correspondências de memória de tradução em comparação ao que ocorre, por exemplo, em projetos de tecnologia da informação, o que leva a uma menor possibilidade de busca por estruturas semelhantes dentro da memória.

Voltando à taxonomia, vemos que, na categoria *terminologia*, a porcentagem sobre o total é de 10% em T1 e 6% em T2, correspondendo a cinco alterações no primeiro e uma no segundo. O número absoluto de T2 é perfeitamente aceitável para uma tradução de nível profissional. T1, por sua vez, demonstrou a necessidade de uma revisão terminológica mais precisa após a conclusão da tradução. Acredito que isso se deva, por um lado, ao fato de que eu ainda não tinha tanta experiência na tradução de patentes quando efetuei a primeira tradução, por isso o maior número de erros do que em T2; por outro, acredito que a pressão do prazo de entrega também possa ter contribuído para uma tradução mais apressada em T1 do que em T2, cujo foco maior foi a qualidade e não a produtividade. De qualquer forma, como veremos em detalhes nas duas avaliações realizadas no jogo 3, o padrão de T1 e T2

está adequado às expectativas de qualidade de tradução de pedidos de patente.

5.2.3. Análise comparativa

Para complementar a avaliação qualitativa, foi gerado de um documento de comparação entre os documentos (vide fig. 14, abaixo). O texto em azul corresponde ao texto traduzido anos atrás na ferramenta de memória de tradução (T1). O texto em vermelho corresponde à versão atual, produzida no Lilt (T2). O texto em preto corresponde aos trechos coincidentes entre ambos os documentos.

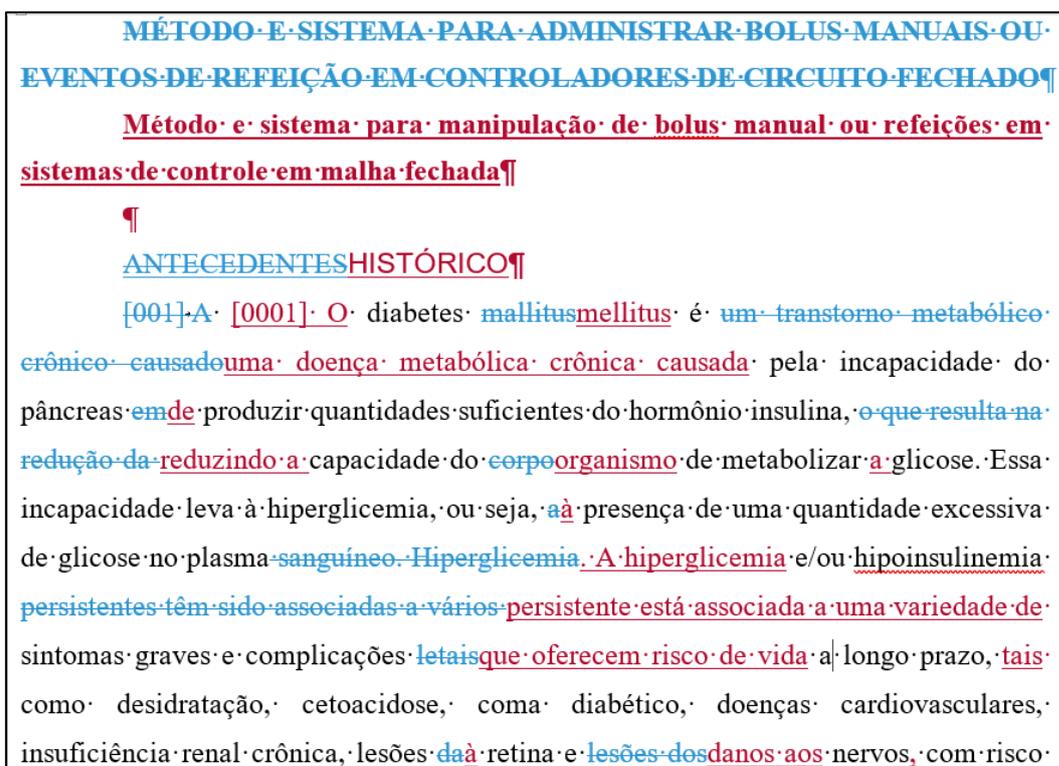


Figura 12. Comparação entre T1 e T2

De início, pode-se ver as diferenças terminológicas entre os dois documentos, a começar pelo título: “administrar bolus” e “manipulação de bolus”, “eventos de refeição” e “refeições”, “controladores de circuito fechado” e “sistemas de controle em malha fechada”. É provável que algumas diferenças se devam à finalidade e à dinâmica de cada um dos jogos.

T1 é reflexo de um texto produzido por um tradutor profissional com limite de prazo, ou seja, cuja finalidade era ser entregue com a melhor qualidade possível no prazo acordado. É possível que, devido a restrições de tempo, eu tenha sido

menos exigente na validação de termos durante a tradução do T1 do que ao fazer a pós-edição do T2, que teve grande foco na qualidade e incluiu uma etapa de revisão com pesquisa atenta dos termos encontrados no pedido de patente. É o caso da expressão “sistemas de controle em malha fechada”, validada após consulta a páginas especializadas em tratamentos para a diabetes, por exemplo. Por outro lado, a opção “administrar bolus manuais” de T1 é mais adequada ao contexto do que “manipulação de bolus manual”.

Talvez a opção pelo verbo manipular em T2 se deva às sugestões da tradução automática do Lilt, pois a expressão no texto-fonte é “handle manual boluses”. Embora não seja possível confirmar este fato por conta do recurso de sugestões de tradução automática a nível de palavra no Lilt, uma pesquisa posterior parece corroborar essa hipótese. Ao buscar o verbete “handle” no Dicionário inglês-português Webster¹⁶, vi que o primeiro conjunto de definições é “manusear, manejar e manipular”, enquanto “administrar” aparece no antepenúltimo conjunto de termos, e ainda assim no mesmo campo semântico de “guiar” e “governar”. Logo, é plausível que “manipular” tenha aparecido como a primeira sugestão da tradução automática do Lilt durante a pós-edição do T2.

Há, como observado acima, uma série de escolhas terminológicas distintas ao longo dos dois documentos. Para avaliar o quanto as escolhas do T2 podem ser fruto das sugestões da TA ou de uma maior preocupação terminológica, elaborei uma tabela comparativa com comentários sobre cada escolha, segundo o melhor do meu conhecimento, após pesquisa terminológica sobre cada um dos termos. Alguns resultados dessa comparação são exibidos no quadro a seguir, com observações sobre as estratégias de pesquisa de validação de termos.

¹⁶ Webster. Dicionário inglês-português. Antonio Houaiss (ed.), 18ª edição, Rio de Janeiro: Record, 2008.

Original	T1	T2	Comentário
blood glucose	glicose sanguínea	glicemia	Em busca na internet, “glicemia” é consideravelmente mais comum em contextos técnicos; termo provavelmente sugerido pela TA do Lilt.
external biologic	biofármacos exógenos	agentes biológicos externos	Não encontrei a expressão “biofármacos exógenos” em busca na internet, mas encontrei “insulina exógena”; o termo “agentes biológicos externos” não se refere a produtos como a insulina, mas a fungos, bactérias e insetos, no contexto de tecnologia agrícola.
[drug] delivery	aplicação	liberação	No contexto da bioquímica e de aparelhos do tipo mencionado no pedido, o termo “liberação” é o mais adequado.
infusion	infusão	perfusão	Embora “perfusão” seja mais frequente, as pesquisas indicam que os dois são de uso corrente. A opção “perfusão” foi sugerida pela TA do Lilt, eu provavelmente não a teria considerado sem a sugestão.
closed-loop controller	controladores em circuito fechado	sistema de controle em malha fechada	Não foram encontrados resultados na busca pela expressão “controlador em circuito fechado” + insulina no Google Acadêmico ¹⁷ ; os resultados para “controle em malha fechada” + insulina apontam para sistemas de controle de glicemia.
model-based controller	controle de modelo preditivo	controlador preditivo baseado em modelo	Após busca no Google Acadêmico, só foram encontradas ocorrências de “controlador preditivo baseado em modelo”. O termo foi encontrado após uma longa pesquisa para validação.

Quadro 5. Comparação terminológica (T1 e T2)

¹⁷ O Google Acadêmico é uma ferramenta de pesquisa de literatura acadêmica. Disponível em: <<https://scholar.google.com.br/>>. Acesso em: 26 out. 2018.

Outras diferenças entre as traduções estão relacionadas à ordenação da frase, como nos dois extratos exibidos na tabela abaixo. Os trechos ilustram as partes em que houve coincidência de estruturação da frase entre o texto-fonte e a tradução.

Texto-fonte	T1	T2
Because restoration of endogenous insulin protection is not yet possible, a permanent therapy is necessary which provides constant glyceemic control	Como ainda não é possível restaurar a produção endógena de insulina, é necessária uma terapia permanente que ofereça controle glicêmico constante	Como a restauração da produção endógena de insulina ainda não é possível, é necessária uma terapia permanente que forneça controle glicêmico constante
In such a scenario, the model in the MPC may predict a quiescent, steady glucose trend for the near future	Neste cenário, o modelo no MPC pode prever uma tendência latente e constante para a glicose para o curto prazo	Neste cenário, o modelo do CPM pode prever uma tendência de glicose inerte e invariável para o curto prazo

Quadro 6. Comparação entre trechos (T1 e T2)

Os extratos ilustram que não houve uma prevalência em T1 ou T2 da manutenção da estrutura das frases do texto-fonte, o que parece indicar que não houve um maior impacto das sugestões da TA na estruturação do texto traduzido. A comparação de outros trechos de T1 e T2 com o texto-fonte não apontou para qualquer impacto negativo da manutenção da estrutura do texto original na qualidade do texto-alvo.

De resto, a análise de outros extratos apontou para a prevalência do uso do singular em T1 e do plural em T2 e reconheceu diferenças em termos não especializados ao longo das duas traduções, tais como “atualmente” e “no momento”, “no entanto” e “contudo”, “de forma que” e “para que”, entre outras.

Assim, tomando o conjunto dos aspectos avaliados, o jogo 2 mostra como os lances do jogo da tradução mudam conforme as circunstâncias e como uma dinâmica diferente pode alterar a forma com o jogador (o tradutor) tenta chegar ao objetivo. Uma dinâmica em que a exigência de produtividade tem maior peso pode levar a uma relativização da qualidade, o oposto do que ocorreu na dinâmica da tarefa de pós-edição do jogo 2, que tinha particular preocupação com a qualidade. Assim, segundo esta avaliação, T2 mostrou mais qualidade que T1, exibindo ainda um nível de produtividade alinhado aos melhores desempenhos da literatura analisada para o presente trabalho.

5.3. Jogo 3

Vimos no item anterior como o mesmo tradutor se comportou de forma diferente em jogos de tradução com o mesmo *objeto*, mas *objetivos* distintos. E como se comportariam dois tradutores *diferentes* ao jogar o jogo da tradução com o mesmo objeto, ou seja, o mesmo texto-fonte? Como cada um deu seus lances em vista do objetivo do jogo, que era produzir um pedido de patente com qualidade suficiente para depósito no INPI? O que a tradução do documento *de facto*, que passaremos a chamar aqui de pedido depositado (PD) pode dizer sobre a qualidade das duas pós-edições de TA realizadas no jogo 1?¹⁸ Essas são perguntas que o jogo 3 se propõe a discutir ao comparar os pedidos depositados à pós-edição da TA gerada tanto pelo Lilt quanto pelo Google Tradutor.

5.3.1. Lilt x PD 1: análise qualitativa

Para analisar o jogo 3, usei os mesmos critérios de análise do jogo 2. Considerei um extrato de cerca de mil palavras para identificação e categorização das edições, sem penalização. Assim, com base nos resultados da divisão por categorias exibidos na figura abaixo, faço em seguida a comparação entre as alterações efetuadas no extrato da tradução final produzida na tarefa de pós-edição do Lilt, doravante chamada de PE-Lilt, ao mesmo trecho do pedido de patente depositado no INPI, doravante chamado de PD 1.

¹⁸ Embora não seja possível confirmar a *dinâmica* de tradução usada pelo tradutor do pedido de depósito — uso de MT, de TA, tradução sem auxílio de ferramentas —, resta confirmado que o *objetivo* da tradução foi cumprido, pois o documento traduzido foi aceito pelo INPI. Logo, é um documento válido, produzido por um profissional, e por isso foi usado como modelo de comparação.

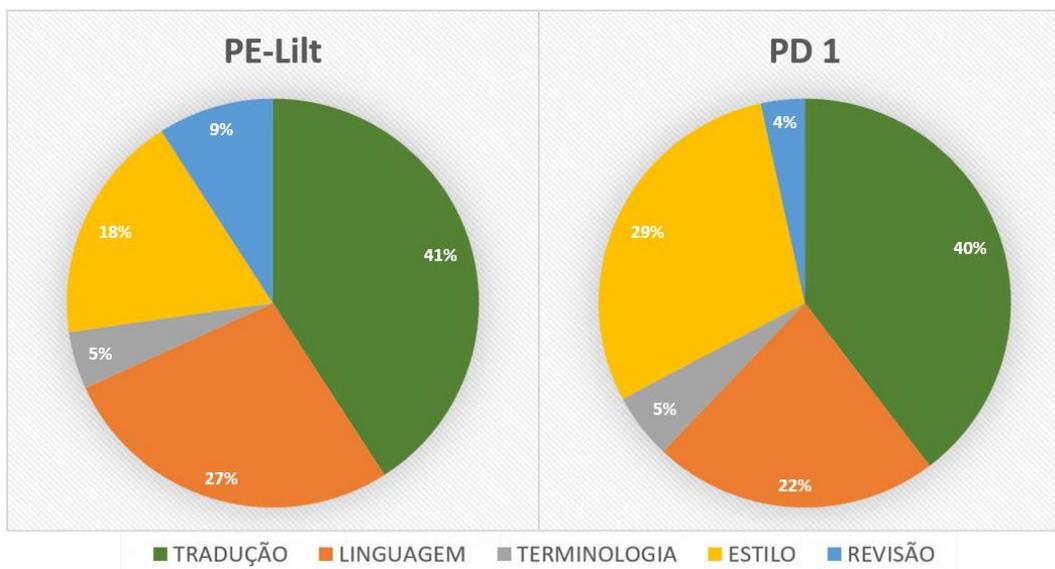


Figura 13. Categorização das edições feitas (PE-Lilt e PD 1)

Ao comparar visualmente o resultado da classificação das alterações da PE-Lilt e do PD 1, vemos uma considerável semelhança entre ambos, como, por exemplo, na categoria *tradução*, responsável por 41% do total de alterações no Lilt e 40% no pedido depositado. Entretanto, as diferenças começam a aparecer se considerarmos o total geral de alterações: 22 na PE-Lilt e 58 no PD 1.

Na categoria *tradução*, temos um total de nove alterações na pós-edição do Lilt e 23 no pedido depositado. Visto que a PE-Lilt é uma tradução sem segunda revisão, ainda podemos considerar o desempenho aceitável. Ainda mais se levarmos em conta que três das alterações recaem na categoria *local* e dizem respeito a correções para adequação ao padrão bibliográfico brasileiro, que foram percebidas na segunda revisão para preenchimento da planilha de alterações. Por sua vez, com 23 erros em mil palavras, o pedido depositado poderia ser considerado uma tradução de menor qualidade. No entanto, estamos diante do pedido de patente *de facto*, ou seja, conforme depositado no INPI, o que demonstra que o padrão de qualidade esperado (ou, pelo menos, aceito) para uma tradução de patente profissional é consideravelmente mais baixo que o exigido, por exemplo, no modelo de avaliação de uma tradução de TI (tecnologia da informação).

No PD 1, a segunda categoria em porcentagem de alterações foi *estilo*, com 29% e 17 ocorrências. Na PE-Lilt, houve apenas quatro ocorrências (18% do total de alterações). Essas alterações se referem, principalmente, a ajustes no tempo verbal: uma para usar o pretérito em um trecho que falava dos resultados percentuais de um estudo sobre o peso de gatos, outras duas para usar o presente

em vez do futuro composto em uma análise de causa e efeito. Por trabalhar com o conceito de pós-edição completa, que busca qualidade similar à tradução humana, senti a necessidade, ao avaliar o pedido depositado, de alterar vários trechos em busca de maior fluência e clareza, como se vê no quadro abaixo.

Original	PD 1	Correção
It is reported that half of the [...] population	É registrado que mais da metade da população	Relatórios apontam que mais da metade da população
in humans as well as companion animals	em seres humanos, assim como em animais de estimação	em seres humanos e animais de estimação
AAFCO recommendations are based on studies in growing kittens	As recomendações da AAFCO são baseadas em estudos em gatinhos em crescimento	As recomendações da AAFCO são baseadas em estudos com gatos em crescimento
If the animal is underweight, the animal would be fed	Se o animal está abaixo do peso, o animal deveria ser alimentado	Se estiver abaixo do peso, o animal deve ser alimentado

Quadro 7. Exemplos de correções da subcategoria *estilo* (PD 1)

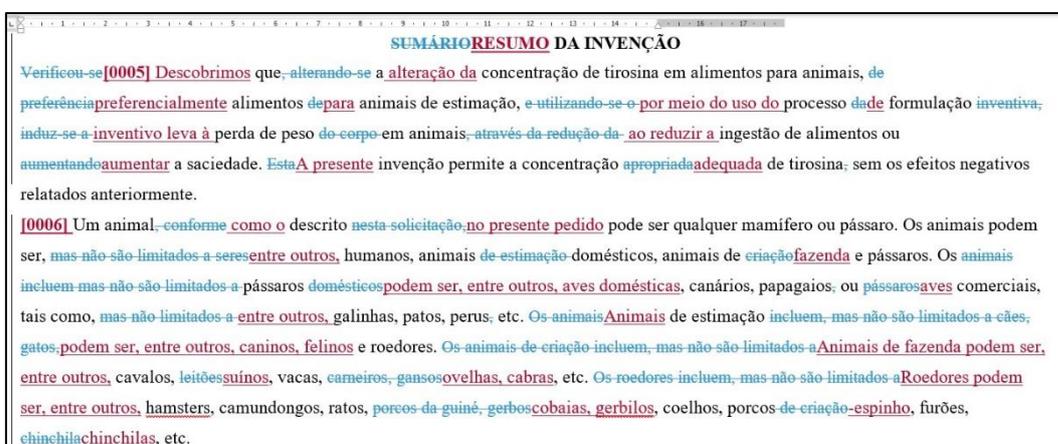
A porcentagem da categoria *linguagem* foi de 27% na PE-Lilt (6 ocorrências) e 22% no PD 1 (13 ocorrências). Em ambos os documentos, as alterações se dividiram entre várias subcategorias, sendo que as categorias *concordância*, *preposições*, *ortografia* e *pontuação* tiveram ocorrências em ambos. O PD 1 também teve três alterações na subcategoria *paralelismo*, tais como “através da redução da ingestão de alimentos ou **umentando** a saciedade”, alterada para “através da redução da ingestão de alimentos ou **do aumento** da saciedade”.

É relevante mencionar que alguns trechos de PD 1 apontam para falta de cuidado com a revisão do conteúdo. Alguns exemplos: na subcategoria *preposição*, ao usar “pesquisa **em** gatos” em vez de “**com** gatos” e “estudos **em** gatinhos” em vez de “**com** gatinhos”); na categoria *tradução*, ao misturar inglês e português no mesmo trecho (“AAFCO Cat Food Nutrient Profiles for Adult Maintenance para fenilalanina” e “NHLBI Obesity Education Initiative Expert Panel sobre a identificação, a avaliação, e os tratamentos de excesso de peso e obesidade em adultos”). São problemas que parecem indicar o uso de tradução automática *sem revisão* em uma época em que a TA ainda não estava tão desenvolvida (o depósito foi feito no INPI em dezembro de 2004).

A categoria *terminologia* teve apenas uma alteração na pós-edição do Lilt e três no pedido depositado (5% das ocorrências em ambos). A ocorrência comum a ambos foi a expressão “tyrosine level”, traduzida como “nível de tirosina” em vez de “teor de tirosina”. O PD 1 ainda sofreu alterações em “material seco”, ajustado para “matéria seca”, e “cabelo” como tradução de “hair”, quando no caso seria “pelo” ou “pelagem”. Esse é outro exemplo de tradução que parece TA não editada. Por fim, na categoria *revisão*, houve duas ocorrências em cada documento. Houve uma alteração comum a ambos, relativa à colocação equivocada de hífen na expressão “não essencial”, que deixou de ter hífen após o Acordo Ortográfico de 2012¹⁹.

5.3.2. Lilt X PD 1: análise comparativa

A figura abaixo, que mostra o início da seção que descreve a invenção, é uma reprodução da tela de comparação de documentos do Microsoft Word. O texto em vermelho corresponde à PE-Lilt e o azul corresponde ao PD 1. Como se pode ver, há uma grande variação entre uma tradução e outra. Uma parte considerável das diferenças, no entanto, está relacionada ao repertório cada tradutor. Exemplos: “verificou-se” ou “descobrimos”, “de preferência” ou “preferencialmente”, “através da redução” ou “ao reduzir”, “apropriada” ou “adequada” são termos equivalentes e poderiam ser usados de maneira intercambiável na tradução. Essa variação é compreensível e esperada, e reflete os lances que cada tradutor dá no jogo da tradução.



¹⁹ Vide: <<http://www.academia.org.br/nossa-lingua/busca-no-vocabulario>>. Acesso em: 12 out. 2018.

Figura 14. Comparação entre PE-Lilt e PD 1

Porém, ainda antes, logo no início do documento, vê-se uma discrepância nos títulos de cada tradução. No PD 1, temos “Composição comestível, métodos de controle do peso corporal de um animal, de reduzir a ingestão de ração e o peso do corpo de um animal com excesso de peso e obeso, de manter o peso de um animal, de aumentar a ingestão de alimento e o peso corporal de um animal com peso abaixo do normal, e, suplemento para animais, petisco para animais ou refeição para animais”. Na PE-Lilt, temos “Tirosina, aminoácido não essencial dietético, regula o peso corporal de animais ao regularizar o apetite ou a ingestão de alimentos”, que corresponde à tradução do título original. É comum que pedidos de patentes mudem seus títulos de país a país, conforme as exigências de cada órgão nacional de propriedade industrial. No documento em análise, é provável que o título do pedido depositado tenha sido definido pelo escritório de patentes, não pelo tradutor, ou talvez até tenha sido mudado por exigência do INPI. No caso da PE-Lilt, o título é meramente uma tradução do título original, sem qualquer adaptação.

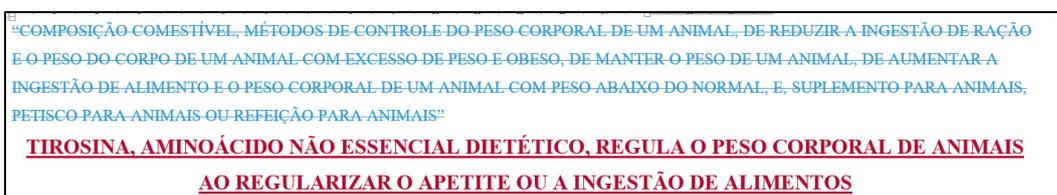


Figura 15. Diferença entre títulos (PE-Lilt e PD 1)

Outra diferença significativa entre PD 1 e PE-Lilt foi a tradução dos títulos de obras de referência, tais como artigos científicos, citadas no pedido original. No PD 1, o tradutor não adotou uma abordagem consistente, alternando entre trechos traduzidos e não traduzidos. Na PE-Lilt, houve uma avaliação caso a caso: títulos de obras de referência foram mantidos em inglês, ao passo que nomes de instituições ou títulos de relatórios explicativos foram traduzidos para ajudar na compreensão do público leitor. Outro aspecto que cabe citar foi o ajuste do espaçamento entre numerais para o padrão brasileiro de citações bibliográficas na PE-Lilt. A normatização dos títulos de obras foi um aspecto relevante no jogo de pós-edição, tanto que a taxonomia abrigou esse tipo de alteração na subcategoria *local*. O quadro a seguir compara a tradução do PD 1 e da PE-Lilt de alguns segmentos do documento.

Original	PD 1	PE-Lilt
Tyrosine injection improves appetite, cognition, and exercise tolerance in activity anorexia in mice	A injeção de tirosina melhora o apetite, a cognição, e exercita a tolerância na anorexia de atividade em camundongos	A injeção de tirosina melhora o apetite, a cognição e a tolerância a exercícios em camundongos com anorexia induzida por atividade física
STUDY 1: Six adult cats (over 1 year of age) were fed ad libitum a canned cat food that was nutritionally complete and balanced	ESTUDO 1: Seis gatos adultos (com mais de 1 ano de idade) foram alimentados “ ad libitum ” por um alimento para gato enlatado que foi completado e balanceado nutricionalmente	ESTUDO 1: Seis gatos adultos (acima de 1 ano de idade) foram alimentados em livre demanda com uma comida de gato enlatada nutricionalmente completa e equilibrada
Complete blood count (CBC) and serum chemistry parameters were within the normal range except mean corpuscular hemoglobin concentration (MCHC) (33.7%) and alanine aminotransferase (ALT) (94.2 u/l)	Os parâmetros de contagem completa do sangue (CBC) e da química do soro estavam dentro da faixa normal, exceto a concentração média corpuscular de hemoglobina (MCHC) (33,7%) e a aminotransferase de alanina (ALT) (94,2 u/l)	Os hemogramas (HGM) e os parâmetros químicos do soro ficaram dentro da faixa normal, com exceção da concentração de hemoglobina corpuscular média (CHCM) (33,7%) e da alanina aminotransferase (ALT) (94,2 u/l)

Quadro 8. Extratos de traduções (PD 1 e PE-Lilt)

Na linha 1, vemos um exemplo de solução da PE-Lilt que funciona melhor que a do PD 1 e outro em que as soluções se equivalem. A expressão “exercise tolerance” foi traduzida, na saída da TA, como “tolerância ao exercício”. Na PE-Lilt, foi corrigida para “tolerância a **exercícios**”. Apesar do erro de ortografia, essa expressão é a mais comumente usada em português brasileiro neste contexto. No PD 1, a expressão aparece como “exercita a tolerância”, uma tradução que, equivocadamente, reconhece o substantivo “exercise” como verbo. No caso da expressão “activity anorexia”, por outro lado, as duas traduções, “anorexia de atividade”²⁰, no PD 1, e “anorexia induzida por atividade física”^{21 22}, na PE-Lilt, parecem ser aceitáveis no jargão da área.

A linha 2 traz uma expressão latina muito usada em inglês, mas pouco usual no português: “ad libitum”. Em inglês, no contexto de alimentação, a expressão significa “sem restrição, livremente”²³. Como se pode ver, ela foi mantida em latim e entre aspas no PD 1 e traduzida para “livre demanda” na PE-Lilt, uma solução que preserva o sentido original. Note-se, também, o trecho “um alimento para gato enlatado” do pedido depositado: por causa do uso da preposição “para”, tem-se a impressão de que o enlatado é o gato, não o alimento.

Por fim, a linha 3 traz alguns exemplos em que as escolhas terminológicas da PE-Lilt foram melhores que as do PD 1, como “hemograma”²⁴ para “complete blood count”, “concentração de hemoglobina corpuscular média (CHCM)”²⁵ para “mean corpuscular hemoglobin concentration (MCHC)” e “alanina aminotransferase”²⁶ para “alanine aminotransferase”. Todos esses termos são mais frequentes que os escolhidos no PD 1, embora “contagem completa do sangue” também seja uma opção aceitável para o primeiro termo da lista. O termo “serum

²⁰ Vide: <http://www.scielo.br/scielo.php?pid=S0102-37722008000300012&script=sci_arttext>. Acesso em 24/10/18.

²¹ Vide: ALMEIDA, Paola E. M. et al., Comportamento Alimentar e Transtorno Alimentar: uma discussão de variáveis determinantes da anorexia e da bulimia. Rev. Bras. de Ter. Comp. Cogn., 2014, Vol. XVI, no. 1, 21-29.

²² Vide: ALMEIDA, P. Transtornos alimentares e atividade esportiva, p. 80. Em: Rubio, Katia. Psicologia do esporte: teoria e prática. São Paulo, Casa do Psicólogo, 2003.

²³ Vide: <https://www.thefreedictionary.com/ad+libitum+feeding>. Acesso em: 04/10/18.

²⁴ Vide: FAILACE, Renato. Hemograma: manual de interpretação. Artmed Editora, 2015.

²⁵ Vide: PAIVA, Adriana A.; RONDÓ, Patrícia HC; GUERRA-SHINOHARA, Elvira M. Parâmetros para avaliação do estado nutricional de ferro. Revista de Saúde Pública, v. 34, p. 421-426, 2000.

²⁶ GOUVEIA, Ericson Cavalcanti et al., Identificação de ponto de corte no nível sérico da alanina aminotransferase para rastreamento da hepatite C em pacientes com insuficiência renal crônica em hemodiálise. Rev. Soc. Bras. Med. Trop., v. 37, n. 1, p. 18-21, 2004.

chemistry parameters” foi traduzido como “parâmetros químicos do soro” na PE-Lilt e “parâmetros [...] da química do soro” no PD 1. Neste caso, a melhor opção, de acordo com a minha pesquisa, seria usar “parâmetros da química sérica”²⁷.

Após a comparação entre os documentos, observou-se que o nível de qualidade da pós-edição no Lilt foi superior ao do pedido depositado. Na PE-Lilt, houve maior preocupação com a pesquisa terminológica, o que levou à escolha de termos mais pertinentes no contexto de produtos alimentícios para animais. A PE-Lilt também mostrou um texto mais fluente, já que, na maioria dos casos, o PD 1 permaneceu muito próximo à estrutura do inglês, o que criou frases mais truncadas em português. Por fim, como se viu no início desta seção, o repertório dos dois tradutores é bastante diverso, o que se revelou até nas escolhas mais simples.

Essa diferença de qualidade na comparação entre a pós-edição da TA do Lilt e a tradução do PD 1 parece apontar para o fato de que o nível de qualidade exigido por empresas que encomendam traduções de pedidos de depósitos de patente não é tão alto quanto o nível buscado em nossa tarefa de pós-edição. Nesse caso, uma explicação possível para o aceite de uma qualidade menor pode estar no fato de que o tempo é um fator essencial quando se trata de pedidos de patente, pois é a data do depósito inicial que define a prioridade de um pedido sobre outro equivalente. Possíveis problemas de tradução poderiam ser resolvidos em uma segunda etapa do processo de análise do mérito do pedido de patente em vista de exigências feitas pelo INPI com base na Lei de Propriedade Industrial. Em todo caso, a comparação entre a pós-edição da TA do Google Tradutor e o pedido depositado do Doc. 2, feita a seguir, pode trazer nova luz sobre a questão da qualidade exigida de pedidos de patentes.

5.3.3. GNMT x PD 2: análise taxonômica

Na segunda parte do jogo 3, vamos fazer a comparação entre a divisão por categorias da pós-edição realizada no memoQ com sugestões da TA do Google Tradutor (doravante PE-GNMT) e da tradução do pedido de patente do Doc. 2, conforme depositado no INPI (doravante PD 2). Novamente, analisamos trechos de

²⁷ ALENCAR, Phillipe Nogueira Barbosa et al., Efeito da homeopatia na função salivar e na morfologia de glândulas parótidas de ratos irradiados. 2013.

cerca de mil palavras de cada documento, seguindo uma prática de avaliação usual da indústria da tradução.

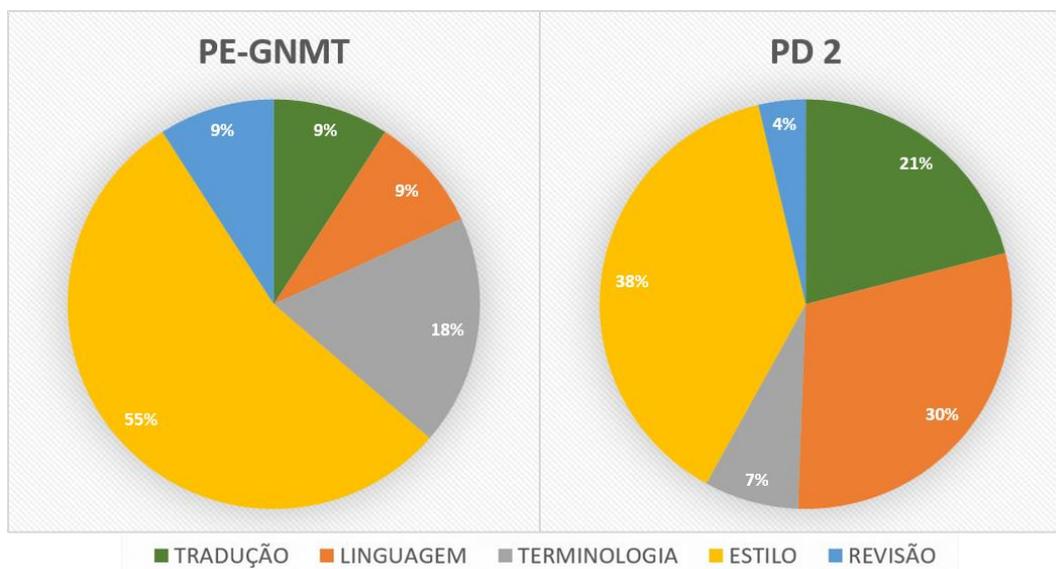


Figura 16. Categorização das edições feitas (PE-GNMT e PD 2)

Ao comparar os gráficos, vemos o predomínio de alterações na categoria *estilo*, com mais da metade das ocorrências na pós-edição do Google Tradutor e 38% no pedido depositado. Poderia ser um indício de qualidade para ambos os documentos, mas a disparidade no total de alterações em cada um deles é um dado decisivo. A PE-GNMT sofreu apenas 11 alterações no total (seis delas em *estilo*). O PD 2, por sua vez, foi o documento mais alterado entre os quatro avaliados no jogo 3, com 81 alterações no total, mais de sete vezes o número de mudanças da PE-GNMT. Desse total, 31 alterações recaíram na categoria *estilo*, o que indica uma grande necessidade de edições para melhorar a fluência e a clareza da tradução.

Além disso, foi preciso alterar o pedido depositado em praticamente todas as subcategorias de *linguagem*, que teve a segunda maior proporção de edições (30%, com 24 alterações no total). Houve seis mudanças classificadas na categoria *paralelismo*, tais como, por exemplo, a presença e a ausência de artigo na mesma sentença ou o uso alternado de singular e plural em enumerações. Também houve cinco mudanças em *pontuação*, notadamente por mau uso de vírgulas, e seis em *preposições*, como, por exemplo: “o animal jovem é apresentado **com** uma nova estrutura social”, em vez de “apresentado **a** uma nova estrutura social”. A PE-GNMT teve uma única alteração em *linguagem*, na subcategoria *paralelismo*.

Na categoria *tradução*, a diferença também foi significativa, com 9% na pós-edição no Google Tradutor e 21% no pedido depositado. Em termos absolutos, isso significa um erro na PE-GNMT e 17 no PD 2, mais um resultado que aponta para problemas do documento depositado no INPI. Apenas uma alteração de tradução no pedido depositado não recaiu na subcategoria *erro de tradução* e se refere a um termo que não foi adaptado para a norma brasileira, com inserção na categoria *local*.

A pós-edição no Google Tradutor teve duas alterações na categoria *terminologia*, o que corresponde a 18% do total de 11 mudanças, enquanto o pedido depositado teve seis, que correspondem a 7% do total de edições. Os números dessa categoria mostram — de forma um tanto exagerada, na verdade — a distorção causada pela leitura apenas dos números, sem se considerar outras variáveis. Embora seja uma virtude da tradução desse documento, o baixo número de alterações na PE-GNMT acabou causando muita distorção nos resultados percentuais. A seção a seguir traz uma comparação entre os documentos, nos moldes da análise anterior, que contrapôs a PE-Lilt e o PD 1. A intenção é ilustrar as diferenças entre a PE-GNMT e o PD 2 para corroborar a visão de superioridade do primeiro produzida pela categorização das alterações nos documentos.

5.3.4. GNMT x PD 2: análise comparativa

De início, a tradução do pedido depositado 2 mostrou mais pontos em comum com a pós-edição do GNMT do que ocorrera entre o pedido depositado 1 e a pós-edição no Lilt. As estruturas das frases da tradução original e da minha pós-edição pareciam mais próximas, como se pode ver na figura abaixo. Como antes, o texto em azul significa a tradução do pedido depositado; em vermelho, a pós-edição no Google Tradutor; em preto, o texto coincidente entre ambos os documentos.

RAÇÃO EXTRUDADA, ÚMIDA	ALIMENTO EXTRUDADO ÚMIDO E MACIA, MACIO
PARA ANIMAIS¶	
<u>ANTECEDENTES</u> FUNDAMENTOS DA INVENÇÃO¶	
<p>O desmame apresenta muitos desafios para um animal jovem. Esses Entre esses desafios incluem uma está a mudança abrupta de uma dieta líquida para uma outra sólida, que contém ingredientes que podem não ser facilmente digeríveis para o animal jovem.</p> <p>Imediatamente após o desmame, o sistema digestivo do animal tem de precisa se adaptar a um novo regime de alimentação com em relação à secreção de enzimas enzimática. Além disso, o animal jovem é apresentado com a uma nova estrutura social. Combinados, esses efeitos rompem prejudicam a ingestão de nutrientes que é necessária para manter a integridade e a função do intestino intestinal. A ingestão reduzida de água e ração comida após o início do desmame pode contribuir para a inflamação intestinal, que pode afetar a estrutura do vilosidades das vilosidades e da cripta. Essas interrupções em Problemas na ingestão de</p>	

Figura 17. Comparação entre PE-GNMT e PD 2

A continuação da análise, no entanto, revelou que as coincidências de texto no início não se manteriam ao longo do documento. Os trechos seguintes mostram grande alternância entre textos em azul, vermelho e preto. Embora o nível de coincidência de texto tenha sido, de fato, um pouco maior na comparação entre PD 2 e PE-GNMT, as divergências em escolhas terminológicas surgiram com a mesma frequência em ambas as comparações. O quadro abaixo apresenta algumas dessas escolhas. Na coluna da direita, avalio os termos usados em ambos os documentos, com comentários sobre qual deles se adequa melhor ao contexto, segundo o melhor do meu conhecimento.

Texto-fonte	PD 2	PE-GNMT	Comentários
nugget	grão	nugget	O termo “grão” está relacionado a “grain”, ou seja, a semente ou fruto de um cereal. “Nugget” não foi traduzido na PE-GNMT por aproximação ao sentido do pedido de patente original: um alimento pequeno e arredondado.

pelleted diets	dietas em pelotas	dietas peletizadas	Em busca no Google Acadêmico, encontramos 316 ocorrências de “dietas peletizadas” e nenhuma de “dietas em pelotas”
bleachable fancy tallow	sebo seletado que pode ser branqueado	sebo	Não consegui confirmar uma tradução consagrada para o termo em inglês. Após buscas no Google Acadêmico e em páginas brasileiras do setor de alimentos para animais, me pareceu que o uso do termo “sebo” atende às necessidades do pedido de patente. Não encontrei referências em português para “sebo seletado que pode ser branqueado”, parece ser uma tradução literal
average daily feed intake (ADFI)	ingestão média diária de ração (ADFI)	ingestão diária média de alimento (IDMA)	Embora ambas as opções sejam adequadas, a tradução na PE-GNMT alterou a sigla para a expressão em português
nursery pigs	leitões de viveiro	leitões de creche	O termo de uso corrente no contexto de criação de animais é “leitões de creche” ²⁸

Quadro 9. Comparação terminológica (PE-GNMT e PD 2)

²⁸ Vide: OLIVEIRA S. J. et al., 2009. Observação de diferentes graus de lesões em estômagos e úlcera gástrica em leitões de creche. Isolamento de *Arcobacter cryaerophilus*. *Acta Scientiae Veterinariae*. 37(2): 119-123 e DE SOUZA, Enio Tonon et al., 2015. Efeito da forma física da ração no desempenho de leitões na fase de creche.

Vê-se, assim, uma maior preocupação com a terminologia na pós-edição do Google Tradutor, o que aponta para o fato de que o pedido depositado 2 segue a mesma dinâmica e tem a mesma finalidade do pedido depositado 1, qual seja, produzir uma tradução de qualidade aceitável no menor prazo possível. Assim, a pressão do tempo leva o tradutor a fazer escolhas mais apressadas pela necessidade de alta produtividade. Vemos, então, que o uso de tradução automática tem potencial para exercer um impacto positivo sobre a qualidade de traduções de patentes. O uso de sistemas de TA contribui para uma maior produtividade do tradutor em termos de palavras traduzidas por hora, o que oferece ao profissional da tradução a possibilidade de dedicar mais tempo à pesquisa de termos.

Além disso, como visto em Guerberof (2014), o tradutor tende a confiar mais na qualidade do texto ao trabalhar com memória de tradução, fazendo menos pesquisas terminológicas; ao trabalhar com tradução automática, por sua vez, parece haver maior desconfiança em relação à qualidade do texto traduzido, o que estimula o tradutor a ser mais criterioso nas pesquisas para validação de terminologia. Dessa forma, a tradução automática oferece uma equação com dois pontos positivos: por um lado, ao aumentar produtividade, oferece mais tempo para pesquisas terminológicas; por outro, ao suscitar uma atitude mais crítica em relação ao texto traduzido automaticamente, leva o tradutor a se preocupar mais com a validação dos termos.

5.4. Resumo

O jogo 1 do presente capítulo mostrou que a tradução automática representa um avanço para o tradutor autônomo em termos de produtividade, sem perda perceptível de qualidade. Ao contrário, a comparação das pós-edições desse jogo com a tradução dos documentos efetivamente depositados no INPI apontam para uma melhor qualidade da PE.

É necessário, no entanto, relativizar os resultados em termos da *dinâmica* e da *finalidade* de cada um desses jogos de tradução: traduções profissionais sofrem uma pressão por prazos curtos de entrega, que pode ser exacerbada na área de patentes, em que, como já mencionado, a data de depósito é decisiva para a prioridade de um pedido sobre outro. Assim, a *finalidade* da tradução profissional

de patentes é entregar o pedido o quanto antes com a melhor qualidade de tradução possível, enquanto, nos jogos desta dissertação, a finalidade era obter a melhor qualidade possível, sem preocupação com prazos. Mesmo assim, a tradução automática permitiu uma produtividade superior ou, no mínimo, similar a estudos com uso de memória de tradução.

O jogo 2 permitiu um aprofundamento ainda maior na avaliação do impacto da pressão dos prazos ao trazer uma comparação entre duas traduções do mesmo texto produzidas por mim com um grande intervalo de tempo (vários anos) entre ambas. De início, a divisão percentual das edições em cada categoria parecia apontar para uma superioridade da tradução antiga (T1) sobre a atual (T2), dado o menor percentual da primeira em categorias como *tradução* (21% contra 27% do total de edições) e *linguagem* (17% contra 27% do total) e, em contraponto, ao maior percentual (46% contra 27%) em *estilo*, que indicaria um predomínio de alterações não relacionadas a “erros”, mas a melhorias no texto.

Ocorre, no entanto, que os números absolutos de ambos os documentos contavam uma história bastante diferente: T1 havia sofrido 48 alterações no total, mais de três vezes o número de edições feitas em T2, apenas 15. Assim, o grande percentual de alterações de *estilo* em T1 corroborava os 22 problemas de fluência de um texto que continha oito erros de *tradução* e dez de *linguagem*, contra apenas quatro em cada categoria em T2. A análise comparativa apontou, também, para uma superioridade da tradução recente, mas que requer uma relativização da qualidade por conta da pressão do tempo sobre a qualidade e de uma eventual menor experiência minha ao traduzir o pedido de patente alguns anos atrás.

A categorização apresentada na segunda parte do jogo 1 mostrou sua *utilidade* nas tarefas de análise das alterações realizadas nos jogos 2 e 3. Além disso, a comparação com estudos semelhantes da literatura apontou para uma semelhança entre a minha taxonomia e outras existentes, validando a categorização e valorizando o que havia de diferente, como a criação de uma categoria, *revisão*, que prevê a categorização de alterações feitas em uma segunda análise do conteúdo para apontar possíveis equívocos de correção. Esta categoria pode ser útil, por exemplo, em avaliações que prevejam múltiplos avaliadores, com uma segunda rodada em que cada avaliador avaliasse a avaliação de um de seus pares.

A comparação entre traduções profissionais de documentos que foram efetivamente depositados como pedidos de patente no jogo 3 confirmou o impacto

da pressão do tempo sobre a qualidade da tradução e também mostrou como a expectativa *real* de qualidade estava abaixo da que me impus em um jogo de pós-edição completa de TA. Por exemplo, as muitas alterações na categoria *estilo*, tanto no PD 1 quanto no PD 2, bem como a diferença considerável na soma de edições da categoria *tradução* nas pós-edições e nos pedidos depositados (dez ocorrências em PE-Lilt/GNMT somados, contra 40 em PD 1/2) ilustram um cuidado meu com a fluência do texto que, em grande medida, não reflete o cuidado dos tradutores profissionais dos textos.

A figura 18, abaixo, mostra os números absolutos de alterações nos documentos analisados nos jogos 2 e 3 deste capítulo, uma comparação visual que, em minha opinião, corrobora a maior qualidade dos documentos traduzidos nas sessões de pós-edição da saída de TA deste trabalho. Como se pode ver, todos os documentos traduzidos com a pós-edição de sugestões de TA tiveram menos alterações que os documentos traduzidos profissionalmente.

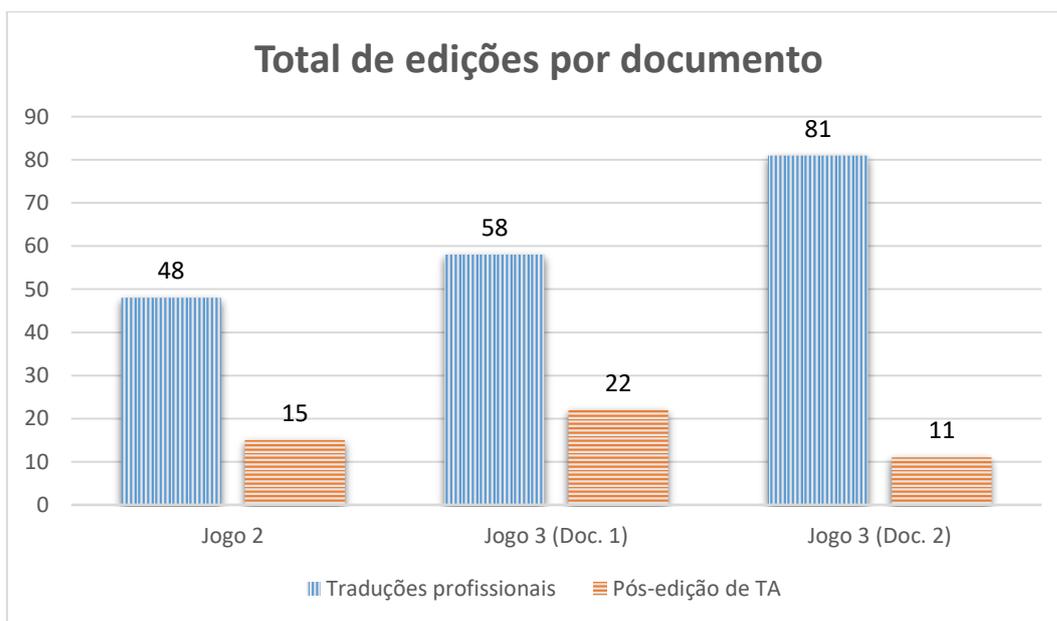


Figura 18. Comparação do volume de edições nos jogos 2 e 3

Se a considerável diferença entre eles em termos de números absolutos já parecia apontar para uma maior qualidade dos primeiros, mesmo sem a necessária análise e relativização dos números, os dados comparativos entre os documentos pós-editados e as traduções profissionais confirmaram a impressão inicial. A análise revelou maior cuidado com estilo e fluência nos primeiros, bem como menos alterações de tradução e de terminologia.

A avaliação ampliada, considerando a produtividade, a categorização e a comparação baseada em trechos das traduções, ofereceu um panorama bastante completo das diferenças entre as traduções e também permitiu avaliar como a percepção de qualidade para a tradução de um documento em nível profissional é perpassada pela tensão entre o tempo disponível e o uso pretendido para o documento, entre outras variáveis, como bem apontaram O'Brien (2012) e Way (2018).

Os resultados obtidos nas análises dos meus jogos confirmaram a importância de se relativizar tanto números absolutos quanto percentagens ao analisar e comparar documentos distintos. Muitas vezes, um critério adotado pode causar distorções na forma como os resultados são apresentados. E acredito que o aspecto mais relevante de todos seja alertar para a necessidade de se analisar com cuidado todos os parâmetros que possam ser adotados em uma avaliação, inclusive as motivações de quem está fornecendo os dados. Toda análise é uma narrativa, e toda narrativa reflete a ideologia de quem está contando a história.

A visão de imparcialidade imputada à ciência é transposta para o texto que se propõe a relatar a experiência científica [...]. Parece-nos evidente que tudo o que foi dito do cientista e do texto científico em geral se aplica ao linguista em particular, verdadeiro manipulador de dados empíricos, cuja inquestionabilidade se vê reforçada hoje pelo uso da estatística e do computador (Coracini, 2003, p. 21).

A narrativa deste trabalho é, assim, minha visão pessoal dos jogos de avaliação, ou seja, de um profissional que tem, por um lado, autoridade para discutir os resultados e apresentar comentários e opiniões relevantes. Por outro, não posso ser absolutamente “racional, impessoal, objetivo” (idem, p. 20), mas um indivíduo com “desejos, fobias, ideologias (experiências e conhecimentos prévios) que integram e determinam” minhas crenças e ações (idem, p. 23). O que espero, com minhas análises, é ter dado uma contribuição relevante para a avaliação de tradução automática pela ótica de um tradutor profissional.

6 Considerações finais

O presente estudo se propôs a avaliar o desempenho e a qualidade da tradução automática de pedidos de patentes do inglês para o português produzida por dois dos principais sistemas de TA disponíveis comercialmente neste momento para tradutores autônomos: Lilt e Google Tradutor. A motivação para a escolha do tema teve uma razão prática: apresentar uma avaliação empírica da tradução automática, com base na ótica de um tradutor profissional.

O estudo consistiu em uma série de três jogos de avaliação de tradução automática. A dinâmica incluiu uma etapa quantitativa para avaliação da produtividade do tradutor ao usar os sistemas de TA, realizada nos jogos 1 e 2, e uma etapa qualitativa, que abrangeu diferentes dinâmicas, como se vê a seguir:

- Jogo 1: pós-edição da saída de TA de dois pedidos de tradução de patentes na área de rações para animais, cada um traduzido em um sistema diferente, e posterior criação de uma taxonomia das edições feitas, definida segundo a ótica de um tradutor profissional autônomo que usa a TA como ferramenta de tradução assistida por computador.
- Jogo 2: comparação entre traduções realizadas pelo mesmo tradutor em momentos distintos, uma com o uso de memória de tradução, outra com o uso de tradução automática.
- Jogo 3: comparação entre as traduções produzidas no processo de pós-edição do jogo 1 e as traduções dos documentos de pedido de patente depositadas no INPI.

A base teórica do presente estudo foi a filosofia da linguagem de Ludwig Wittgenstein, tomando por base os jogos de linguagem descritos nas *Investigações filosóficas*. Alinhados a essa visão, os experimentos não tiveram a pretensão de ser absolutamente controlados, ainda que pese alguma tentativa de sistematicidade, mas funcionarem como jogos de linguagem de tradução, tradução automática, pós-edição, criação de taxonomias e avaliação em que eu fui o principal jogador. Refletiram, assim,

a minha visão e a minha prática dos lances, das dinâmicas e da finalidade de cada jogo de tradução automática.

O jogo 1 mostrou, com os jogos da pós-edição, que o uso da tradução automática como ferramenta auxiliar é uma estratégia útil para o aumento da produtividade do tradutor sem perda da qualidade do texto-alvo. As médias de produtividade do Lilt e do GNMT nas tarefas de pós-edição — 755 e 693 palavras por hora, respectivamente — se mostraram superiores às médias de produtividade diária de um tradutor divulgadas pela Common Sense Advisory (2012) — 335 palavras/hora — e aos experimentos com sistemas de memória de tradução sem correspondência (“*no match*”) realizados em Bowker (2005) e Guerberof (2014) — 459 e 647,4 palavras/hora, respectivamente.

Os resultados quantitativos apontaram para um desempenho superior da pós-edição do jogo 2 quando comparada à literatura: a produtividade atingiu 907,1 palavras/hora, quando considerada sem revisão, um desempenho que só fica abaixo das médias obtidas em Guerberof (2014) para pós-edição de tradução automática e para correspondências de memória de tradução entre 84% e 95% — 1.032,6 e 957 palavras/hora, respectivamente.

Os indícios de manutenção da qualidade descritos no jogo 1 foram corroborados pela comparação entre duas traduções minhas do mesmo texto, uma feita com um sistema de memória de tradução (T1), outra feita na plataforma online do Lilt (T2), durante o jogo de avaliação do jogo 2. A comparação mostrou que T2 foi superior a T1 no que diz respeito às escolhas terminológicas, sem perda perceptível de fluência e clareza. A possibilidade de que as sugestões da TA do Lilt influenciassem a tradução, orientando as escolhas de termos e a estruturação das frases de forma semelhante ao original, não se confirmou. De modo geral, houve maior esforço de validação dos termos em T2 e a estruturação do texto foi muito semelhante em ambas as traduções, indicando que meu estilo de tradução se manteve.

A maior qualidade percebida em T2 se deve, em parte, à pressão que o tempo exerce sobre o trabalho do tradutor. O cumprimento de prazos é um ponto nevrálgico da indústria de tradução de hoje, e o disseminação do uso de ferramentas computacionais ao longo das últimas décadas exerceu uma forte pressão pela redução dos prazos propostos para a entrega de traduções, um movimento que teve como consequência a relativização do conceito de qualidade de tradução, com diferentes níveis de exigência em vista do tempo, da utilidade, da natureza e da precibilidade do texto, conforme abordado no capítulo 3 deste trabalho (vide citações a O’Brien, 2012 e Way, 2018).

O jogo 3, por sua vez, traz um exemplo concreto dessa relativização de qualidade. Quando comparados às traduções que fiz com o auxílio da saída da TA do Lilt e do Google Tradutor, os pedidos de patente efetivamente depositados no INPI mostraram mais problemas terminológicos e menor preocupação com fluência. Como aventado anteriormente, a pressão do tempo é uma característica premente da indústria de patentes: uma data de depósito mais antiga pode significar a precedência de um pedido sobre outro. E essa precedência pode significar um aumento substancial nos lucros da empresa que obtiver a concessão de patente. Quando editei a saída de TA para este estudo, a qualidade teve precedência sobre o tempo, interessava produzir o melhor texto possível dentro do meu jogo de tradução com auxílio de TA. Quando o tradutor traduziu a patente para depósito, a precedência era do tempo. Natural que o esforço de validação de termos, por exemplo, tivesse um peso muito maior na dinâmica do meu jogo do que na tradução profissional.

Assim, é uma contribuição desta dissertação o desenvolvimento e a aplicação de métricas que põem em xeque a crença – comum entre tradutores, inclusive – de que o uso de tradução automática prejudica a qualidade do trabalho final do tradutor. Pelo contrário: os resultados da comparação entre as pós-edições feitas para este trabalho e traduções feitas em contextos profissionais evidenciam a qualidade superior das primeiras. Logo, além do ganho de tempo, a TA também permitiu um comprovado ganho de qualidade de tradução.

A etapa de criação de uma taxonomia das edições feitas durante a avaliação do jogo 1 tinha por finalidade gerar não apenas uma classificação que refletisse a visão de um tradutor sobre o processo de pós-edição de TA, mas também uma categorização que fosse útil para tradutores e professores de tradução, por exemplo. No caso destes, o primeiro aspecto, e o mais visível, é mostrar a estudantes de tradução como a TA pode ser uma ferramenta útil para diversos usos do tradutor autônomo. Outro aspecto relevante é a possibilidade de analisar as alterações feitas no processo de pós-edição para mostrar os pontos fortes e também as limitações da TA quando comparada a outras ferramentas computadorizadas disponíveis para o tradutor. Ou, ainda, apresentar a estudantes, por meio das tabelas que comparam as diferentes dinâmicas de tradução — TA, PE, tradução humana — as nuances e sutilezas do trabalho do tradutor. E, por fim, a discussão sobre os usos da TA também fomenta a atualíssima discussão sobre os diferentes níveis de qualidade esperados de uma tradução no mercado de hoje e sobre o panorama para a profissão de tradutor em vista da tecnologia.

A taxonomia deste estudo apresentou uma visão que se alinhou às classificações de estudos anteriores e seguiu a ideia de que, para ser informativa e útil, uma categoria precisa ser abrangente. Também buscamos fornecer uma descrição clara das categorias, de forma a propiciar o maior nível de consenso possível entre avaliadores. Esperamos que essa taxonomia possa ser útil para promover um bom nível de concordância entre os avaliadores que porventura venham a se utilizar dela para produzir avaliações de tradução automática centradas no trabalho do tradutor no futuro.

Por fim, outra contribuição deste trabalho é a abordagem da avaliação como um jogo de linguagem, uma visão não-logocêntrica que dificilmente aparece no contexto da avaliação de TA. Propus uma avaliação que discute números e categorias, mas também impressões, que traz a parcialidade e assume a impossibilidade de um total controle das variáveis em jogo. Esta dissertação comunica, enfim, uma perspectiva sobre o jogo da tradução automática. “Qual a utilidade dela?”, pode-se perguntar, em estilo wittgensteiniano. Ora, respondemos, é útil a todos que queiram ter a visão de um tradutor profissional sobre a TA.

As referências analisadas neste trabalho trazem, entre outras informações, descrições sobre desempenho de sistemas, sobre desempenho de grupos de tradutores e sobre categorizações de erros. Aqui trazemos esses aspectos e incluímos a visão de um indivíduo que usou a tradução automática como ferramenta em um recorte do tempo. É um testemunho do uso da TA hoje, no ano de 2018. E o testemunho do hoje ajuda a contextualizar o uso de ontem e, talvez, a vislumbrar o uso do amanhã. É, por fim, um depoimento a favor da TA como ferramenta. Que seja útil a outros tradutores profissionais que pretendem trazer a TA para a sua prática diária, o que, em nossa opinião, é parte indissociável do futuro do trabalho de tradução a médio e, possivelmente, a longo prazo.

7

Referências bibliográficas

ALMEIDA, Giselle. **Translating the post-editor**: an investigation of post-editing changes and correlations with professional experience across two Romance languages. Tese (Doutorado) - Dublin City University, Dublin, 2013.

ALVES, Regysane. **A crítica de traduções na teoria e na prática**: o caso da “Versão Brasileira”. Dissertação (Mestrado) - PUC-Rio, Rio de Janeiro, 2009.

AZIZ, Wilker; SPECIA, Lucia. Fully automatic compilation of Portuguese-English and Portuguese-Spanish parallel corpora. In: **Proceedings of the 8th Brazilian Symposium in Information and Human Language Technology**. 2011.

BARREIRO, Anabela; MOTA, Cristina. e-PACT: eSPERTo Paraphrase Aligned Corpus of EN-EP/BP Translations. In: **Tradução em Revista**, v. 22, 2017, pp. 88 et. seq.

BENTIVOGLI, Luisa; BISAZZA, Arianna; CETTOLO Mauro; FEDERICO, Marcello. Neural versus phrase-based machine translation quality: a case study. In: **arXiv preprint arXiv:1608.04631**, 2016.

BENTIVOGLI, Luisa; BERTOLDI, Nicola; CETTOLO Mauro; FEDERICO, Marcello; NEGRI, Matteo; TURCHI, Marco. On the evaluation of adaptive machine translation for human post-editing. In: **IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)**, v. 24, n. 2, 2016, pp. 388-399.

BOWKER, Lynne. Towards a methodology for a corpus-based approach to translation evaluation. In: **Meta**, v. 46, n. 2, 2001, pp. 345-364.

_____. **Computer-aided translation technology**: A practical introduction. University of Ottawa Press, 2002.

_____. Productivity vs quality. A pilot study on the impact of translation memory systems. In: **Localisation Focus**, v. 4, n. 1, 2005, pp. 13-20.

CADWELL, Patrick; CASTILHO, Sheila; O'BRIEN, Sharon; MITCHELL, Linda. Human factors in machine translation and post-editing among institutional translators. In: **Translation Spaces**, v. 5, n. 2, 2016, pp. 222-243.

CASELI, Helena. O uso de corpora paralelos para a criação de um tradutor automático estatístico. In: VIANA, V.; TAGNIN, S. E. O. **Corpora na Tradução**. São Paulo: HUB Editorial, 2015, pp. 243-267.

CASTILHO, Sheila et al. Approaches to Human and Machine Translation Quality Assessment. In: **Translation Quality Assessment**. Springer, Cham, 2018, pp. 9-38.

CHAMPOLLION, Yves. Machine Translation Acceptance Among Professional Linguists: Are We Nearing the Tipping Point? In: **Proceedings of AMTA 2016**, Vol. 2: MT Users' Track. Austin, Texas, EUA, out/nov, 2016, pp. 82-85.

CHO, Kyunghyun; VAN MERRIENBOER, Bart; BAHDANAU, Dzmitry; BENGIO, Yoshua. On the properties of neural machine translation: Encoder-decoder approaches. In: **arXiv preprint arXiv:1409.1259**, 2014.

CORACINI, Maria José. O cientista e a noção de sujeito na lingüística: expressão de liberdade ou submissão. In: **O signo desconstruído: implicações para a tradução, a leitura e o ensino**. 2. ed. Campinas: Pontes, 2003.

COSTA, Angela; LUÍS, Tiago; COHEUR, Luísa. Translation errors from English to Portuguese. In: **Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC)**, 2014, pp. 1231-1234.

DILLINGER, Mike. Introduction. In: BALLING et al. (Orgs.). **Post Editing of Machine Translation: Processes and Applications**. Cambridge Scholars Publishing, 2014.

DRUGAN, Joanna. **Quality in professional translation: Assessment and improvement**. A&C Black, 2013.

ELLIS, John Martin. **Language, thought and logic**. Evanston, Illinois: Northwestern University Press, 1993.

ESTILL, Daniel Argolo. **De artesanato a indústria: a tradução globalizada**. Tese (Doutorado) - PUC-Rio, Rio de Janeiro, 2017.

FREIGANG, Karl-Heinz. Machine-aided translation. In: BAKER, Mona. **Routledge Encyclopedia of Translation Studies**, 1998, pp. 134-136.

FREITAS, Bianca. **O dizer em português: diálogos entre tradução, descrição e lingüística computacional**. Dissertação (Mestrado) - PUC-Rio, Rio de Janeiro, 2016.

GLOCK, Hans Johann. **Dicionário Wittgenstein**. Tradução de Helena Martins. Rio de Janeiro: Jorge Zahar Editor, 1998.

GUERBEROF, Ana. Productivity and quality in MT post-editing. In: **MT Summit XII-Workshop: Beyond Translation Memories: New Tools for Translators MT**, 2009.

_____. Correlations between productivity and quality when post-editing in a professional context. In: **Machine Translation**, v. 28, n. 3-4, 2014, pp. 165-186.

GUHA, Jyoti; HEGER, Carmen. Machine translation for global e-commerce on ebay. In: **Proceedings of the AMTA**, 2014, pp. 31-37.

HUTCHINS, John. **The history of machine translation in a nutshell**. Disponível em: <<http://hutchinsweb.me.uk/Nutshell-2005.pdf>>. Acesso em: 02 out. 2017

_____. Machine Translation: General Overview. In: MITKOV, R. (Org.). **The Oxford Handbook of Computational Linguistics**. Oxford University Press, 2005.

JOHNSON, M; SCHUSTER, M; QUOC, V. LE; KRIKUN, M; WU, Y; CHEN, Z; THORAT, N; VIÉGAS, F; WATTENBERG, M; CORRADO, G; HUGHES, M; DEAN, J. Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation. In: **arXiv:1611.04558v2 [cs.CL]**, 2017.

KAY, Martin. Introduction. In: MITKOV, R. (Org.). **The Oxford Handbook of Computational Linguistics**. Oxford University Press, 2005.

KOEHN, Philipp et al. Moses: Open source toolkit for statistical machine translation. In: **Proceedings of the 45th annual meeting of the ACL on interactive poster and demonstration sessions**. Association for Computational Linguistics, 2007, pp. 177-180.

KOEHN, Philipp. **The State of Neural Machine Translation (NMT)**. Disponível em: <<https://omniscien.com/state-neural-machine-translation-nmt/>>. Acesso em: 13 out. 2018.

LOMMEL, Arle; USZKOREIT, Hans; BURCHARDT, Aljoscha. Multidimensional quality metrics (MQM): A framework for declaring and describing translation quality metrics. In: **Tradumática**, n. 12, 2014, pp. 455-463.

LUONG, T.; CHO, K.; MANNING, C. **Neural Machine Translation**. Disponível em: <<https://nlp.stanford.edu/projects/nmt/Luong-Cho-Manning-NMT-ACL2016-v4.pdf>>. Acesso em: 18 ago. 2017.

MARTINS, Débora; CASELI, Helena. Anotação manual de erros de tradução automática em textos traduzidos de inglês para português do Brasil. In: **NILC-TR-13-02**, 2013.

MARTINS, Helena. Sobre a estabilidade do significado em Wittgenstein. In: **Veredas Revista de Estudos Linguísticos**, v. 4, n. 2, 2000, pp. 19-42.

O'BRIEN, Sharon; ROTURIER, Johann; ALMEIDA, Giselle. Post-Editing MT Output. Views from the researcher, trainer, publisher and practitioner. In: **Machine Translation Summit XII (MTS)**, 2009.

O'BRIEN, Sharon; BALLING, Laura; CARL, Michael; SIMARD, Michael; SPECIA, Lucia. Foreword. In: **Post-editing of machine translation: Processes and applications**. Cambridge Scholars Publishing, 2014.

O'BRIEN, Sharon. Towards a dynamic quality evaluation model for translation. In: **The Journal of Specialised Translation**, v. 17, 2012, pp. 55-77.

PAPINENI, K., ROUKOS, S., WARD, T., & ZHU, W. J. BLEU: A Method for Automatic Evaluation of Machine Translation. In: **Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)**. Filadélfia: 2002, p. 311-318.

PENKALE, Sergio; WAY, Andy. Tailor-made Quality-controlled Translation. In: **Proceedings of Translating and the Computer**, v. 35, 2013.

PICININI, Silvio. When MT does not take translators' jobs away - and may create more jobs. **eMpTy Pages**. Disponível em: <<http://kv-emptypages.blogspot.com.br/2016/07/when-mt-does-not-take-translators-jobs.html>>. Acesso em: 16 dez. 2016.

_____. Tradução automática e tradutores: ideias sobre o futuro. In: **IX Congresso Internacional da Associação Brasileira de Tradutores (ABRATES)**: Rio de Janeiro, 2018.

PLITT, M.; MASSELOT, F. A Productivity Test of Statistical Machine Translation Post-Editing in a Typical Localisation Context. In: **The Prague Bulletin of Mathematical Linguistics**, n. 93, jan. 2010, Praga, pp. 7-16.

SANTOS, Diana. A tradução na sociedade do conhecimento. In: **IX Seminário de Tradução e Técnica em Língua Portuguesa**, Lisboa: Instituto Franco-Português, 13 de novembro de 2006.

SCHLEIERMACHER, Friedrich. Sobre os diferentes métodos de tradução. Tradução de Margarete von Mühlen Poll. In: **Scientia Traductionis**, n. 9, 2011.

SNOVER, M., DORR, B., SCHWARTZ, R., MICCIULLA, L., MAKHOUL, J. A study of translation edit rate with targeted human annotation. In: **Proceedings of association for machine translation in the Americas**, v. 6, 2006.

SPARCK JONES, Karen. Natural language processing: a historical review. In: **Computational Linguistics**, 2001.

TURING, Alan. Computing machinery and intelligence. In: **Mind**, v. 59, n. 236, 1950, pp. 433 et. seq.

WAY, Andy. Traditional and emerging use-cases for machine translation. In: **Proceedings of Translating and the Computer**, v. 35, 2013.

_____. Quality expectations of machine translation. In: **arXiv preprint arXiv:1803.08409**, 2018.

WEAVER, Warren. Translation. In: **Rockefeller Foundation**, 1949.

WILKS, Yorick. The History of Natural Language Processing and Machine Translation. In: **Encyclopedia of Language and Linguistics**. Amsterdã: Kluwer, 2005.

WITTGENSTEIN, Ludwig. **Investigações filosóficas**. Tradução de José Carlos Bruni. Os pensadores. 2. ed. São Paulo: Abril Cultura, [1953] 1979.

WU, Y., SCHUSTER, M., CHEN, Z., LE, Q. V., NOROUZI, M. Google's neural machine translation system: Bridging the gap between human and machine translation. In: **arXiv preprint arXiv:1609.08144**, 2016.

Anexo 1

Extrato da planilha de registro de edições do presente trabalho (jogo 1).

Texto original	Saída da TA	Texto pós-editado	Comentários
[00037] Kibbles can also be made from the dough using a baking process, rather than extrusion, where the dough is placed into some sort of mold prior to dry heat processing.	[00037] Kibbles podem ser feitas a partir da a massa usando um processo de cozedura, em vez de extrusão, onde a massa é colocada em uma espécie de bolor prévia ao calor seco de processamento.	[00037] Os biscoitos também podem também ser feitos da massa através de um processo de cocção em forno, em vez de extrusão, no qual a massa é colocada em algum tipo de molde antes do processamento por calor seco.	1. Tradução/Não traduzido: “kibbles” 2. Tradução/Erro de tradução: “da a massa”; 3. Terminologia: “cozedura” 4. Linguagem/Preposição: “onde a massa” 5. Terminologia: “bolor”; 6. Tradução/Erro de tradução: “prévia ao calor seco de processamento”
[00038] Animal treats of the present invention can be prepared by extrusion or baking similar to dry food as is common for commercial pet food.	[00038] Animal trata da presente invenção pode ser preparado por extrusão ou assando semelhantes para a comida seca como é comum para comercial petshop.	[00038] As guloseimas para animais da presente invenção podem ser preparadas por um processo de extrusão ou cocção em forno similar ao da comida seca, como é comum para alimentos comerciais para animais.	1. Tradução/Erro de tradução: “Animal trata da presente invenção” 2. Tradução/Erro de tradução: “ou assando semelhantes” 3. Tradução/Erro de tradução: “comercial petshop”
Other processes may also be used to either coat tyrosine on the exterior of existing treat forms, or inject it inside a treat.	Outros processos podem também ser utilizados para também casaco tirosina no exterior existente de tratar formulários, ou injetar dentro de um tratamento.	Outros processos também podem ser usados para revestir de tirosina a parte externa de formas de guloseimas existentes, ou injetá-la dentro da guloseima.	1. Preferencial: “podem também” 2. Tradução/erro de tradução: “utilizados para também casaco tirosina no exterior existente de tratar formulários, ou injetar dentro de um tratamento”

Anexo 2

Exemplos de traduções do PD 1 e da PE-Lilt (jogo 3).

Código de cores:

- **Amarelo** significa melhores soluções na PE-Lilt
- **Verde** significa melhores soluções no Golden 1
- **Cinza** significa soluções aceitáveis em ambas as traduções
- **Azul** significa soluções problemáticas em ambas as traduções.

Original	PD 1	PE-Lilt
A treat or an animal snack is usually defined as a product that is given to an animal that entices them to eat usually during a non-meal time.	Um petisco ou uma refeição leve para animal usualmente é definida como um produto que é dado a um animal que é atraído para comer o mesmo usualmente durante um tempo fora das refeições normais.	Guloseimas ou petiscos para animais são usualmente definidos como um produto dado a um animal para estimulá-lo a comer, geralmente fora do horário normal de alimentação.
Examples of substances with a stabilizing effect optionally contained in the animal food product of the invention are preservatives, antioxidants, synergists and sequestrants, packaging gases, stabilizers, emulsifiers, thickeners, gelling agents, and humectants.	Exemplos de substâncias com um efeito estabilizante opcionalmente contidas no produto de alimento para animal da invenção, são conservantes, antioxidantes, sinergistas e seqüestrantes, gases para embalagem, estabilizantes, emulsificantes, espessantes, agentes de geleificação, e umectantes	Exemplos de substâncias com efeito estabilizador opcionalmente contidas no produto alimentício para animais da invenção são conservantes, antioxidantes, agentes sinérgicos e sequestrantes, gases de embalagem, estabilizantes, emulsificantes, espessantes, gelificantes e umectantes.
Other processes may also be used to either coat tyrosine on the exterior of existing treat forms, or inject it inside a treat.	Outros processos poderão também ser utilizados para revestir a tirosina no exterior de formas existentes de refeições, ou injetar a mesma dentro de uma refeição.	Outros processos também podem ser usados para revestir de tirosina a parte externa de formas de guloseimas existentes, ou injetá-la dentro da guloseima.
Weekly average ME intake was about 130 kcal/day (Table 1)	A ingestão média semanal de ME era em torno de 130 kcal/dia (tabela 1)	Média semanal de ingestão de EM de cerca de 130 kcal/dia (Tabela 1)

Anexo 3

Exemplos de traduções de termos no PD 2 e na PE-GNMT, com comentários (jogo 3).

Código de cores:

- **Amarelo** significa melhores soluções na PE-Lilt
- **Verde** significa melhores soluções no Golden 1
- **Cinza** significa soluções aceitáveis em ambas as traduções
- **Azul** significa soluções problemáticas em ambas as traduções.

Texto-fonte	Golden 2	PE-GNMT	Comentário
villus	vilo	vilosidade	O termo “vilo” não foi encontrado em português; os termos “vilosidade” e “villus” têm a mesma definição segundo os dicionários Houaiss ²⁹ e Merriam-Webster online ³⁰
shelf stability	estabilidade de armazenagem	estabilidade de prateleira	As duas opções estão corretas
yellow grease	banha amarela	gordura amarela ³¹	Minhas pesquisas indicaram para o uso de “gordura amarela” no contexto de alimentos. Não encontrei referências a “banha amarela” nesse contexto no Google Acadêmico
torula yeast	levedura de tórula	levedura de torula ³²	“torula” não tem acento nesse contexto
livestock	animais domésticos	gado	Neste contexto, as opções mais adequadas seriam os termos “semoventes” ou “animais de fazenda”; as opções das traduções limitam o sentido original da palavra “livestock”
weanling pigs	leitões recém-desmamados	porcos recém-desmamados	A expressão “recém-desmamados” é adequada, porém a PE usou porcos em vez de leitões e não usou o hífen

²⁹ Dicionário Houaiss eletrônico. Rio de Janeiro: Objetiva, 2009.

³⁰ Disponível em: <<https://www.merriam-webster.com/dictionary/nugget>>. Acesso em: 27 out. 2018.

³¹ Vide: LÓPEZ, S. E.; LÓPEZ, J.; STUMPF JUNIOR, W. Parâmetros séricos de vacas leiteiras na fase inicial de lactação suplementadas com diferentes fontes de gordura. Arch. Latinoam. Prod. Anim., 12(3), 2004, pp. 96-102. Outra fonte: <http://www.cnpsa.embrapa.br/sgc/sgc_publicacoes/publicacao_n5r89e7u.pdf>. Acesso em: 27 out. 2018.

³² Vide: <<https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/910280/1/CursoManejodePragas.NiltonSanchez.pdf>>. Acesso em: 27 out. 2018.

Anexo 4

Número de alterações por categoria da taxonomia (T1 e T2) (jogo 2).

CAT. PRINCIPAL	CAT. SECUNDÁRIA	T1	T2
TRADUÇÃO	Adição		
	Erro de tradução	8	4
	Erro do original		
	Local		
	Exclusão		
	Ordem errada		
	Não traduzido		
LINGUAGEM	Concordância	1	1
	Preposições	4	
	Determinantes	2	2
	Ortografia		
	Paralelismo	2	
	Pontuação	1	1
	Registro		
TERMINOLOGIA		5	1
ESTILO		22	4
REVISÃO	Erro de correção		1
	Correção desnecessária		1
	Revisão incorreta		
	Falta de correção	3	
	TOTAL	48	15

Anexo 5

Número de alterações por categoria da taxonomia (PE-Lilt e PD 1) (jogo 3).

CAT. PRINCIPAL	CAT. SECUNDÁRIA	PE-Lilt	PD 1
TRADUÇÃO	Adição		
	Erro de tradução	5	21
	Erro do original		
	Local	3	2
	Exclusão		
	Ordem errada		
	Não traduzido	1	
LINGUAGEM	Concordância	3	3
	Preposições	1	3
	Determinantes		
	Ortografia	1	1
	Paralelismo		3
	Pontuação	1	3
	Registro		
TERMINOLOGIA		1	3
ESTILO		4	17
REVISÃO	Erro de correção		1
	Correção desnecessária		
	Revisão incorreta	1	1
	Falta de correção	1	
	TOTAL	22	58

Anexo 6

Número de alterações por categoria da taxonomia (PE-GNMT e PD 2) (jogo 3).

CAT. PRINCIPAL	CAT. SECUNDÁRIA	PE-GNMT	PD 2
TRADUÇÃO	Adição		
	Erro de tradução	1	16
	Erro do original		
	Local		1
	Exclusão		
	Ordem errada		
	Não traduzido		
LINGUAGEM	Concordância		2
	Preposições		6
	Determinantes		4
	Ortografia		1
	Paralelismo	1	6
	Pontuação		5
	Registro		
TERMINOLOGIA		2	6
ESTILO		6	31
REVISÃO	Erro de correção	1	3
	Correção desnecessária		
	Revisão incorreta		
	Falta de correção		
	TOTAL	11	81