

**A TRADUÇÃO AUTOMÁTICA  
E O PAPEL DO TRADUTOR COMO PÓS-EDITOR**

**LEILA LACERDA BAIA**

Relatório Final de Estágio Profissional apresentado à Escola Superior de  
Educação de Bragança para obtenção do Grau de Mestre em Tradução

Orientado por:

**Isabel Augusta Chumbo**

**Bragança  
Dezembro, 2022**



# **A TRADUÇÃO AUTOMÁTICA E O PAPEL DO TRADUTOR COMO PÓS-EDITOR**

**LEILA LACERDA BAIA**

Relatório Final de Estágio Profissional apresentado à Escola Superior de Educação  
de Bragança para obtenção do Grau de Mestre em Tradução

Orientado por:  
**Isabel Augusta Chumbo**

**Bragança**  
**Dezembro, 2022**

## **Agradecimentos**

A conclusão deste curso só foi possível pela presença e atuação de Deus em minha vida. Sem Ele nada do que empreendi poderia ter sido realizado de maneira tão especial.

Agradeço à minha família (pai, mãe, Hideu, Léa, Zan, Júlia, Elizeu e Paulo) que, mesmo à distância, se fez presente através da torcida, incentivo e orações, sempre acompanhando meu desempenho e caminhada em terras portuguesas.

À família Cañete (Daniela, Jaime, Carlos Hugo e Maria Clara) por abrir as portas da sua casa e me acolher no período em que estive a realizar meu estágio profissional em Barcelona. Sou imensamente grata pelo suporte oferecido em um momento de tanta vulnerabilidade e incertezas por conta da pandemia.

Aos amigos e irmãos em Cristo, em especial à querida Paula Campos, que demonstraram o valor de uma verdadeira amizade e deixaram claro que eu nunca estive sozinha.

À professora Isabel Chumbo pelo incentivo, paciência e suporte em todos os momentos do desenvolvimento deste trabalho.

À professora Cláudia Martins por se disponibilizar a me buscar em Espanha no momento do tão esperado retorno à Portugal após o estágio em Barcelona.

À Quicksilver Translate por proporcionar a experiência do estágio profissional, pelo acolhimento e oportunidade de conhecer o funcionamento de uma agência de tradução na prática.

À querida Mariana Casseiro pelo apoio certo na elaboração desse trabalho.



## **Resumo**

A pós-edição tem sido uma atividade cada vez mais presente no cenário da prestação de serviços de tradução e isto ocorre, principalmente, devido à demanda crescente do mercado da tradução no mundo. Segundo a estimativa da Statista Inc e CSA, o mercado global da tradução arrecadou entre 2019 e 2020 cerca de 50 mil milhões de dólares, o que sugere um valor de arrecadação de 20 mil milhões de euros somente para os Estados-Membros da União Européia. Há décadas que já não é possível conceber o trabalho do tradutor sem o uso de ferramentas computacionais e, recentemente, está cada vez mais presente o uso da tradução automática, processo conhecido como Googlish. Sendo a pós-edição o processo de melhoria da tradução automática com o mínimo de trabalho manual e devido às exigências do mercado da tradução quanto à otimização de tempo e aumento da produtividade, é perceptível que o tradutor precisa estar preparado para assumir a atribuição de pós-editor. Este trabalho se propõe a apresentar o histórico e a análise da evolução da tradução automática e a contextualizar o seu uso na perspectiva atual do tradutor. A partir da experiência adquirida na realização do estágio profissional na empresa Quicksilver Translate, e por meio da análise dos trabalhos desenvolvidos, especialmente o trabalho de pós-edição, foi possível observar a necessidade de formação específica do tradutor para o aperfeiçoamento das competências linguísticas, de tradução especializada, de gestão terminológica, competências informáticas e o aperfeiçoamento em tecnologia linguística, que é a base do desenvolvimento das ferramentas de apoio ao tradutor. Por meio dessa análise fica evidente que a tecnologia é a maior aliada do tradutor que está preparado para gerir as ferramentas e utilizá-las de maneira a atingir seu maior potencial no intuito de fornecer uma tradução com a credibilidade e qualidade esperadas pelo cliente.

Palavras-chave: tradução automática, ferramentas de tradução assistida por computador, pós-edição, qualidade, produtividade.

## **Abstract**

Post-editing has been an increasingly relevant activity in the translation service provider scenario, mainly due to the growing demand of the translation market worldwide. According to the estimate of Statista Inc and CSA, the global translation market has billed about \$50 billion between 2019 and 2020, which suggests a revenue figure of €20 billion for EU member states alone. It has been decades since it has been possible to conceive the translator's work without the use of computer tools, and recently the use of machine translation, a process known as Googlish. Since post-editing is the process of improving machine translation with minimal manual work, and due to the demands of the translation market for time optimization and increased productivity, it becomes apparent that the translator needs to be prepared to take on the assignment of post-editor. This report proposes to present the history and analysis of the evolution of machine translation and to contextualize its use in the translator's current perspective. Based on the experience acquired during the professional internship at Quicksilver Translate, and through the analysis of the work carried out, especially regarding post-editing, I was able to observe the need for specific training for translators to improve their linguistic skills, specialized translation skills, terminology management skills, computer skills and in linguistic technology, which is the basis for the development of the translator's support tools. Through this analysis, it becomes obvious that technology is the translator's greatest ally, and that the translator must be prepared to manage the tools and use them in such a way as to reach their greatest potential, to provide a translation with the credibility and quality expected by the client.

**Keywords:** machine translation, computer-assisted translation tools, post-editing, quality, productivity.



## Índice

Agradecimentos .....	i
Resumo .....	iii
Abstract.....	v
Siglas e Abreviaturas .....	ix
Introdução.....	1
1. Tradução e Tecnologia .....	4
1.1 Período de Germinação (1967-1983).....	6
1.2 Período de Crescimento Constante (1984-1992) .....	10
1.3 Período de Rápido Crescimento (1993-2003) .....	11
1.4 Período de Desenvolvimento Global (2004-2013) .....	13
1.5 Tradução Automática na Era da Inteligência Artificial .....	16
1.6 Desafios e Limitações da Tradução Automática .....	19
1.7 Aplicações da Tradução Automática .....	21
1.8 Abordagens à Tradução Automática.....	22
1.8.1 Abordagem Baseada em Regras.....	23
1.8.2 Abordagem Baseada em <i>Corpus</i> .....	24
1.8.3 Tradução Automática Neural .....	25
2. A Pós-Edição.....	28
2.1 Definição de Edição, Pré-Edição, Pós-Edição e Revisão .....	29
2.2 Produtividade e Esforço Cognitivo.....	33
2.3 O Ensino da Pós-Edição .....	35
2.4 Garantia da Qualidade da Tradução Automática .....	39
3. O Estágio Profissional.....	47
3.1 A Entidade de Acolhimento.....	47
3.2 As Tarefas Desenvolvidas no Estágio Profissional .....	48
3.2.1 Tradução.....	49
3.2.2 Glossário.....	58
3.2.3 Memória de Tradução .....	59
3.2.4 Formatação .....	60
3.2.5 Digitação .....	62
3.2.6 Pós-Edição.....	62
3.2.7 Legendagem .....	64

4. Reflexão crítica .....	66
Conclusão .....	69
Anexo 1 – Norma ISO 18587:2017 - Serviços de tradução - Pós-edição .....	76
Anexo 2 - Categorias de erros de pós-edição – TAUS (DQF-MQM).....	86

### **Índice de Figuras**

Figura 1 – Mapa Conceitual de Holmes–Toury sobre estudos de tradução .....	4
Figura 2 – Diagrama de Venn – Tradução Automática Contemporânea .....	17
Figura 3 – Página de criação das MTs por pares linguísticos no memoQ .....	60
Figura 4 – Exemplo da planilha encaminhada para a realização da formatação.....	61
Figura 5 – Ecrã de designação de revisor no memoQ .....	63
Figura 6 – Exemplo de relatório de severidade TAUS emitido pelo memoQ.....	64

### **Índice de Tabelas**

Tabela 1 - Tabela comparativa das abordagens da TA ao longo do tempo.....	27
Tabela 2 - Categorias de Erros do Modelo LISA QA .....	43
Tabela 3 - Categorias de Erros do Modelo SAE J2450.....	43
Tabela 4 – Categorias de Erros do modelo TAUS DQF-MQM.....	44
Tabela 5 – Níveis de Severidade do modelo TAUS DQF-MQM .....	45
Tabela 6 – As diretrizes de pós-edição do TAUS .....	45
Tabela 7 – Atividades desenvolvidas no estágio profissional .....	49

## **Siglas e Abreviaturas**

ALPAC – *Automatic Language Processing Advisory Committee*

ALPS – *Automated Language Processing Systems*

ATA – *American Translators Association*

BDT – Bases de Dados Terminológicas

CEO - *Chief Executive Officer*

DQF – *Dinamic Quality Framework*

EN>PT – Inglês com tradução para o Português

ENIAC – *Electronic Numerical Integrator and Computer*

ES>PT – Espanhol com tradução para o Português

ETOC – *Easy to Consult*

EUATC – *European Union Association of Translation Companies*

IA – Inteligência Artificial

LC – Língua de chegada

LISA – *Localization Industry Standards Association*

LP – Língua de partida

LQA – *Linguistic Quality Assurance*

MIT – *Massachusetts Institute of Technology*

MQM – *Multidimensional Quality Metrics*

MT – Memória de Tradução

PE – Pós-edição

PLN – Processamento de Linguagem Natural

PT-BR – Língua Portuguesa com a variante brasileira

PST – Prestador de Serviços de Tradução

QA – *Quality Assurance*

SAE – *Society of Automobile Engineers*

SRX – *Segmentation Rules eXchange*

TA – Tradução Automática

TAC – Tradução Assistida por Computador

TAN – Tradução Automática Neural

TAUS – *Translation Automation User Society*

TBX – *Term-Base eXchange*

TR – Teoria da Relevância

TMX – *Translation Memory eXchange*

XML – *Extensible Markup Language*

XLIFF – *XML Localization Interchange File Format*

## Introdução

Qual não foi minha surpresa quando, no primeiro dia do curso de graduação (Bacharelado em Tradução e Interpretação), a professora da disciplina de teoria da tradução confrontou-nos com a seguinte máxima: *Traduttore, traditore*. Tal expressão tem o intuito de instigar o pensamento crítico quanto à atuação do tradutor e fazer-nos questionar sobre os parâmetros e possíveis limites na realização deste ofício.

Na alegoria do Mito da Caverna de Platão, Sócrates e Glauco dialogam a respeito de uma caverna onde alguns homens foram aprisionados desde a infância, amarrados pelas pernas e pescoços, de maneira que estavam resignados a ficar imóveis e a olhar apenas para frente. Atrás deles, arderia, a certa distância e fora da caverna, um fogo que projetava na parede à frente deles, as sombras de tudo o que acontecia no ambiente externo e, desta forma, tomavam as sombras por realidade, não tendo consciência de sua ignorância. Sócrates afirma que, se um deles fosse liberto e conduzido para fora da caverna, em princípio, continuaria a acreditar que as sombras que via antes seriam mais verdadeiras do que os objetos reais que passaria a enxergar diretamente. Mas que, após um período de adaptação, reconheceria finalmente a verdade, e sentiria piedade de seus companheiros de cárcere. No entanto, se tivesse a oportunidade de retorno à caverna, veria com mais dificuldade que seus companheiros, por ter voltado ao mundo das sombras após ter sido apresentado à luz. E pior, ao tentar convencer seus companheiros de que durante toda a vida teriam visto apenas sombras, seria chamado de louco, pois seria sumariamente desacreditado.

Essa alegoria é útil para que entendamos o papel do tradutor que, de certa forma atua como aquele prisioneiro que foi liberto e é tido como traidor por apresentar aos companheiros monolíngues, o que só ele viu ao ter contato com uma língua estrangeira. Portanto, a máxima *Traduttore, tradittore* nos leva à reflexão de quem, quando e como o tradutor irá trair. Trairá o autor, o leitor ou a si mesmo? Quais trechos e contextos pedem uma traição? Quais escolhas e qual discurso escolherá? Quais princípios prevalecerão?

Jakobson (1976), em seu ensaio sobre os aspectos linguísticos da tradução, classifica a tradução em três espécies: tradução intralingual ou reformulação, tradução interlingual e tradução intersemiótica. Traduzir exige reflexão e prudência e a escolha das palavras nunca é aleatória. Para além do conhecimento linguístico, traduzir não é somente a simples troca de uma palavra da língua A para outra da língua B, não é uma conta exata, é um processo de recodificação, “onde o tradutor recodifica e transmite a mensagem de outra fonte” (p. 65). O foco desse trabalho é a tradução interlingual e, em específico, o uso de ferramentas computacionais, nomeadamente, a tradução automática como geradora de material traduzido com o objetivo com vistas à realização da pós-edição.

As ferramentas tecnológicas de auxílio ao tradutor já não são novidade, especialmente na era digital na qual estamos inseridos. Todo o indivíduo, que assim o desejar, pode acessar qualquer tipo de conteúdo à distância de um clique, tanto por computadores pessoais, como pela tela de seus *smartphones*. Dicionários, gramáticas e enciclopédias impressas acabaram por migrar para o formato digital para a difusão por meio da internet devido à facilidade no acesso e à rapidez com que as atualizações podem ser disponibilizadas. A rede de comunicação mundial hoje transformou o mundo no que chamamos de aldeia global, através da qual as barreiras físicas não impedem a interação entre povos e culturas e, acima de tudo, com o crescimento do comércio virtual, a necessidade de disponibilização de todo e qualquer material traduzido, tornou-se imprescindível.

Este trabalho parte da análise do contexto histórico que permeia desde o início da ideia até a implementação do uso da tecnologia informática como ferramenta de auxílio ao tradutor e desenvolve o tema central pautado na evolução da tradução automática que, mesmo após avanços importantes, necessita da intervenção do tradutor para conferir o nível de qualidade humana.

Na sequência, apresenta o desenvolvimento do estágio curricular inserido no âmbito do curso de Mestrado em tradução, o qual foi realizado na empresa Quicksilver Translate, localizada em Sant Cugat del Vallès, Barcelona, Espanha.

A escolha pela realização do estágio profissional veio ao encontro do desejo de observar e participar do funcionamento de uma empresa de tradução, o qual permitiu que pudesse experimentar a realização de outras atividades que não somente a tradução. A somatória do conjunto de atividades desenvolvidas, as quais serão apresentadas

detalhadamente no capítulo sobre o estágio curricular (capítulo 3), juntamente com a observação do trabalho dos gestores, permitiu que os objetivos iniciais fossem plenamente alcançados e superou as expectativas no sentido de promover a prática profissional e o aprendizado dessa área até então desconhecida por mim, a pós-edição, e de entender a sua importância nos estudos da tradução.

Por fim, apresenta uma reflexão crítica que permite demonstrar que o profissional da tradução necessita de atualização contínua não só com relação às questões linguísticas, por óbvio, mas sobretudo deve buscar o conhecimento especializado consoante às questões tecnológicas envolvidas no processo tradutório moderno.

## 1. Tradução e Tecnologia

Atualmente, é comum relacionarmos o uso de ferramentas tecnológicas ao desenvolvimento do trabalho do tradutor, especialmente, quando as tecnologias informáticas são parte integrante da vida em sociedade e proporcionam um nível de interação e conexão cada vez maiores. Há algumas décadas, a bancada de trabalho do tradutor ostentava outros tipos de ferramentas de apoio, nomeadamente, dicionários e gramáticas físicos e uma boa máquina de escrever. Posteriormente, com o valor reduzido na produção de componentes eletrônicos, a perspectiva de compra de *desktops* pessoais se tornou possível e conduziu, naturalmente, ao desenvolvimento de softwares específicos que têm tornado o trabalho do tradutor mais produtivo e eficiente.

De acordo com o mapa conceitual desenvolvido por Toury em 1995, com base no artigo “The Name and Nature of Translation Studies” de James S. Holmes, publicado em 1972, a tradução automática (TA) situa-se dentro do campo de estudos aplicados da tradução na área de assistência à tradução, que vai desde o uso de dicionários, glossários e gramáticas até ao desenvolvimento dos recursos de auxílio à tradução, tais como ferramentas de tradução assistida por computador (TAC) na forma de softwares, memórias de tradução (MT), tradução aplicada ao ensino de línguas e também o uso da linguística de *corpus* (Pym, 2016).

A figura seguinte ilustra o mapa conceitual dos estudos da tradução:

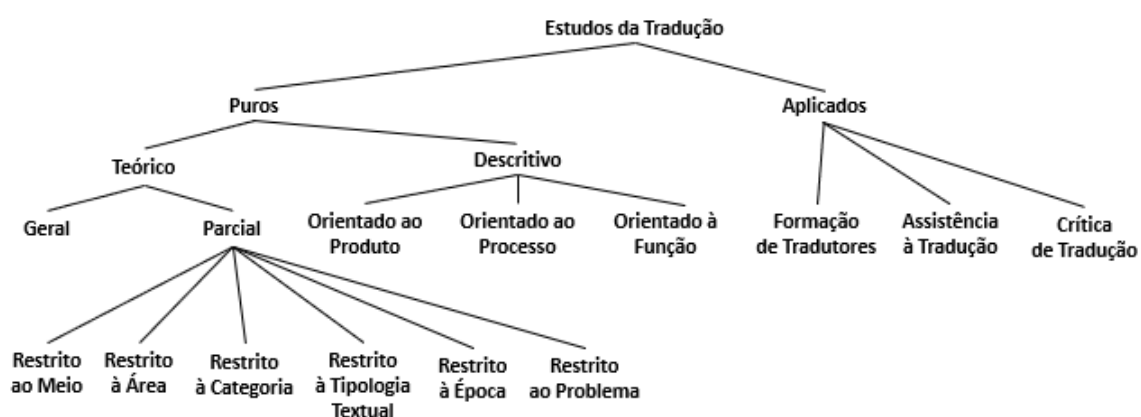


Figura 1 – Mapa Conceitual de Holmes–Toury sobre estudos de tradução

Fonte: Gideon Toury, *Descriptive Translation Studies – And Beyond* (1995), figura elaborada pela autora deste relatório.

Primeiramente, faz-se necessária a distinção do que de fato é TA, tradução mecânica ou *Machine Translation*, pois pode ser facilmente confundida com a TAC, mas referem-se a ferramentas distintas, conforme segue:

In MT, translation is performed by the computer, with no human intervention in the process (although, as will be seen, there may be need for human intervention before or after it); in computer-aided translation, translation is performed by a professional with the aid of a range of translation tools to help them (Forcada, 2010, p. 215).

A TA é uma ferramenta que pode ser utilizada em uma das fases do processo de tradução, através da qual a tradução é realizada totalmente pelo computador por meio da análise e interpretação dos enunciados inseridos na ferramenta em língua natural e na posterior transformação em texto-alvo com eventual necessidade de pré ou pós-edição. A TAC, por sua vez, é uma ferramenta que permite que os usuários possam armazenar textos-fonte previamente traduzidos, os quais são alinhados em segmentos com os respectivos textos-alvo que resultam na criação de MT e bases de dados terminológicas (BDT) que venham a facilitar a realização de traduções futuras (Bowker & Fisher, 2010).

Antes mesmo do desenvolvimento de um mecanismo eletrônico de TA, Hutchins e Lovtskii (2000) apontam para o que podemos chamar de primeira tentativa para o desenvolvimento de um dispositivo de tradução. No ano de 1933, o cientista russo Petr Troyanskii criou uma máquina para a seleção e impressão de palavras durante o processo de tradução de uma língua para a outra (Hutchins e Lovtskii, 2000). A máquina era manipulada por um operador que selecionava uma palavra do texto, encontrava uma carta correspondente, tirava uma fotografia e datilografava as características morfológicas em uma máquina de escrever, como uma espécie de dicionário mecanizado. O cartão e o filme fotográfico eram unidos de maneira a serem utilizados simultaneamente, criando um conjunto da palavra e sua morfologia para consulta futura (Hutchins & Lovtskii, 2000).

A história da TA com o uso de dispositivo eletrônico tem seu início logo após a invenção do primeiro computador (ENIAC - *Electronic Numerical Integrator and Computer*), no ano de 1946. A partir desse momento, a evolução e desenvolvimento dos estudos para a aplicação da TA são divididos em quatro períodos, segundo Sin-wai, 2015. O primeiro período é conhecido como germinação e vai do ano de 1967 a 1983. O segundo período é compreendido entre os anos de 1984 e 1993 e é nomeado como o

período de crescimento constante. O terceiro período vai de 1993 a 2003 e é conhecido como a década de rápido crescimento. O quarto período é compreendido entre os anos de 2004 e 2013, no qual o crescimento global recebe destaque.

A partir do ano de 2014 podemos classificar a TA como pertencente à era digital, período em que a tecnologia tem papel fundamental dentro das organizações, as quais substituem determinadas tarefas, antes realizadas de forma manual ou com total controle humano, por tecnologias que automatizam processos e melhoram seu desempenho.

### **1.1 Período de Germinação (1967-1983)**

Os anos que antecedem o período de germinação são marcados, primeiramente, pela atuação do matemático Warren Weaver, diretor da Divisão de Ciências Naturais da Fundação Rockefeller, o qual foi desafiado a desenvolver um dispositivo para tradução que fosse semelhante às calculadoras numéricas, pois era especialista no desenvolvimento de dispositivos eletrônicos para a realização de cálculos matemáticos. Em março de 1947, Weaver contactou Norbert Wiener, professor do Instituto de Tecnologia de Massachusetts (MIT), linguista e especialista em cibernética, para apresentar a proposta da criação do dispositivo de tradução mecânica (Hutchins, 1986). O sistema proposto tinha como base o uso de criptografia, processamento automático das línguas naturais, a qual teve seu uso inicial no período das grandes guerras, através da codificação de mensagens visando impedir o acesso dos exércitos inimigos a informações estratégicas. Na mesma altura, Weaver foi apresentado ao Dr. Andrew Booth, cientista britânico com interesse na investigação dos computadores eletrônicos recém-inventados. Booth conhecia Alan Turing, especialista em computadores do Laboratório Nacional de Física da Grã-Bretanha e entusiasta dos estudos para uso de computadores para fins de tradução. A partir do contato com Weaver, Booth regressa a seu país com a ideia da criação de um dicionário mecanizado e então conhece Richard H. Ritchens, diretor assistente do *Commonwealth Bureau of Plant Breeding and Genetics* em Cambridge, o qual havia idealizado e realizado testes para a tradução de títulos e resumos de artigos científicos e, para tal, realizou testes com um equipamento de cartões perfurados que, em princípio, contemplava a tradução palavra por palavra (Hutchins, 1986).

Em 1949, Weaver escreveu um memorando que foi enviado a cerca de 30 individualidades, o qual foi intitulado *Translation* e que visava instigar o início da pesquisa do uso de computadores para a tradução de documentos. Segundo Hutchins (1986), o documento elaborado por Weaver apresentou quatro propostas de apoio à elaboração do dispositivo. A primeira reside na questão dos múltiplos significados e da ambiguidade. A segunda proposta parte do pressuposto de que há elementos lógicos na linguagem e que, por este motivo, um computador poderia ser programado de modo a prever todas as variáveis possíveis baseado na estrutura neural do cérebro humano. A terceira retoma a aplicabilidade dos métodos criptográficos e baseia-se na teoria da informação de Claude Shannon (1948), a qual ocupa-se das propriedades estatísticas básicas da comunicação e estuda a quantificação, armazenamento e comunicação da informação. A quarta proposta, baseia-se na crença de que, assim como pode haver características lógicas comuns em todas as línguas, também pode haver universais linguísticos, características universais da linguagem em geral (Hutchins, 1986).

A primeira demonstração pública do funcionamento de uma máquina de tradução aconteceu em janeiro de 1954 em Nova York, por meio da parceria entre a IBM e a Universidade de Georgetown (Austermühl, 2001). O equipamento, chamado de *701 Translator*, era um computador de grandes dimensões que transformou cerca de 60 frases do idioma russo para o inglês em poucos segundos. O experimento foi realizado a partir de textos da área química e a escolha pelo idioma russo não foi aleatória, pois deveu-se ao facto de que o número de estudantes era ínfimo em comparação ao grande volume de material textual russo produzido na altura, especialmente no que tange à ciência e tecnologia. Este feito estimulou o investimento na pesquisa e desenvolvimento de dispositivos de tradução eletrônica, não só nos EUA, mas em outros países (Austermühl, 2001). Após esse marco relevante, o governo norte americano investe na pesquisa para o aprimoramento do dispositivo e, num primeiro momento conta com o comando do filósofo, matemático e linguista israelita Yehoshua Bar-Hillel, que visita centros de TA por aproximadamente dois anos para observar o que se estava a fazer ao redor do mundo e, no início da década de 1960, escreve um relatório, cujo principal ponto reside na ausência de métodos para a resolução do problema da ambiguidade e que a TA estava longe de atingir a qualidade de um ser humano (Hutchins, 1986).

No esforço da busca por uma solução para os temas abordados por Bar-Hillel e seus antecessores e, na expectativa de liderar a corrida pelo desenvolvimento do melhor dispositivo de tradução automática, o governo norte-americano cria no ano de 1964 o ALPAC (*Automatic Language Processing Advisory Committee*), um comitê multidisciplinar composto por sete membros, sendo eles os linguistas Eric P. Hamp (Universidade de Chicago) e Charles F. Hockett (Universidade de Cornell), os investigadores de tradução automática David G. Hays (RAND Corporation) e Anthony G. Oettinger (Universidade de Harvard), o psicólogo John B. Carroll (Universidade de Harvard) e seu presidente, o engenheiro John R. Pierce (Laboratórios Bell Telephone); além desses, contavam com a colaboração de investigadores ativos em TA, tais como, Paul Garvin e Jules Mersel (Bunker-Ramo Corporation), Gilbert King (Itek Corporation e anteriormente IBM) e Winfred P. Lehmann (Universidade do Texas) (Hutchins, 1996).

Inicialmente, o comitê concordou que o apoio à pesquisa em TA poderia ser justificado em duas bases: (1) pesquisa em um campo intelectualmente desafiador que fosse amplamente relevante para a missão da agência de apoio e (2) pesquisa e desenvolvimento com uma clara promessa de efetuar reduções de custo antecipadas, ou melhorar substancialmente o desempenho, ou reunir uma necessidade operacional. Um ponto importante a ser observado é que o relatório se concentrou nas necessidades militares do governo norte americano na análise e digitalização de documentos em língua russa, não estava preocupado com outros usos potenciais ou usuários de sistema de TA ou com qualquer outro par linguístico (Hutchins, 1996).

O relatório destacou a importância dos estudos para o avanço no desenvolvimento de softwares informáticos e considerou necessária a investigação básica do desenvolvimento de métodos informáticos para lidar com a linguagem como ferramenta de apoio ao linguista de modo que possa descobrir e afirmar generalizações e tipos complexos de teorias. Por fim, reforçou o potencial de melhora da qualidade da tradução no caso do uso de ferramentas de auxílio ao tradutor, como a TAC, para tornar a tradução mais rápida e mais barata (Hutchins, 1996).

Após a divulgação do relatório ALPAC os avanços na área sofreram um sério golpe que fez com que os estudos de TA permanecessem estagnados por mais de uma década, não só nos Estados Unidos, como também na Europa e Rússia. Mas esse foi o momento importante para a ascensão da TAC e mostrou o quanto TA e TAC estão intimamente

ligadas. A TAC, no entanto, não seria possível sem o apoio de conceitos e softwares relacionados, além do desenvolvimento das MT, termo este que foi grandemente difundido entre os anos 1960 e 1980, conforme aponta Hutchins (1996).

Em 1978, Alan Melby, do Grupo de Investigação de Tradução da Universidade de Brigham Young realizou uma pesquisa sobre TA e desenvolveu um sistema interativo de tradução, ALPS (*Automated Language Processing Systems*), que já incorporava o conceito de MT. Seguindo a mesma linha de estudos, Peter Arthern propôs no ano seguinte, numa conferência realizada pela Comissão Europeia, o método de tradução por recuperação de texto (Sin-wai, 2015).

Em 1980, Martin Kay publicou um artigo intitulado “The Proper Place of Men and Machines in Language Translation”, por meio do qual propôs a criação de um sistema de TA em que a exibição do ecrã é dividida em duas janelas, onde o texto a traduzir aparece na janela superior e a tradução seria realizada na janela inferior. Além disso, era possível editar o texto e seleccionar palavras para consulta ao dicionário. Em vista do nível de capacidade de processamento de texto naquela altura, sua proposta foi inspiradora e exerceu grande impacto no desenvolvimento da TAC nos anos seguintes (Sin-wai, 2015). Por essa razão, Kay é considerado o precursor de sistemas de tradução interativa. É evidente que a implantação da MT ao sistema de TA foi a contribuição mais importante para o crescimento da TAC nos sistemas de primeira geração, porém a deficiência na geração de frases com correspondência total minimizava a reutilização das MT, o que não favorecia a agilidade do processo (Sin-wai, 2015). Sendo assim, por volta de 1980, de acordo com Sin-wai (2015), alguns pesquisadores começaram a recolher e armazenar amostras de tradução com o intuito de partilhar seus recursos de tradução. Restringidos pelo espaço limitado no hardware, pelo custo elevado na construção de uma base de dados bilíngue, pela ausência de estudos mais aprofundados sobre algoritmos de alinhamento de dados bilíngues, a tecnologia da MT chegou ao limite. Como resultado, não surgiu nenhum sistema TAC que pudesse ser comercializado e a tecnologia desenvolvida até então não causou o impacto esperado na indústria da tradução (Sin-wai, 2015).

## 1.2 Período de Crescimento Constante (1984-1992)

O início desse período foi marcado pela atuação das primeiras companhias desenvolvedoras de sistemas comerciais de TAC: Trados, na Alemanha e STAR Group, na Suíça (Sin-wai, 2015). Mais tarde, essas duas empresas tiveram um impacto marcante no desenvolvimento da TAC em nível mundial. O nome Trados vem de TRAnslation and DOcumentation Software e foi criada inicialmente como fornecedora de serviços linguísticos para o desenvolvimento de um projeto de tradução solicitado pela IBM. No processo de atendimento da demanda de traduções da IBM foi desenvolvido o software Trados GmbH, que é considerado como o ponto de partida do período de crescimento constante da TAC (Garcia & Stevenson, 2005). De igual importância foi a fundação da empresa STAR Group, cujo nome é um acrônimo formado por Software, Translation, Artwork e Recording. Em 1986, abriu seu primeiro escritório na Alemanha e passou a atuar mais ativamente no mercado da localização de softwares. Na sequência desenvolveu softwares para a gestão de informação e memórias de tradução. Com o crescimento do mercado de exportação foi levada a abrir filiais no Japão e na China e ainda hoje desempenha papel importante na indústria da tecnologia da tradução (Sin-wai, 2015).

Em 1988, teve início a comercialização de sistemas TAC, quando a filial japonesa da IBM lançou a ferramenta ETOC (*Easy to Consult*), que era um dicionário eletrônico atualizado que oferecia a consulta palavra por palavra. Não era possível realizar consultas por frases ou sentenças, porém ao introduzir uma frase o sistema tentava extrai-la do seu dicionário. Embora o sistema não utilizasse o termo memória de tradução e a base de dados continuasse a ser chamada de dicionário, possuía, essencialmente, as características básicas da MT que conhecemos atualmente, de acordo com Sumita & Tsutsumi (1988), citados por Sin-wai (2015).

Em 1988, Trados desenvolveu o editor para processador de textos TED e nesse mesmo período a empresa foi dividida de modo a ficar responsável somente pelo desenvolvimento de softwares de tradução. A outra parte, responsável pelos serviços de tradução, foi adquirida pela empresa INK na Holanda. Após dois anos, Trados lançou sua primeira ferramenta de gestão terminológica multilíngue para o sistema operacional DOS que permitia a opção de personalização da ferramenta pelo utilizador (Wassmer, 2011).

Em 1991, a empresa STAR Group lançou a ferramenta Transit, a qual já utilizava internamente desde 1987. Também baseada no sistema operacional DOS, a ferramenta já apresentava módulos semelhantes aos utilizados nas ferramentas TAC atuais, tais como gestão terminológica, MT e gestão de projetos, entre outros. Desde então foram desenvolvidos produtos adicionais para a implementação e automatização, tais como o TermStar, WebTerm, GRIPS, MindReader, Spider e Star James (<http://www.star-group.net>). Ainda em 1991, foi fundada a empresa PROMT, na Rússia, cujo foco inicial foi a tecnologia de TA e mais tarde começou a fornecer uma gama completa de soluções de tradução: sistemas de serviços de TA, dicionários, sistemas de MT e de mineração de dados (<http://www.promt.com>). Porém, um dos eventos de maior destaque na altura foi o lançamento da primeira ferramenta TAC que marcou o início da comercialização de sistemas de tradução, Trados, em 1992 (Sin-wai, 2015).

O final desse período foi marcado pela abertura de duas empresas especializadas na produção de softwares para tradução, ambas no Reino Unido: a SDL International (<http://www.sdl.com>), como fornecedora de serviços para a globalização e a TA Software Technology Ltd, especializada na produção de softwares de tradução com foco na língua árabe, sendo responsável, posteriormente, pelo desenvolvimento de uma série de produtos de tradução automática (árabe e inglês), sistema híbrido MT e TM, Xpro7 e motor de tradução online (<http://www.atasoft.com>) (Sin-wai, 2015).

### **1.3 Período de Rápido Crescimento (1993-2003)**

Este período é reconhecido como período de rápido crescimento devido ao lançamento de mais sistemas comerciais, ao desenvolvimento de mais funções integradas, ao domínio dos grandes sistemas tendo o Windows como base operacional, ao suporte de mais formatos de documentos, ao suporte de mais línguas para tradução, e ao domínio do Trados como líder de mercado. Até então, havia apenas três sistemas disponíveis no mercado, o Translator's Workbench II do Trados, o IBM Translation Manager/2, e o Star Transit 1.0. No período seguinte foram desenvolvidos cerca de 20 sistemas para venda, isso significa um aumento seis vezes maior do que no período anterior. Os sistemas mais conhecidos desenvolvidos nesse período são Déjà Vu, Eurolang Optimizer (Brace, 1994),

Wordfisher, SDLX, ForeignDesk, Trans Suite 2000, Yaxin CAT, Wordfast, Across, OmegaT, MultiTrans, Huajian, Heartsome e Transwhiz (Sin-wai, 2015).

As ferramentas TAC do primeiro e segundo períodos foram equipadas com componentes básicos tais como MT, gestão terminológica e editor de tradução. Gradualmente foram integrados mais componentes e mais funções às ferramentas TAC, as quais merecem destaque, as ferramentas de alinhamento de textos, TA e gestão de projetos. A TA foi integrada em sistemas TAC para realizar a tradução de segmentos não encontrados nas MT. Os sistemas TAC desenvolvidos antes de 1993 eram executados no sistema operacional DOS ou OS/2 da IBM, mas em 1993 surgiram versões de sistemas TAC executados a partir do sistema operacional Windows e, desde então, tornaram-se o fluxo dominante. Como exemplos das primeiras ferramentas que migraram do DOS para o Windows, temos o IBM Translator's Workbench e o Trados GmbH (Sin-wai, 2015).

Além da migração do sistema DOS para o Windows, este também foi o período em que as ferramentas TAC foram aperfeiçoadas de modo a aceitarem mais formatos de documentos, por meio da inclusão de ferramentas com o suporte para ficheiros de forma direta ou por meio do uso de filtros, como o Adobe InDesign, o HTML, Microsoft PowerPoint, Excel, Word, entre outros. Uma das ferramentas que passou a suportar todos os formatos de ficheiros amplamente utilizados pela comunidade de tradução, foi o Trados 6.5, por exemplo. Outro facto relevante, foi o aumento na quantidade de línguas suportadas pelas ferramentas, muito embora a MT seja, de certa forma, independente das línguas traduzidas. Em 1992, a ferramenta Translator Workbench Editor suportava somente cinco línguas europeias: alemão, inglês, francês, italiano e espanhol, porém, na mesma altura, a ferramenta Translator's Manager da IBM já suportava 19 línguas, incluindo chinês, coreano e outros conjuntos de códigos de caracteres compatíveis com o sistema operacional OS/2 (Sin-wai, 2015).

Muito do que diz respeito à evolução dos processos de armazenamento, intercâmbio de dados e processamento em qualquer língua deve-se à implementação do uso do Unicode (padrão que permite aos computadores a representação e manipulação, de forma consistente, de textos de qualquer sistema de escrita existente – <http://home.unicode.org>), o que permitiu aos programadores de sistemas TAC a resolução gradual dos obstáculos que surgiram no processamento de linguagem. Após o lançamento do Microsoft Windows 2000, o número de ferramentas TAC com o apoio do Unicode cresceu visivelmente, com

o lançamento do Transit 3.0 em 1999, MultiTerm e WordFisher 4.2.0 em 2000, WordFast Classic em 3.34 em 2001, entre outros (Sin-wai, 2015). O fim desse período foi marcado pelo domínio da ferramenta Trados como líder de mercado e isso ocorreu, especialmente, pelo investimento na construção de uma base tecnológica sólida por meio do desenvolvimento de uma gama de softwares de tradução que foram agregados à ferramenta, pela consolidação do uso das MT, pela ampla gama de ficheiros suportados pela ferramenta e pela possibilidade de outras ferramentas processarem ficheiros gerados pela ferramenta Trados (Sin-wai, 2015).

#### **1.4 Período de Desenvolvimento Global (2004-2013)**

O último período evidenciou o lançamento de novas ferramentas TAC e a implementação de melhorias por meio de avanços tecnológicos no desenvolvimento dessas ferramentas. A ampliação na oferta e a disponibilidade de pacotes personalizados facilitou o acesso por conta dos preços mais competitivos. Com o crescimento da indústria da tradução, o fluxo natural foi o aumento da quantidade de empresas desenvolvedoras de ferramentas TAC no mercado da tradução no período de desenvolvimento global (Sin-wai, 2015).

Um dos primeiros países a despontar nesse período foi a Hungria, quando em 2004 criou a empresa Kilgray Translation Technologies, a qual no ano de 2005 lançou a primeira versão da ferramenta memoQ, sendo que a principal característica dessa versão era a possibilidade de criação de projetos de servidor (<http://memoq.com>).

Outro exemplo foi a criação da Rozetta Corporation, no Japão. A empresa desenvolveu a ferramenta TraTool, um sistema de tradução computadorizada com MT, ferramenta de alinhamento integrada, base de dados terminológicos e um dicionário. Embora tenha sido disponibilizado comercialmente, não apresentou melhorias consideráveis desde sua primeira versão (<http://www.tratool.com>).

No Reino Unido, a ATA lançou um sistema de tradução com MT para a língua árabe, o Xpro7, que possuía a função de TA (<http://www.atasoft.com>). A SDL International, lançou a ferramenta SDLX 2004, com características inovadoras para a época, tais como: integração com sistemas empresariais por meio de ficheiros em formato TMX (*Translation Memory Exchange* - criados especificamente para o intercâmbio de

informações entre ferramentas TAC), gestão de terminologia online, gestão de fluxo de trabalho multilíngue, adaptação de novos formatos de ficheiros e tradução baseada em conhecimento (<http://www.sdl.com>).

Nos Estados Unidos da América, a empresa Systran lançou a ferramenta Systran Professional Premium 5.0, que disponibilizava MT integrada com suporte de TMX, ferramenta de pós-edição e análise de qualidade (<http://www.systransoft.com>).

O ano de 2006 foi significativo para os avanços na área da tecnologia voltada à tradução em todo o mundo, com destaque especial para a Europa, Reino Unido e Estados Unidos. Na Alemanha, a empresa Across Systems GmbH lançou a ferramenta Corporate Translation Management 3.5, que marcou o lançamento mundial do software Across (<https://www.across.net/>). No Reino Unido, a SDL International lançou o SDL Trados 2006, cuja principal inovação foi o suporte para Quark, InDesign CS2 e Java (<http://www.sdl.com>).

Nos Estados Unidos, a Apple lançou a Apple Trans, um editor de textos concebido para tradutores, cuja principal inovação era o uso de corpora online, além disso, ajudava os utilizadores a localizar páginas de websites (<http://developer.apple.com>).

Ainda nos Estados Unidos, a empresa Lingotek, criadora de motores de busca linguística, lançou a versão beta de um serviço colaborativo visando melhorar a eficiência dos tradutores de maneira significativa. A ferramenta funcionava através do uso do motor de pesquisa linguística por meio de uma base de dados de conteúdo multilíngue contendo grande volume de traduções. Dessa forma, os utilizadores poderiam aceder à base de dados para encontrar combinações de palavras mais específicas visando a reutilização. E essa pesquisa baseada no significado das palavras ajudava a manter o melhor estilo, tom e terminologia. Além disso, o motor de pesquisa da Lingotek suportava ficheiros para diversas extensões de documentos (PDF, HTML, DOC, por exemplo), permitindo aos utilizadores o carregamento de seus ficheiros diretamente na ferramenta, além do suporte à ficheiros de MT compatíveis com a extensão TMX (<http://www.lingotek.com>).

Certamente, foi um período importante para o desenvolvimento e expansão das ferramentas de tradução em nível mundial. Sin-wai (2015) aponta o que podemos destacar como os avanços mais importantes neste período:

- 1) A compatibilidade das ferramentas com o sistema operacional Windows e com o Microsoft Office. Na altura, das 67 ferramentas TAC disponíveis no mercado,

somente uma não era compatível com o Windows. Isso demonstra que as ferramentas foram desenvolvidas buscando a consonância com os avanços tecnológicos disponíveis no mercado da informática. O Wordfast 5.51j, por exemplo, foi lançado em abril de 2007, três meses após o lançamento do Windows Vista, e WordFast 5.90v foi lançado em julho de 2010 justamente na sequência do lançamento do Microsoft Word 2007 e 2010.

- 2) A integração do controle do fluxo de trabalho em sistemas TAC foi um passo importante, pois além da reciclagem ou reutilização de traduções de textos previamente traduzidos e da gestão terminológica, foram acrescentadas as funções gestão de projetos, verificação ortográfica, garantia de qualidade e controle de conteúdo. Como exemplo dessa inovação podemos citar o SDL Trados Studio 2011 ao qual foi agregada a função de verificação ortográfica para o maior número de línguas e a ferramenta PerfectMatch 2.0 para acompanhar as alterações realizadas em documentos originais. A maioria dos sistemas disponíveis na altura também realizava o trabalho de verificação de correspondência com segmentos circundantes idênticos encontrados no documento ou na MT.
- 3) A disponibilidade de sistemas em rede ou online também foi um grande marco deste período. O rápido desenvolvimento de novas tecnologias da informação permitiu que os sistemas pudessem ser armazenados em servidores, na web ou mesmo em nuvem, que possui maior capacidade de armazenamento de dados. Como exemplos dessa tecnologia podemos destacar as ferramentas Lingotek Collaborative Translation Platform, SDL World Server, and XTM Cloud.
- 4) A adoção de novos formatos de ficheiros específicos para a indústria da tradução favoreceu o intercâmbio de dados entre diferentes sistemas TAC. Essa mudança aconteceu através da atuação da *Localization Industry Standards Association* (LISA), que desempenhou um papel significativo no desenvolvimento e promoção de normas de intercâmbio de dados, tais como o SRX (*Segmentation Rules eXchange*), TMX (*Translation Memory eXchange*), TBX (*Term-Base eXchange*) e XLIFF (XML Formato de ficheiro de intercâmbio de localização) (<http://en.wikipedia.org/wiki/XLIFF>).

Por fim, Sin-wai (2015) afirma que o crescimento e o rápido desenvolvimento da tecnologia voltada à tradução são notáveis, especialmente, a partir da década de 1980 com o surgimento dos sistemas TAC. Neste momento temos à disposição mais de 100 sistemas desenvolvidos em diferentes países, incluindo China, Japão e Índia, e países da América do Norte, como Canadá e Estados Unidos. Hoje, o uso de ferramentas TAC é uma realidade para mais de 200 000 tradutores na Europa e mais de 6 000 grandes empresas no mundo (Sin-wai, 2015). Segundo pesquisa realizada pelo próprio autor (Sin-wai, 2015), pode-se observar que, cerca de três sistemas TAC foram produzidos todos os anos entre 1984 e 2012. Também foi constatado que, no mesmo período, quase metade dos sistemas de tradução foram desenvolvidos na Europa, o que representa 49,38%, enquanto 27,16% deles foram produzidos na América. O levantamento dessas informações ressalta a perspectiva positiva do mercado da tradução e aposta no desenvolvimento crescente de mais e melhores ferramentas de apoio ao tradutor.

### **1.5 Tradução Automática na Era da Inteligência Artificial**

Após o período de desenvolvimento da TA teve início o fenômeno da convergência tecnológica, a qual representa a viabilidade de desenvolvimento e comercialização em grande escala de soluções de tecnologia, que alterou de maneira significativa a forma como interagimos com os meios de comunicação e uns com os outros. Nesse novo modelo de interação cultural, os meios de comunicação não são apenas “extensões do homem” como afirmou McLuhan (1964), mas permitem com que a comunicação eletrônica se estenda a todos os âmbitos da vida, seja na rotina diária da casa para o trabalho, das escolas para os hospitais e até do entretenimento às viagens, conforme pontua o sociólogo Manuel Castells (2000, p. 394) (Bielsa, 2022).

Nesse contexto, a tradução automática contemporânea é mencionada frequentemente em conjunto com uma série de outros conceitos relacionados, incluindo inteligência artificial, aprendizagem de máquinas, redes neurais artificiais e aprendizagem profunda, alguns dos quais podem ser difíceis de diferenciar para os leigos no assunto. Fontes como Goodfellow et al. (2016) usam um diagrama Venn para explicar como eles se relacionam entre si, conforme podemos visualizar na imagem abaixo:

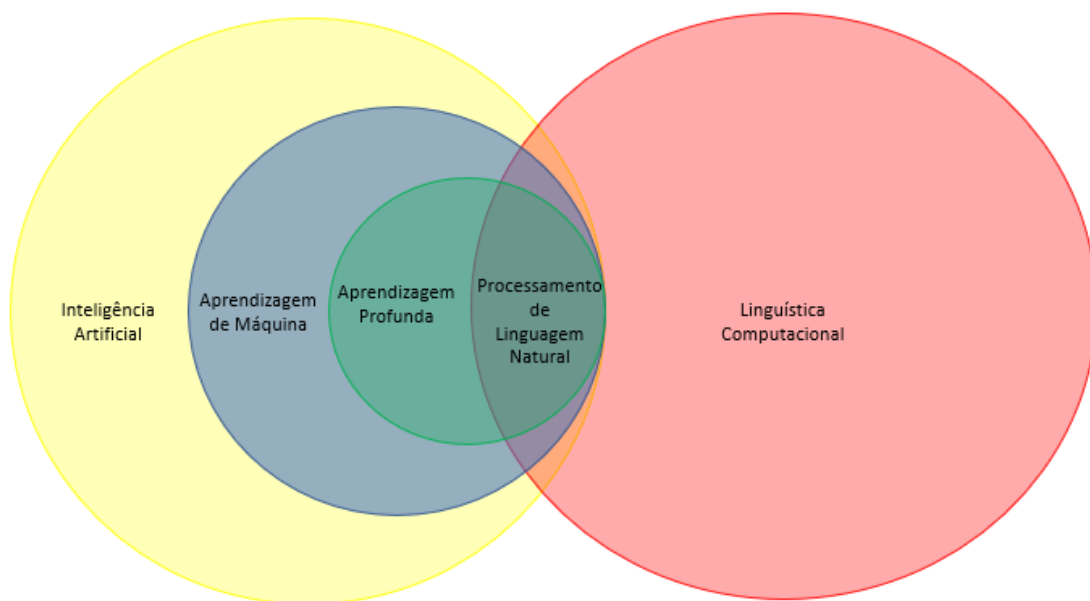


Figura 2 – Diagrama de Venn sobre Tradução Automática Contemporânea  
 Fonte: Deep Learning (Goodfellow et al., 2016)  
 Elaborado pela autora deste relatório.

A IA é a categoria mais geral e, segundo Kenny (2022), é definida como o ramo da ciência da computação que visa criar máquinas, ou mais especificamente, programas de computador que podem resolver problemas do tipo que normalmente exigiriam inteligência humana. As máquinas em questão não têm necessariamente que pensar como os humanos, mas são desenvolvidas para agir de forma semelhante a um humano e esse tipo de máquina pode ser projetado para resolver problemas definidos de forma bastante restrita, como reconhecer rostos, por exemplo; tais objetivos são o material da IA estreita, também conhecida, como IA fraca (Kenny, 2022). Em linhas gerais, a IA forte tem um algoritmo complexo que a ajuda a atuar de maneira independente em diferentes situações, enquanto todas as ações em IAs fracas são pré-programadas por um humano, portanto a tradução, como praticada por tradutores profissionais, requer o tipo de inteligência a que a IA forte aspira, mas que tal inteligência ainda permanece além da capacidade dos sistemas de tradução automática (Kenny, 2022).

O aprendizado de máquinas se baseia na premissa de que ao invés de dizer a um programa de computador tudo o que precisa saber desde o início, é melhor deixar a máquina adquirir seus próprios conhecimentos e a máquina o faz observando como o problema que se pretende resolver já foi resolvido no passado. Os problemas de tradução

e suas soluções podem ser capturados em nível de segmento nas unidades de tradução armazenadas em memórias de tradução e outros corpora paralelos. Estas unidades de tradução constituem os dados de treinamento dos quais os sistemas de tradução automática contemporâneos aprendem e é por isso que esses sistemas são geralmente categorizados como dados; aprender com os dados é o que distingue o aprendizado por máquina de outros tipos de IA (Kenny, 2022).

A aprendizagem profunda é o subconjunto da aprendizagem por máquina que é baseada em redes neurais artificiais e o processo de aprendizagem é profundo porque a estrutura das redes neurais artificiais consiste em múltiplas entradas, saídas e camadas ocultas, onde cada camada contém unidades que transformam os dados de entrada em informações que a próxima camada pode utilizar para uma determinada tarefa prevista e, dessa forma, uma máquina pode aprender através do seu próprio processamento de dados (Google, 2016).

Em uma rede neural profunda, as camadas externas correspondem às entradas e saídas da rede e são visíveis para o analista. As camadas intermediárias, ou ocultas, são menos abertas, o que confere ao aprendizado profundo uma reputação de opacidade, o que leva alguns comentaristas a usar a palavra "magia" para descrever o funcionamento interno de redes neurais profundas. Essa visão "mística" da TAN é acrescentada a quando grandes empresas de tecnologia relatam seus sucessos na construção de traduções multilíngues modelos, às vezes envolvendo centenas de idiomas, e que podem lidar com tradução entre idiomas para os quais não havia dados de treinamento bilíngues "diretos" (Google, 2016; Kenny, 2022).

O processamento de linguagem natural (PLN) refere-se ao ramo da ciência da computação, especificamente o ramo da IA, preocupado em dar aos computadores a capacidade de entender textos e palavras faladas da mesma forma que os seres humanos. O PLN combina a linguística computacional, a qual é baseada em regras da linguagem humana, com modelos estatísticos, de aprendizagem mecânica e de aprendizagem profunda. Juntas, estas tecnologias permitem aos computadores o processamento da linguagem humana na forma de dados de texto ou voz e 'compreende' seu significado completo, com a intenção e o sentimento do orador ou escritor. O PLN conduz programas de computador que traduzem texto de um idioma para outro, respondem a comandos falados e resumem grandes volumes de texto rapidamente, em tempo real (IBM, 2020).

O Google Translate e o DeepL são exemplos de tecnologia de PLN. A tradução eficaz tem que capturar com precisão o significado e o tom do idioma de entrada e traduzi-lo para texto com o mesmo significado e impacto desejado no idioma de saída. As ferramentas de tradução automática estão fazendo bons progressos em termos de precisão. Uma ótima maneira de testar qualquer ferramenta de tradução automática é traduzir o texto para um idioma e depois voltar para o original. Um exemplo clássico muito citado para o teste das ferramentas de TA é a seguinte frase: “*The spirit is willing but the flesh is weak*” do inglês para o russo, cuja tradução foi “*The vodka is good but the meat is rotten*”. Com os avanços da ferramenta, o resultado atual é “*The spirit desires, but the flesh is weak*”, o que não é perfeito, mas inspira muito mais confiança na tradução de inglês para russo (IBM, 2020).

Além do uso e do desenvolvimento da IA nos estudos voltados ao aprendizado de máquina, tem sido observado o crescimento de pesquisa nas mais diversas áreas, como na área de tomada de decisões e na área de representação do conhecimento.

### **1.6 Desafios e Limitações da Tradução Automática**

É sabido que a tradução crua, produzida por um sistema de tradução automática, é muito diferente do resultado apresentado pela tradução realizada por profissionais. Pois, sendo a tradução um gesto de re(criação) característico do ser humano, não poderia ser reproduzida por nenhuma máquina, ainda que antropomórfica. Porém, entende-se que a combinação do humano com o mecânico pode reduzir significativamente os custos do processo de tradução (Forcada, 2010).

Os desafios encontrados pela tradução automática são classificados em quatro grupos por Arnold (2003):

- 1) A forma não determina completamente o conteúdo, pois um texto pode dar margem a diferentes interpretações, especialmente, pela questão da ambiguidade que acontece quando uma ou mais palavras possuem mais do que uma interpretação (ambiguidade lexical). Isto ocorre quando a frase tem mais de uma estrutura sintática possível, como no exemplo: “Eu vi o João passar pelo banco”.

A palavra banco pode ter duas interpretações possíveis: banco como instituição financeira ou como o local onde podemos nos sentar.

A resolução de um problema de ambiguidade por parte da máquina não é uma tarefa simples. Enquanto o profissional da tradução recorre à observação do contexto, ao conhecimento de mundo, às expectativas e crenças para descartar interpretações errôneas, a máquina recorre apenas a processos programáveis para processar a informação e extrair do texto o elemento ambíguo.

- 2) O conteúdo não determina completamente a forma, pois há diversas maneiras de se expressar em uma língua em contextos específicos.
- 3) Cada língua possui uma estrutura diferente para transmitir a mesma interpretação. Ainda, segundo Arnold (2003), utiliza-se como exemplo a frase “Eu gosto de bicicletas”, cuja tradução para o catalão é “M'agraden les bicicletes”. Podemos observar que o objeto direto na frase em português “bicicletas” foi transformado no tema da frase em catalão “les bicicletes” que requer um verbo plural na 3.<sup>a</sup> pessoa “agraden” e que, o pronome em português “Eu” transformou-se num pronome de objeto proclítico “M”. Embora este exemplo seja relativamente simples, demonstra que as estruturas utilizadas pelas diferentes línguas podem ser tão divergentes que uma tradução simples, palavra-por-palavra, soaria ininteligível ou mesmo errada.
- 4) Além dos problemas citados acima, Arnold (2003) também ressalta o problema de descrição. Não é possível expressar todos os mecanismos relacionados aos processos mentais envolvidos no ato da tradução. A maior parte desses processos são resolvidos com o uso de métodos que, em geral, têm de fazer simplificações radicais ou até mesmo reformulações completas no processo tradutório.

Os problemas apresentados são visivelmente reduzidos se os pares linguísticos estão relacionados por afinidades morfológicas, sintáticas e semânticas que simplificam a concepção destes sistemas e permitem obter traduções de fácil compreensão e correção. Também há textos mais fáceis de traduzir por seu conteúdo e formato, tais como cartas comerciais e textos técnicos, por exemplo; porém há textos para os quais a tradução

automática torna-se completamente inaplicável, como os textos publicitários ou poéticos (Arnold, 2003).

### **1.7 Aplicações da Tradução Automática**

Conforme afirma Forcada (2010), a tradução automática se propõe, basicamente, a dois propósitos: assimilação e disseminação, conforme vemos na sequência.

Na assimilação, recorreremos à tradução automática quando a LP não é de fácil compreensão e queremos ter uma ideia aproximada do conteúdo do texto, sua essência. Neste caso, os erros de tradução não têm tanta relevância se o sistema conseguir transmitir o sentido geral do texto. O nível de detalhe alcançado está diretamente ligado ao sistema utilizado e aos pares linguísticos envolvidos, pois há línguas que dispõem de uma estrutura mais rica devido ao volume de utilizações, o que permite o aperfeiçoamento do sistema ao longo do tempo e a apresentação de um resultado cada vez mais refinado.

Na disseminação, os textos são traduzidos automaticamente como etapa intermediária na produção de um documento a ser publicado. Neste caso, o resultado bruto produzido pela máquina tem que passar pelo processo de pós-edição, ou seja, será revisto e corrigido por um profissional qualificado. A substituição da tradução totalmente realizada por tradutor para uma tradução realizada com o auxílio de máquina tem sido cada vez mais implementada e pode ser feita em diferentes fases do processo:

1. Tradução automática seguida de pós-edição: o profissional edita o texto bruto de forma a torná-lo adequado. Há vantagens nesse processo se o custo da tradução automática somado ao custo do serviço de pós-edição seja inferior ao custo da tradução humana, a considerar custos adicionais em decorrência da formação profissional, do fluxo de trabalho e do volume de traduções.
2. Tradução automática após pré-edição: quando estão envolvidas várias línguas-alvo pode ser vantajoso pré-editar o texto-fonte (Hutchins & Somers, 1992), pois cada edição bem escolhida do texto-fonte pode evitar o elevado volume de pós-edições em mais de uma língua (nota-se que a pós-edição não é descartada por completo). Porém, os pré-editores precisam ser treinados para prever possíveis

problemas que a tradução automática possa vir a apresentar, e isto certamente irá onerar o custo do processo.

3. Línguas controladas: a pré-edição repetitiva pode ser evitada ao se definir uma língua controlada (Nyberg et al, 2003; Arnold et al, 1993, p. 147; Hutchins & Somers, 1992, p. 151; Lockwood, 2000) a ser utilizada pelos autores. Trata-se de uma variante da língua-fonte com restrições lexicais e sintáticas para evitar problemas com a tradução automática. A adaptação de uma língua controlada é, sem dúvida, uma atividade que necessita de investimento financeiro e só será rentável se o processo de pré-edição for repetitivo, caso contrário, não é tão indicado.

### **1.8 Abordagens à Tradução Automática**

De acordo com Forcada (2010), há duas abordagens principais para a TA: desde os primórdios da junção entre linguística e informática até a década de 1990 a abordagem dominante era conhecida como tradução baseada em regras ou no conhecimento, por meio da qual equipas compostas por especialistas em informática e em tradução desenvolviam softwares dotados de mecanismos de análise gramatical e da compilação de dicionários para a transformação de estruturas de frases e afins em formatos que podiam ser processados por TA (p. 218), cujas ferramentas de destaque eram Systran e Apertium. Entretanto, desde o início da década de 1990 foi introduzida uma nova abordagem, baseada em *corpus*, através da qual softwares de tradução “aprendem a traduzir” por meio da análise de textos bilíngues onde milhões de frases/segmentos em um idioma são alinhadas com suas contrapartidas no idioma alvo, tal como conhecemos no processo de TAC. As ferramentas em destaque nesse período eram Microsoft e Google Translator.

Uma breve comparação entre essas duas abordagens mostra que os sistemas baseados em regras levam mais tempo para serem construídos, tendo em vista todas as variáveis que precisam ser agregadas ao software, porém tornam-se mais viáveis se observarmos que a abordagem baseada em *corpus* necessita de um grande volume de textos bilíngues previamente alinhados. Portanto, podemos observar a dificuldade de aplicação deste, em se tratando de idiomas minoritários ou de menos recursos sem extensos corpora bilíngues (Forcada, 2010).

A partir do ano de 2014 uma nova abordagem foi implementada, a TA baseada em redes neurais, a qual tem como base o uso de *corpus* e da TA baseada em estatística para que, através do uso de IA e de cálculos probabilísticos, ofereça uma tradução mais aproximada de um texto alvo esperado (Forcada, 2017), cujas ferramentas de destaque são o DeepL e o Google Translator.

### **1.8.1 Abordagem Baseada em Regras**

Dentre os sistemas de tradução automática baseados em regras, os mais usuais são os sistemas de transferência. Segundo Hutchins & Somers (1992) um sistema de transferência ideal possui três estágios característicos:

- 1) A análise produz, a partir da frase em LP, uma representação intermediária abstrata, em que são estabelecidas classificações e agrupamentos linguísticos que permitem a aplicação das regras gerais de tradução. Como, por exemplo, na tradução do inglês para o português do segmento “a comfortable cushion” consiste em um determinante, de um adjetivo e de um substantivo, é possível aplicar posteriormente uma regra geral que reordena esta sequência de maneira que seja compreensível na LC, determinante – substantivo – adjetivo, “uma almofada confortável”.
- 2) A transferência converte a representação intermediária entregue por análise em uma nova representação intermediária para a LC, ao procurar palavras no dicionário bilíngue (transferência lexical) e, por exemplo, a aplicação da transferência estrutural, conforme vimos no primeiro estágio.
- 3) A geração produz uma frase concreta na LC a partir desta representação intermediária abstrata.

A análise para o par de idiomas A-B pode ser usada como referência para o par de idiomas A-C, já a geração para o par de idiomas A-B pode ser usado como referência para o par de idiomas D-B. As representações intermediárias podem ser mais ou menos complexas dependendo da análise sintática da frase ou até mesmo da representação semântica. Tal análise pode ser tão profunda que é possível se obter uma representação neutra em relação ao idioma, sem a necessidade de transferência.

Tais sistemas são chamados de sistemas interlíngua e, segundo Mitamura (1993), citado por Forcada (2010), têm a vantagem de que nenhum conhecimento bilíngue é necessário para adicionar um novo idioma a um sistema existente e que apenas os módulos de análise e geração do novo idioma devem ser adicionados, mas projetar um sistema interlíngua de propósito, equivale a projetar um modelo completo do mundo real (ou até mesmo de mundos hipotéticos), o que pode restringir a atuação de sistemas interlíngua à realização de traduções de domínio limitado.

### **1.8.2 Abordagem Baseada em *Corpus***

A tradução automática baseada em *corpus*, também é chamada de tradução baseada em dados, e pode ser dividida em dois paradigmas principais: tradução automática baseada em exemplos e tradução automática baseada em estatística. Para que tais abordagens possam ser realizadas, ambos necessitam de um *corpus* de texto paralelo bilíngue com segmentos alinhados. Segue abaixo a distinção dos dois tipos de tradução abarcados pela abordagem baseada em *corpus*.

#### **Tradução Automática Baseada em Exemplos**

A tradução automática baseada em exemplos (Carl & Way, 2003) foi inicialmente formulada por Nagao (1984) como “tradução por analogia” e, geralmente, é descrita como sendo composta por três fases distintas:

- Correspondência: a nova frase a ser traduzida é segmentada e os segmentos são comparados com segmentos idênticos ou similares localizados no *corpus* da LP.
- Alinhamento: os fragmentos correspondentes na LC são combinados de maneira a construir “unidades de tradução”.
- Recombinação: os textos em LC destas “unidades de tradução” são combinados em uma tradução para a nova sentença.

De acordo com Arnold (2003), estas três fases são geralmente vistas como paralelas e complementares aos estágios de análise, transferência e geração as quais vimos anteriormente na abordagem baseada em regras. Importante ressaltar que, quando a nova frase é idêntica a uma frase localizada no *corpus*, sua tradução é recuperada e utilizada,

do mesmo modo como opera um sistema de memória de tradução no software de tradução assistida por computador. Cada uma dessas fases apresenta desafios específicos e ainda são objeto de pesquisa, tais como, a natureza dos segmentos e o processo de segmentação que são cruciais para o alinhamento, especialmente quando os pares linguísticos não são comumente relacionados, e o resultado esperado pode não ser satisfatório por questões de falta de concordância, palavras repetidas, entre outros.

### **Tradução Automática Baseada em Estatística**

A tradução automática estatística teve o início de sua aplicação por meio dos estudos de Brown et al. (1988), citados por Forcada (2010) e, desde então, tem dominado as pesquisas de TA. Baseia-se no paradigma da probabilidade, através do qual diz-se que uma frase na LP e uma frase na LC são a tradução uma da outra com uma certa probabilidade. A situação reside no facto de que qualquer frase na LC possui a probabilidade de ser a tradução de uma frase na LP e a tarefa do sistema de TA é encontrar a frase na LC para qual esta possibilidade se aproxima da probabilidade mais alta possível.

Segundo Forcada (2010), o modelo em questão é geralmente composto por vários componentes, tais como a análise das probabilidades lexicais, onde uma palavra na LP e uma palavra na LC sejam traduções mútuas; da análise das probabilidades de alinhamento, realizada por meio da descrição de processos como o reordenamento de palavras; ou do modelo de probabilidade da LC definir se uma frase é natural ou aceitável independente da frase em LP, entre outros. No modelo “puro” de TA baseado em estatística, em princípio, não há nenhuma abordagem linguística. Os modelos são baseados somente em estimativas estatísticas complexas que fazem uso de corpora bilíngues extensos, alinhados por frases e utilizam segmentos mais longos, mesmo que não sejam unidades sintáticas no sentido linguístico.

### **1.8.3 Tradução Automática Neural**

De acordo com Forcada (2017), a tradução automática neural (TAN) é baseada na TA de *corpus* que aplica uma grande rede neural artificial para prever a probabilidade de uma sequência de palavras ou segmentos. Ao contrário da TA baseada em estatística, que

consome mais memória e tempo, a TAN faz uso de IA para treinar suas partes de ponta a ponta para maximizar seu desempenho. O nome TAN vem do facto de que, as redes neurais artificiais nas quais essa tecnologia se baseia, são compostas de milhares de unidades que se assemelham aos neurônios, na medida em que a entrada ou saída dos dados depende de estímulos recebidos de outros neurônios e da força das conexões ao longo das quais esses estímulos são transmitidos (Forcada, 2017).

A inteligência artificial da TAN lê cada segmento de origem para formar representações possíveis de modo que os resultados se aproximem ao máximo das traduções de referência ou das normas correspondentes. Para tal, treina a rede neural, ou seja, determina o peso ou força de cada uma das ligações entre neurônios de modo que os resultados desejados sejam obtidos. A TAN requer corpora extensos, como os tipicamente utilizados pela TA baseada em estatística e, durante o treinamento, os pesos são modificados de tal forma que o valor de uma função de erro específica ou função de perda descreva até que ponto os resultados da TA estão distantes da tradução de referência. Para que isso ocorra, são utilizados algoritmos de treino que calculam pequenas correções a pesos que são repetidamente aplicados até que a função de perda seja reduzida ao mínimo aceitável. Dessa forma, o sistema atribui a máxima probabilidade a toda a tradução de referência para todos os pares linguísticos fonte/alvo no conjunto de treinamento (Forcada, 2017). A maioria dos sistemas TAN são construídos e treinados de tal forma que se assemelham a um dispositivo de preenchimento de texto como a função de preenchimento automático dos teclados dos telemóveis, a qual informa uma representação da frase de origem, ou mais especificamente, uma representação de cada uma das palavras da frase de origem em seu contexto, construído pela parte codificadora do sistema. O codificador constrói o texto e o decodificador fornece, em cada posição da frase, a probabilidade da palavra ser uma continuação do que já foi produzido e, a melhor tradução é construída a partir da escolha da palavra mais provável em cada posição. Com base nesse princípio, Peris et al. (2017) propuseram o que chamam de TA neural interativa, por meio da qual, o sistema sugere possíveis continuações que podem ser aceites ou ignoradas por tradutores profissionais à medida que o texto-alvo é digitado (Forcada, 2017). Finalmente, para que possamos observar a evolução das abordagens propostas e traçar um comparativo entre elas, a tabela abaixo mostra o resumo da

retrospectiva dos sistemas de TA ao longo do tempo, ressaltando o período de atuação, principais ferramentas utilizadas e pontos de destaque das respectivas abordagens.

<b>TA baseada em regras</b>	<b>TA baseada em <i>corpus</i></b>	<b>TA baseada em redes neurais</b>
1970 – 1990	1990 – 2013	2014 - Atual
Systran, Apertium	Microsoft, Google	Google, DeepL
A "abordagem clássica". Baseia-se em regras linguísticas e dicionários bilingues integrados.	Baseia-se em modelos de tradução estatística baseados na análise de corpora multilingues.	TA baseada em estatística + TA neural para prever a probabilidade de uma sequência de palavras.
Previsível. Complexo e caro.	Imprevisível. Uso de dados disponíveis.	Previsível. Necessita de treino.

Tabela 1 - Tabela comparativa das abordagens da TA ao longo do tempo.

A tabela mostra que a evolução das abordagens da TA está intrinsecamente ligada ao desenvolvimento das inovações tecnológicas no decorrer do tempo. Nota-se que, a partir do momento em que o acesso aos dispositivos informáticos foi popularizado, em meados da década de 1990, o aprimoramento das ferramentas apresentou desenvolvimento significativo por meio do investimento em pesquisas e através do crescimento constante de seu uso empírico. Assim, a adoção de ferramentas de TA e de TAC por parte dos prestadores de serviços linguísticos tem contribuído para a oferta cada vez mais acessível e diversa de ferramentas que acompanham e suportam o crescimento contínuo do mercado da tradução.

O próximo tópico aborda a pós-edição como parte integrante do processo tradutório, o qual, na primeira geração de dispositivos de TA, foi um dos fatores que fez com que o investimento no desenvolvimento dessas ferramentas fosse drasticamente reduzido, pois os primeiros investigadores entendiam que a TA deveria resultar em um texto sem a necessidade de intervenções. Com o desenvolvimento de novas tecnologias e com uma visão mais ampla a respeito dos benefícios, especialmente com relação à agilidade e à produtividade que a TA pode oferecer, a PE voltou a ser amplamente difundida e praticada de modo a trazer luz sobre a mais nova atividade atribuída ao tradutor, a de pós-editor.

## 2. A Pós-Edição

Como vimos até o momento, a TA tem evoluído gradualmente ao longo dos anos através da implementação de novas tecnologias e do desenvolvimento de softwares, tanto os inseridos na estrutura de ferramentas TAC como os softwares disponibilizados de forma autônoma. De igual modo, a indústria da tradução tem apresentado um crescimento no volume de conteúdos traduzidos, bem como sofrido pressões para o aumento da velocidade e da produtividade.

Nesse contexto surge a pós-edição (PE) que se relaciona diretamente com o uso da TA e propõe uma mudança significativa nos paradigmas propostos no período que antecede o uso das ferramentas informáticas. A PE torna-se necessária quando o resultado da TA necessita de ajustes para que o produto apresente a qualidade suficiente para divulgação. No início de seu estudo *Post-editing of Machine Translation: Processes and Applications*, O'Brien et al (2014) relacionam a PE com a utilização da TA do seguinte modo:

Post-editing is possibly the oldest form of human-machine cooperation for translation, having been a common practice for just about as long as operational machine translation systems have existed. Recently however, there has been a surge of interest in post-editing among the wider user community, partly due to the increasing quality of machine translation output, but also to the availability of free, high-quality software for both machine translation and post-editing. (O'Brien et al. 2014:ix)

Assim, a PE é, atualmente, o tópico com abordagem cada vez mais presente no mundo da tradução, e isso se reflete no número expressivo de artigos científicos publicados na área, que só em 2021 chegaram ao número aproximado de 28 500 artigos publicados na plataforma Google Acadêmico, obtidos com as palavras-chave *post-editing machine translation*. Além dos estudos de ganho de produtividade e velocidade, esse tópico envolve outros fatores que devem ser observados, tais como o esforço cognitivo, formação de pós-editores e qualidade. O interesse na TA e PE reflete-se também na realização de conferências, fóruns e *webinars* para a fomentação e reflexão sobre o tema, tais como o Workshop on Post-Editing Technology and Practice realizado bianualmente desde 2012 como parte integrante do MT Summit (<https://amtaweb.org/>), cujo próximo evento será realizado de forma híbrida (online e presencial) em setembro de 2022 na

cidade da Flórida, Estados Unidos da América. A temática do evento abordará os avanços da TA orientada por dados (por exemplo, neural e estatística), TA para idiomas com poucos recursos, TA para comunicação (blogs, redes sociais e bate-papos), bem como avaliação da TA. Outro evento de relevância para a área é o Conference on Machine Translation que em 2022 realizará a sua décima sétima edição a qual será acontecerá no mês de dezembro na cidade de Abu Dhabi, capital dos Emirados Árabes. A temática desse evento abordará assuntos como: PE automática, estimativa de qualidade, tradução geral por máquina e terminologia (<https://machinetranslate.org/>).

## **2.1 Definição de Edição, Pré-Edição, Pós-Edição e Revisão**

A PE é uma das etapas inseridas mais recentemente no processo de tradução realizado pelas empresas prestadoras de serviços linguísticos, nesse contexto é importante definir o papel de cada etapa e de como elas atuam, separadamente ou em conjunto, no intuito de oferecer o resultado mais adequado de acordo com as necessidades dos clientes.

Segundo Pym (2011), a edição é a realização de alterações a um texto numa situação em que a progressão linear é ausente (no caso de um texto gerado automaticamente a partir de TA, por exemplo) ou completa (isto é, a redação ou tradução por completo). A edição pode aplicar-se a traduções ou textos não traduzidos, embora seja preferível o uso dos termos revisão, autorrevisão ou outra revisão para trabalhos de tradução. Quando a saída da TA está a ser corrigida ou alterada, o termo mais apropriado é pós-edição, uma vez que a revisão implicaria que o processo de escrita do texto fosse concluído. Os vários tipos de edições podem ser encontrados em manuais de estilo (edição de cópias, edição estilística, edição estrutural) e podem ser adaptados para se adequarem ao problema que precisa ser resolvido (pp. 80-81).

De acordo com Austerlühl (2001), o processo de pré-edição envolve a preparação de um determinado texto de partida (*input*) a fim de evitar problemas desde o início do processo de tradução. A preparação do texto de partida na língua inglesa, por exemplo, evita a tradução direta de expressões idiomáticas, evita a omissão de pronomes antes de um verbo, evita a omissão de pronomes relativos (por exemplo, *that*, *who*, *whom*), mantém as estruturas das frases mais claras, simples e diretas no intuito de expressar uma ideia por frase, mantém a estrutura oracional sujeito-verbo-objeto a fim de evitar o uso da

voz passiva, evita a divisão de verbos (por exemplo, *look up*) e mantém o inglês padrão, formal, no qual as ligações gramaticais são claramente expressas. Isto implica que podem ser esperados problemas se essas regras de escrita forem ignoradas. Outro ponto frequentemente mencionado em relação à pré-edição, diz respeito à utilização de expressões simples e perfeitamente claras para evitar o problema de ambiguidade semântica e a identificação de texto que deve permanecer sem tradução, como nomes próprios e endereços, por exemplo (pp. 163-164).

O controle da língua proposto pelo processo de pré-edição nos leva a buscar a definição do que se entende por língua controlada ou língua especial. De acordo com Arnold et al. (1994, p.156) a linguagem controlada é “a form of language usage restricted by grammar and vocabulary rules”. Os escritores técnicos, em sua maioria, apresentam uma atitude de resistência quanto ao uso da língua controlada pois alegam que ela restringe a criatividade envolvida no processo tradutório. No entanto, as multinacionais como a Ericsson, a britânica Perkins Engines, entre outras, têm de respeitar o uso de linguagem controlada no intuito de melhorar a legibilidade da documentação dos seus produtos. As ‘Regras de Inglês Controlado’ que os autores técnicos da Ericsson têm de respeitar, por exemplo, incluem a instrução de que uma palavra deve ser atribuída a apenas um significado, que cada palavra deve ser usada em apenas uma classe de palavras, e que as estruturas sintáticas complexas, tais como cláusulas condicionais devem ser evitadas (Schawanke, 1991). Não é surpreendente que os textos originais escritos em línguas controladas geralmente produzam melhores resultados com a TA do que textos escritos em línguas não controladas. No entanto, mesmo que o texto original tenha sido pré-editado ou escrito numa língua controlada, o texto alvo precisa ser pós-editado a fim de se conseguir uma tradução de alta qualidade (Austermühl, 2001).

Em linhas gerais, as línguas especializadas são utilizadas em campos técnicos específicos e são caracterizadas por itens lexicais cuidadosamente definidos e pelo facto de certas estruturas sintáticas serem utilizadas mais frequentemente do que outras, como nos manuais técnicos, por exemplo, onde predominam as formas de instrução, o que significa que o sistema de TA pode ser adaptado a essas estruturas. Um dos exemplos mais conhecidos é o Sistema Météo utilizado no Canadá para a tradução de boletins meteorológicos. Sistemas como esse podem beneficiar o facto de que os textos em línguas especiais tendem a ser menos ambíguos do que texto generalistas (Austermühl, 2001).

Segundo Pym (2011), a pós-edição refere-se ao processo de se fazer correções ou emendas ao texto gerado automaticamente, nomeadamente o resultado de tradução automática. O termo funciona em oposição à pré-edição, que é o principal meio alternativo de melhoria da qualidade da produção de TA. Estes dois termos não fazem muito sentido em situações em que não há geração automática de texto envolvida.

De acordo com Austermühl (2001), Wills (1998, p. 183) argumenta que as tarefas de pré-edição e pós-edição devem ser colocadas nas mãos de tutores profissionais que estejam conscientes das limitações da TA, uma vez que sem essa intervenção profissional o processo é muito demorado e ineficiente.

O pós-editor, segundo Allen (2001), tem como atividade principal editar, modificar e/ou corrigir texto traduzido que tenha sido processado previamente por um sistema de TA, de uma LP para uma LC. Ainda, de acordo com Allen (2003, p. 297), a PE é definida por Veale & Way (1997) como “term used for the correction of the machine translation output by human linguistics/editors”. Wagner (1985), por sua vez, define a PE como a correção de texto previamente traduzido, em oposição à tradução de raiz. Para Somers (2001), a PE é definida como: “tidying up the raw output, correcting mistakes, revising entire, or, in the worst case, retranslating entire sections” (p. 138). A PE também pode ser analisada a partir de outras propostas no que se refere ao ato de correção de uma tradução, conforme afirma O’Brien et al (2014):

(...) in translation, the term of post-editing refers to the act of correcting a translation proposal (from a single word or character to a complete document). If this proposal comes from an MT system, we talk of post-editing MT (PEMT); if it comes from human translation, we talk of human post-editing (HPE). In general, post-editing itself is performed by (human) translators, but it can also be performed automatically: we then talk of automatic post-editing (APE) (p. 26).

Em suma, PE é entendida como o processo de correção, feita por um tradutor pós-editor, de uma tradução previamente realizada por máquina no intuito de melhorar o texto para disseminação.

Para Allen (2003) existem três níveis de pós-edição: a pós-edição ligeira, a pós-edição mínima e pós-edição completa. A pós-edição ligeira é estritamente necessária para a remoção de erros gritantes e significativos, sem considerar questões estilísticas. A pós-edição completa representa a outra extremidade, que implica uma elevada qualidade dos textos resultantes. Já a pós-edição mínima ou intermediária é onde se concentra a maior

parte dos trabalhos de pós-edição, é a categoria que Allen (2003) qualifica como difusa ou abrangente. O termo faz referência há um número mínimo de alterações que são aplicadas a fim de produzir uma qualidade “aceitável”. A definição do que seria uma qualidade “aceitável” é uma questão de debate, pois depende de qual será o público leitor do TC e qual o destino desse texto, se será para distribuição interna ou externa, por exemplo (Depraetere, 2010).

Para Guzmàn (2007), as correções linguísticas devem limitar-se a "corrigir" certos tipos de erros, ou seja, erros gramaticais e de sintaxe (por exemplo, concordância errada (número, género, ordem das palavras que causam problemas gramaticais), erros ortográficos e de pontuação (por exemplo, falta de acentos, capitalização errada), erro de tradução (por exemplo, uso errado de terminologia; frases corretas com uma terminologia diferente significado da fonte).

A revisão é o termo utilizado na norma europeia de qualidade EN 15038 (2006) para quando uma pessoa que não seja o tradutor corrige a tradução. O Prestador de Serviços de Tradução (PST) deve assegurar que a tradução seja revista. A norma define "revisão" como "examinar uma tradução quanto à sua adequação ao objetivo acordado, e esta deve incluir, como requerido pelo projeto, a comparação dos textos de origem e de destino para a coerência terminológica, registo e estilo. Tendo em conta as recomendações do revisor, o PST deve tomar medidas para assegurar que quaisquer correções necessárias sejam implementadas. O termo "revisão" é por vezes utilizado mais vagamente para qualquer processo de revisão realizado por uma pessoa que não o tradutor, e refere-se também ambigualmente a resenhas de livros e revisões gerais. A revisão pode ser dividida em pré-projeto e pós-projeto, onde a revisão pré-projeto ocorre antes do tradutor apresentar a versão final do texto; e a revisão pós-projeto ocorre após a tradução ter sido concluída. A "revisão em projeto" também pode ser dividida em "revisão por sentença", "revisão por parágrafo", e assim por diante, conforme necessário. (Pym, 2011).

É possível que haja confusão entre o que é revisão e pós-edição, pois se trata basicamente do mesmo processo, mas o que define a diferença entre um termo e o outro é que a revisão se refere ao processo de correção de texto produzido pelo profissional de tradução e o processo de pós-edição se refere à correção de texto traduzido por máquina.

## 2.2 Produtividade e Esforço Cognitivo

O fluxo natural dos avanços tecnológicos leva as descobertas do âmbito das pesquisas em laboratório para o domínio público, onde efetivamente é testada quanto à sua utilidade. Inicialmente, isto pode agir como um disruptor - os processos utilizados e aceitos durante anos são perturbados, o que, por sua vez, leva compreensivelmente a perguntas, preocupações ou irritação.

Estudos realizados por Gerberof (2009) e Plitt & Masselot (2010) demonstram que a PE realizada em textos em que a TA foi aplicada pode, de facto, aumentar a produtividade de tradutores profissionais em comparação com a tradução manual realizada “do zero”. No entanto, quando se trata de uma TA deficiente, a atividade não apresenta nível de produtividade satisfatório, especialmente no caso dos pares linguísticos com menor frequência de uso, o que resulta em um maior esforço cognitivo por parte do pós-editor (Fontes, 2013).

Um estudo na Autodesk, relatado por Plitt e Masselot (2010), envolveu doze tradutores profissionais trabalhando do inglês para o francês, italiano, alemão ou espanhol e comparou sua produtividade (palavras traduzidas por hora) quando da PE da TA à tradução manual. Na média, verificou-se que a PE aumentou a produção em 74%, embora houvesse uma variação considerável entre os diferentes tradutores (Plitt e Masselot 2010, p. 10). O estudo de continuação relatado por Zhechev (2014) incluiu outros idiomas-alvo e mostrou grande variação nos ganhos de produtividade entre diferentes pares de idiomas.

Em resumo, a pesquisa sobre a produtividade da pós-edição mostra resultados parcialmente conflitantes: alguns estudos relataram ganhos de produtividade bastante notáveis ao comparar a PE com a tradução manual, enquanto outros não encontraram diferenças significativas. Os ganhos de produtividade parecem depender, pelo menos em parte, de condições específicas. Estas condições estão relacionadas à tradução automática de alta qualidade suficiente, que atualmente é possível para certos pares de idiomas, e sistemas de tradução automática voltados para o tipo de texto específico que está sendo traduzido. A familiaridade dos pós-editores com as ferramentas e processos envolvidos na tarefa de PE também pode ter um papel importante. Além disso, a comparação dos próprios pós-editores mostra variações individuais: os diferentes processos de trabalho

das diferentes pessoas parecem afetar o quanto elas se beneficiam com o uso da TA (Koponen, 2016).

O esforço cognitivo envolvido na PE da TA é um dos pontos-chave através do qual são levantados questionamentos sobre até que ponto esse esforço é aceitável e qual a sua importância em termos de condições de trabalho das pessoas envolvidas no ciclo tradutório. Com base nesse ponto, Thicke (2013), afirma que os pós-editores, em última análise, são os maiores prejudicados em se tratando da má qualidade da TA. Nesse contexto, surge a necessidade do estabelecimento de formas objetivas para a medição do esforço visando a definição do preço que será cobrado. Geralmente, a fixação de preços baseia-se no valor/hora do profissional ou na determinação de uma escala de pontuação semelhante ao *fuzzy match* utilizado pela MT dos sistemas TAC, o que nem sempre condiz com a medida exata do esforço envolvido (Guerberof & Arenas, 2010).

O estudo desse tópico, geralmente, baseia-se na velocidade ou no número de páginas produzidas através da TA pós-editada em comparação com a tradução manual, mas este aspecto revela somente parte do esforço envolvido. De acordo com Krings (2001), o esforço aplicado na pós-edição distingue-se de maneira tridimensional por meio do aspecto do tempo, da técnica e do esforço cognitivo. O aspecto temporal refere-se ao tempo necessário para a PE de um texto em particular; o aspecto técnico refere-se a supressões, inserções e reordenação de texto; o esforço cognitivo trata da extensão e do tipo de processos cognitivos que o tradutor precisa aplicar em uma saída de TA, ou seja, a quantidade de esforço gasta em uma tarefa de PE (O'Brien, 2006).

Uma das teorias que dá embasamento ao estudo do esforço cognitivo é a Teoria da Relevância (TR) a qual postula que a cognição humana é guiada pela relevância, ou seja, essencialmente a busca de informações que sejam relevantes para a obtenção do maior número possível de efeitos cognitivos. Em outras palavras, a TR se baseia na premissa de que nossa cognição procura alcançar os maiores efeitos cognitivos com o mínimo esforço de processamento necessário. De acordo com Wilson (2011, p. 72), o sistema cognitivo humano tende a "seguir um caminho de menor esforço na busca de implicações; interpretações de teste em ordem de acessibilidade, e parar quando suas expectativas de relevância estão satisfeitos" (Alves, et al., 2016).

Na PE, entretanto, devido à maior pressão por conta do tempo escasso, espera-se que o pós-editor economize tempo e esforço cognitivo, pois o sistema de TA gera uma

primeira versão do texto de destino. Assim, ainda seguindo o princípio da relevância, o pós-editor tentará encontrar uma solução suficientemente boa, levando em consideração a modalidade de PE acordada para a realização da atividade - leve ou completa, calculando seus efeitos no ambiente cognitivo do público e, então, aceitando integral ou parcialmente a produção de TA bruta. Mais especificamente, ao mesmo tempo em que se faz a pós-edição de uma saída de TA com a ajuda de uma ferramenta TAC, serão oferecidas ao pós-editor algumas soluções opcionais para um problema específico no caso de ele decidir mudar algo em uma unidade de tradução (Alves, et al., 2016).

Estas soluções oferecidas pela TA e pela ferramenta TAC mudarão o propósito de leitura do pós-editor, já que, a partir de uma mudança inicial, todo o restante da saída é automaticamente alterado, deixando ao pós-editor a decisão de tanto ler para revisão quanto para pós-edição. A este respeito, Jakobsen & Jensen (2008) investigaram os efeitos do tipo de tarefa (leitura para compreensão, para tradução, para tradução visual e para tradução escrita) sobre os movimentos oculares e concluíram que o propósito da leitura teve um efeito claro sobre os movimentos oculares e sobre o tempo de observação. Assim, espera-se que uma tradução que seja resultado de TA juntamente com a ferramenta TAC mostrem um efeito sobre os movimentos oculares relacionados aos processos de PE, reduzindo o esforço de processamento cognitivo exigido durante a revisão de saída (Alves, et al., 2016).

### **2.3 O Ensino da Pós-Edição**

A necessidade de proporcionar formação especializada em pós-edição, baseia-se principalmente na demanda crescente da indústria linguística que anseia por profissionais qualificados. Um estudo encomendado pela Direção Geral de Tradução da União Europeia em 2009, estimou o valor da indústria linguística nos Estados-membros da União Europeia em cerca de 8,4 mil milhões de euros em 2008. Embora já fosse um número considerável, a investigação mais recente de estudos de mercado propõe números que vão até cerca de 50 mil milhões de dólares para o mercado global em 2019-2020 (estimativa da Statista Inc e CSA), o que sugere um valor de cerca de 20 mil milhões de euros para os Estados-membros da UE (ELIS, 2022).

Quando o primeiro inquérito ELIS – *European Language Industry Survey*, que na altura era conhecido como o inquérito EUATC – *European Union Association of Translation Companies* – foi realizado pela primeira vez em 2013, a TA ainda era vista principalmente como uma ameaça e ao mesmo tempo como um desafio, e somente algumas empresas linguísticas a viram como uma oportunidade. Desde então a TA percorreu um longo caminho. Hoje, 65% das empresas participantes do inquérito ELIS veem a melhoria da qualidade da TA como uma oportunidade e não como uma tendência. Mas 35% dos profissionais independentes inquiridos a veem como uma oportunidade e 41% como uma ameaça. O facto é que a tecnologia já está integrada aos processos tradutórios e de acordo com os resultados do ELIS de 2022, 58% das empresas consultadas implementaram a tecnologia e 20% estão a planear fazê-lo, e mais de 70% dos profissionais linguísticos independentes afirmam estar a utilizar a tecnologia até certo ponto. O relatório aponta que os departamentos linguísticos de agências públicas internacionais e empresas privadas estão a atingir ou a aproximar-se dos 90% e os académicos inquiridos relatam que a TA está implementada ou planeada (ELIS, 2022).

Em se tratando de um contexto de formação, para que o conteúdo do curso possa ser determinado, é necessário definir quais as competências que o pós-editor deve ter. O'Brien (2002), por exemplo, argumenta que existe um lado teórico, bem como um lado prático no ensino da pós-edição. A fim de formar os pós-editores, sugere as seguintes disciplinas que devem ser abordadas no componente teórico: Introdução à Pós-edição, Introdução à Tradução Automática, Introdução à Autoria de Linguagem Controlada, Gestão Terminológica Avançada, Linguística de Texto Avançada e Habilidades Básicas de Programação.

O'Brien (2002) ainda especifica qual a abordagem que deve ser dada a cada componente teórica como podemos ver a seguir. A disciplina de Introdução à Pós-Edição, por exemplo, deve abordar a necessidade da PE e os fatores que a diferenciam da tradução e revisão, quais os níveis de PE existentes, como são determinadas as necessidades dos utilizadores, qual a tecnologia podemos aplicar à PE, como fazer a classificação de erros etc. A disciplina de Introdução à Tradução Automática deve cobrir a história dos tipos de sistemas de TA, a descrição dos sistemas comerciais de TA, as metodologias de avaliação, o estado atual dos sistemas disponíveis, incluindo a integração com ferramentas de memória de tradução, e as perspectivas para o futuro. Com relação à disciplina de

Introdução à Autoria de Linguagem Controlada, O'Brien sugere que deve incluir a história das línguas controladas, a descrição das ferramentas de línguas controladas, avaliação de metodologias para ferramentas de línguas controladas, estado atual de conhecimento sobre o tema, integração entre a autoria de línguas controladas e a ferramentas de TA, além das perspectivas futuras para a área. Já a disciplina de Gestão Terminológica Avançada deve ter como objetivo o desenvolvimento de competências básicas de gestão terminológica de modo a gerar discussão sobre os pontos fracos e fortes das ferramentas de gestão terminológica, codificação de dicionário para TA, e salienta como ponto de destaque o intercâmbio terminológico entre ferramentas que utilizam normas de intercâmbio terminológico como o XLT, OLIF e TBX. Com relação à disciplina de Linguística de Texto Avançada, O'Brien sugere que o desenvolvimento do tema deve basear-se nas competências linguísticas básicas do aluno, introduzindo-as nas normas de textualidade, classificação do tipo de texto, e no uso de ferramentas de análise de *corpus* linguístico e de análise de corpus para a análise de tipos de textos. Por fim, sugere que a disciplina de Competências Básicas de Programação introduza o aluno às noções básicas de programação e o instrua na programação macro e numa linguagem de programação adequada ao Processamento de Linguagem Natural, como o software Perl, por exemplo (O'Brien, 2002).

Segundo Depraetere (2010), o nível de conforto com a pós-edição é grandemente aumentado após a pós-edição 100 000 palavras, pois essa é a média sugerida para refletir o número de horas de formação necessárias para transformar o profissional num pós-editor competente. Dado que 100 000 palavras representam um mês de trabalho completo, ficará claro que apenas uma pequena percentagem da formação pode ser alcançada durante os seminários propriamente ditos. O'Brien (2002) salienta que os estudantes devem ser encorajados a praticar a pós-edição tanto dentro como fora das horas destinadas ao curso. A componente prática deve envolver exercícios com várias ferramentas comercialmente disponíveis e diferentes níveis de pós-edição (leve e completa) devem ser praticados. A proposta de O'Brien (2002) representa um ideal que pode não ser fácil de alcançar dado o número de horas de contato e a gama de tópicos a serem abordados na formação em tradução.

Tendo em conta que os programas de formação de tradutores são estruturados de forma diferente em cada instituição, é impossível definir onde se enquadra um curso de especialização como o que foi proposto. No entanto, o sucesso da pós-edição requer um elevado nível de confiança no próprio trabalho do pós-editor, segundo Wagner (1987) citada por Depraetere (2010). Na introdução ao seu livro sobre revisão e edição tradicional, Mossop (2020) relata a sua descoberta de que os estudantes de graduação raramente estão prontos para a auto-revisão ou revisão do trabalho dos outros até que tenham completado um trabalho na prática. Isto também sugere que a experiência e a confiança são ingredientes necessários para a tarefa de pós-edição. Na proposta de curso sugerida, partiu-se do princípio de que os estudantes já tivessem algumas competências prévias, tais como, excelentes competências linguísticas, competências de tradução especializada, linguística básica, gestão terminológica básica, competências informáticas e uma introdução à tecnologia linguística. Isto implica dizer que esse tipo de curso só deveria ser oferecido na última parte de um programa de formação de tradutores ou num programa de pós-graduação, pois é mais provável que os estudantes interessados possuam a experiência e a confiança necessárias para um melhor aproveitamento (O'Brien, 2010).

Acredito que a formação em pós-edição não é imprescindível, mas é uma mais valia para que o profissional possa realizar um trabalho de qualidade, tendo em conta que os temas sugeridos por O'Brien (2002) estão em consonância com o que o mercado da tradução espera, ou seja, um profissional que perceba não só das questões linguísticas envolvidas no processo de pós-edição, mas também do desenvolvimento das ferramentas e competências básicas de programação e processamento de linguagem natural. O modelo de curso a ser oferecido poderia ser por meio de disciplinas optativas dentro de um curso de licenciatura e mestrado ou num curso de pós-graduação. O ponto a seguir tem como objetivo a avaliação da qualidade envolvida nos processos de TA tendo em conta a abordagem da pós-edição realizada pelo profissional da tradução e pelos avanços obtidos na pós-edição realizada por máquina.

## **2.4 Garantia da Qualidade da Tradução Automática**

Conforme afirma Pym (2011), o conceito de qualidade é notoriamente problemático, pois a noção de "alta qualidade" absoluta estabelece o ideal da tradução perfeita. Entretanto, sabemos que as traduções podem e devem ser diferentes para diferentes fins e sob diferentes condições de trabalho, por exemplo, uma tradução feita dentro do prazo pode ser mais aceitável do que uma tradução mais precisa, porém entregue tardiamente.

De acordo com Gouadec (2007), as atividades envolvidas na prestação de um serviço de tradução são organizadas em três fases: pré-tradução, tradução e pós-tradução. A pré-tradução inclui tudo o que ocorre até o momento em que o tradutor recebe o material para a tradução e diz respeito à negociação do trabalho junto ao cliente, à obtenção das especificações corretas e à contratação do serviço. Ainda segundo Gouadec, a tradução, por sua vez, é dividida em três etapas: pré-transferência, transferência e pós-transferência. A pré-transferência inclui todas as operações que levam até a tradução de facto, que inclui a preparação do material, buscas documentais, alinhamento, consolidação de memória, extração de terminologia, entre outros. A transferência refere-se à atividade central do processo, a tradução propriamente dita. A pós-transferência cobre tudo o que seja feito para atender aos requisitos e critérios de qualidade antes da entrega do material traduzido. A pós-tradução abrange todas as atividades que se seguem à entrega do material ao cliente, e isso inclui a possível integração do material traduzido a outros suportes (a inserção de legendas, alterações no layout do texto e a integração a um website, por exemplo), a criação de um ficheiro do projeto onde a terminologia será consolidada para projetos futuros, o pagamento dos profissionais envolvidos etc.

O processo de aferição da qualidade agrega a definição de normas regulamentadoras e o uso de métricas, medidas quantificáveis usadas para analisar o resultado de um processo, a fim de atestar não só a qualidade, mas também a produtividade dos serviços de tradução. As normas regulamentadoras fornecem requisitos para os principais processos, recursos e outros aspectos necessários para a entrega de um serviço de tradução de qualidade que satisfaça as especificações aplicáveis. Já as métricas, asseguraram que um serviço ou produto cumpre um nível mínimo de qualidade de acordo com categorias de erros e níveis de gravidade pré-definidos.

As normas mais utilizadas no âmbito da prestação de serviços linguísticos são: a EN 15038, a ISO 17100 e a ISO 18587. A EN 15038:2006 foi a primeira norma a ser desenvolvida especialmente para prestadores de serviços de tradução e visa unificar a terminologia utilizada no campo da tradução, definir requisitos básicos para os prestadores de serviços linguísticos e criar um quadro para a interação de clientes e prestadores de serviços em termos dos seus direitos e obrigações. Um forte enfoque é dado aos processos administrativos, de documentação, de revisão, bem como às funções dos diferentes especialistas envolvidos no processo de tradução ao longo de sua duração.

A norma ISO 17100:2015, surgiu para substituir a EN 15038 e fornece diretrizes para os principais processos, recursos e outros aspectos necessários para a entrega de um serviço de tradução de qualidade que satisfaça as especificações aplicáveis. O seu conteúdo começa por determinar termos e definições dos conceitos do que é traduzir e o que é tradução, qual é o fluxo de trabalho relacionado ao processo de tradução, qual o produto da tradução, qual é o âmbito da interação entre a empresa prestadora de serviços de tradução e o cliente e qual é o trabalho do intérprete. Com relação aos conceitos relacionados ao fluxo de trabalho e à tecnologia, define o que é TAC e TA, qual é o produto de TA, o que é PE, o que é revisão e a importância do uso de um guia de estilo. Quanto aos conceitos relacionados à língua, define o que é conteúdo do idioma de origem e do idioma de destino, o que é texto, o que é língua natural, o que é registro linguístico, quais as convenções sobre o tipo de texto, qual a delimitação da área temática do texto e qual o significado de *locale*. Com relação aos entes envolvidos no processo de tradução, define quem é o fornecedor de serviços de tradução, o PST, o cliente, o tradutor, o revisor, o gerente de projetos e suas respectivas competências. Concernente aos conceitos relacionados ao controle do processo de tradução, define o que é verificação, o que é documento, o que é registro, correção e ação corretiva. Quanto aos recursos, define quais os recursos humanos, qual a responsabilidade da empresa quanto às tarefas subcontratadas, quais as competências que o tradutor deve ter (competência de tradução, linguística e textual, de investigação, de aquisição e processamento de informações, conhecimentos culturais e técnicos), quais as qualificações esperadas do tradutor, quais as competências esperadas do revisor, qual a competência do gerente de projetos, e qual a importância de que esses profissionais mantenham-se atualizados frente às demandas da área da tradução. Também define quais os recursos técnicos e tecnológicos devem ser

providenciados para a realização dos serviços de tradução, quais os processos e atividades de pré-produção devem ser adotados, define como deve ser o manuseio de informações relacionadas aos projetos dos clientes, como deve ser feita a preparação do projeto, o processo de tradução e o processo de pós-produção. A produção em bruto da tradução e a pós-edição estão fora do escopo dessa norma.

A norma ISO 18587:2017 (ver anexo 1) fornece diretrizes específicas para o processo de pós-edição humana completa dos resultados da tradução automática e das competências dos pós-editores, bem como define os limites dos tipos de PE oferecidos no mercado da tradução: PE leve e a PE completa. Destina-se a ser utilizado por PSTs, seus clientes, e pós-editores e, é aplicável apenas a conteúdos processados por sistemas TA. A norma determina os termos e definições relacionados aos conceitos de TA, língua e conteúdo, pessoas, organizações e tradução.

De acordo com a norma ISO 18857:2017, os objetivos da pós-edição são garantir:

- A compreensibilidade da produção pós-editada;
- A correspondência do conteúdo do idioma de origem e conteúdo do idioma de destino;
- O cumprimento dos requisitos e especificações de PE definidos pelo PST.

Esses objetivos podem ser alcançados assegurando que os seguintes critérios sejam atendidos:

- Consistência terminológica/lexical, assim como o cumprimento da terminologia de domínio;
- Uso de sintaxe padrão, ortografia, pontuação, diacríticos, símbolos especiais e abreviações e outras convenções ortográficas da língua-alvo;
- Conformidade com quaisquer normas aplicáveis;
- Formatação correta;
- Adequação para o público-alvo e para o propósito do conteúdo do idioma de destino;
- Cumprimento dos acordos cliente-PST.

As ferramentas TAC contam com um processo de controle de qualidade próprio. O memoQ, por exemplo, oferece dois níveis de controle de qualidade: o QA – Quality Assurance, que executa uma verificação de possíveis erros durante o processo de tradução e LQA – Linguistic Quality Assurance, que categoriza os tipos de erros e cria a estatística de erros em todo o projeto. As verificações predefinidas de QA incluem a busca por segmentos ou termos não traduzidos, inconsistências de números ou pontuação, espaços, letras maiúsculas e minúsculas, problemas com *tags* e expressões regulares – *Regex*, e poder ser criadas outras verificações de acordo com a necessidade do usuário. Através das configurações de LQA é possível adicionar outro nível à verificação por meio da categorização dos erros com a atribuição de níveis de gravidade e pontos de penalidade. As categorias podem ser personalizadas de modo a produzir mensagens de controle de qualidade com os pontos de penalidade anexados às categorias de erros a partir da qual pode ser criada uma estatística sobre a qualidade linguística do projeto em questão e uma lista de erros com suas categorias e gravidades.

Associações como a LISA – Localization Industry Standards Association (que encerrou suas atividades no ano de 2011), SAE - Society of Automotive Engineers e a TAUS - Translation Automation User Society, desenvolveram seus próprios sistemas de métricas, os quais têm sido largamente utilizados pela indústria da tradução, pois trabalham de maneira integrada ao controle de qualidade LQA das ferramentas TAC.

No sistema de métricas LISA QA há três níveis de gravidade para os erros: crítico, importante e leve. Quanto mais grave é o erro, maior é o número de pontos atribuído a ele, para refletir a sua severidade. Erros leves são os menos graves. Cada erro leve encontrado vale um ponto. Os erros importantes são o segundo tipo mais grave. Cada erro importante vale cinco pontos, ou seja, é igual a cinco erros leves. Erros críticos são o tipo mais sério de erro. Mesmo que apenas um erro crítico seja encontrado, a tradução fica abaixo da métrica imediatamente. Erros críticos valem um ponto a mais que o máximo de pontos de erros permitido. Por exemplo, três erros leves em uma tradução de 200 palavras fariam com que ela ficasse abaixo da métrica, já que dois é o máximo de pontos de erros permitido. No entanto, dois erros leves não fariam uma tradução de 2 000 palavras ficar abaixo da métrica, pois o número máximo de pontos de erros permitido é 18. Basicamente, o número máximo de pontos de erros permitido é proporcional ao número de palavras traduzidas. Essa proporção é menor do que 1% do número total de palavras

(<https://www.terralingua.com.br/qa/?ln=pt>). As categorias de erro do modelo LISA QA estão descritas na tabela 2:

<p><b>Erro de Tradução</b></p> <p><b>Precisão</b></p> <ul style="list-style-type: none"> <li>• Omissões</li> <li>• Adições</li> <li>• Referências Cruzadas</li> <li>• Cabeçalhos/rodapés</li> </ul> <p><b>Terminologia</b></p> <ul style="list-style-type: none"> <li>• Aderência ao glossário</li> <li>• Contexto</li> </ul> <p><b>Idioma</b></p> <ul style="list-style-type: none"> <li>• Gramática</li> <li>• Semântica</li> </ul>	<ul style="list-style-type: none"> <li>• Pontuação</li> <li>• Ortografia</li> </ul> <p><b>Estilo</b></p> <ul style="list-style-type: none"> <li>• Estilo geral</li> <li>• Registro/tom</li> <li>• Variante do idioma/escrita</li> </ul> <p><b>País</b></p> <ul style="list-style-type: none"> <li>• Normas específicas de cada país</li> <li>• Adequação local</li> <li>• Normas da empresa</li> </ul> <p><b>Consistência</b></p>
---	---

Tabela 2 - Categorias de Erros do Modelo LISA QA

O sistema SAE J2450 foi criado originalmente para aferir a qualidade do serviço de tradução para a indústria automotiva e posteriormente passou a ser utilizado pela indústria da tradução em geral. Fornece um sistema de pontuação de erros por meio do qual a qualidade pode ser medida independentemente da língua de origem, da língua de destino e da forma como a tradução é executada. O sistema métrico SAE J2450 apresenta sete categorias de erro primário e duas classificações de erro (nomeadamente, leve ou grave). O avaliador tem primeiro de determinar em qual das sete categorias um erro se enquadra, e depois fazer um julgamento sobre se o erro é grave ou leve. Finalmente, uma pontuação global é calculada de acordo com as pontuações ponderadas em todas as sete categorias. Uma pontuação elevada indica que a tradução é de baixa qualidade (<https://www.sae.org/standardsdev/j2450p1.htm>). Segue a tabela abaixo das categorias de erro do modelo SAE J2450.

Nome da Categoria	Sub-Classificação	Peso
Termo Errado (TE)	grave	5/2
Erro Sintático (ES)	leve	4/2
Omissão (OM)		4/2
Erro de Estrutura de Palavra ou Acordo (EA)		4/2
Erro de Ortografia (EO)		3/1
Pontuação (P)		2/1
Erros Diversos (ME)		3/1

Tabela 3 - Categorias de Erros do Modelo SAE J2450

A TAUS desenvolveu o sistema de métricas DQF-MQM (Dynamic Quality Framework – Multidimensional Quality Metrics) (ver Anexo 2) que aperfeiçoou os

sistemas existentes pois apresenta-se como uma planilha dinâmica, ou seja, que considera diversas variáveis, tais como o tipo de conteúdo, a função comunicativa, os requisitos do usuário final, o contexto e a qualidade, a qual é considerada dinâmica, uma vez que os requisitos de qualidade de tradução mudam consoante o tipo de conteúdo, a finalidade do texto e o público-alvo (Meer et al., 2017, p. 12). A TAUS realizou uma análise no 1º trimestre de 2011 para rever os modelos de avaliação existentes (O'Brien et al., 2011) e isto mostrou que os modelos de QA existentes são relativamente rígidos. Para a maioria, as categorias de erro, sanções aplicadas e limiares de aprovação/reprovação são as mesmas, independentemente dos parâmetros de comunicação envolvidos. Os modelos são também de natureza tão detalhada que a sua aplicação é morosa e a avaliação só pode ser feita para uma pequena amostra de palavras. A ferramenta de tipologia de erros TAUS DQF-MQM oferece uma forma padronizada de categorizar e contar os erros de tradução, utilizando os critérios habitualmente utilizados na indústria para a exatidão, linguagem, terminologia, estilo e padrões consoante a cada país. O sistema métrico TAUS DQF tem oito categorias de erro divididos em subcategorias e cinco níveis de severidade, conforme tabela abaixo:

<b>Tipo de Erro</b>	<b>Subcategoria</b>	<b>Tipo de Erro</b>	<b>Subcategoria</b>
Precisão	Adição Omissão Erro de tradução Tradução em excesso Tradução em falta Texto não traduzido Correspondência de MT incorreta	Design	Comprimento Formatação local Markup Texto em falta Truncagem/expansão de texto
Fluência	Pontuação Ortografia Gramática Registro gramatical Inconsistência Link/cruzamento de referência Codificação de caracteres	Convenção	Formato do endereço Formato da data Formato da moeda Formato de medição Tecla de atalho Formato telefônico
Terminologia	Inapropriada Estilo da empresa Estilo incoerente Estilo terceirizado Unidiomático <sup>1</sup>	Veracidade	Referência específica da cultura
		Outros	Quaisquer outras questões

Tabela 4 – Categorias de Erros do modelo TAUS DQF-MQM

<sup>1</sup> O conteúdo é gramatical, mas não idiomático.

<b>Níveis de Severidade</b>
Crítico
Grave
Leve
Neutro
Kudos <sup>2</sup>

Tabela 5 – Níveis de Severidade do modelo TAUS DQF-MQM

Nesse contexto, a TAUS também fornece uma série de diretrizes no sentido de orientar o pós-editor referente às diferenças entre a pós-edição ligeira e a pós-edição completa, conforme tabela abaixo:

<b>Pós-edição ligeira</b>	<b>Pós-edição completa</b>
Visar uma tradução semanticamente correta.	Visar uma tradução correta do ponto de vista gramatical, sintático e semântico.  Assegurar que a terminologia seja traduzida corretamente e que os termos não traduzidos pertençam à lista de “termos a não traduzir” do cliente.
Certificar-se de que nenhuma informação tenha sido adicionada ou omitida acidentalmente.	Garantir que nenhuma informação tenha sido adicionada ou omitida acidentalmente.
Editar qualquer conteúdo ofensivo, inapropriado ou culturalmente inaceitável.	Editar qualquer conteúdo ofensivo, inapropriado ou culturalmente inaceitável.
Usar o máximo da produção bruta de TA o mais possível.	Usar o máximo possível da produção bruta de TA.
Aplicar as regras básicas de ortografia	Aplicar as regras básicas de ortografia, pontuação e hifenização.
Não há necessidade de implementar correções que são de natureza apenas estilística.	
Não há necessidade de reestruturar as frases somente para melhorar o fluxo natural de o texto.	
	Assegurar-se de que a formatação esteja correta.

Tabela 6 – As diretrizes de pós-edição do TAUS

De acordo com Görög (2014), o sistema de métricas TAUS DQF-MQM consiste em uma base de conhecimentos sobre avaliação da qualidade composta pelas melhores prática, relatórios, modelos e uma série de ferramentas que permite a avaliação de tradutores tanto humana quanto automática. Além disso, a plataforma permite a redução

<sup>2</sup> A autora deste trabalho optou por não traduzir essa expressão, a qual, de acordo com o contexto, pode significar o respeito, o elogio, o enaltecimento ou o reconhecimento por um feito.

da subjetividade da opinião do revisor por meio de um fluxo de trabalho padronizado e oferece a possibilidade de comparação de traduções de projetos anteriores através do armazenamento de dados.

Portanto, parafraseando Pym (2011) sobre a noção de “alta qualidade”, assim como a beleza está nos olhos de quem vê, a qualidade está nos olhos do responsável pela pós-edição dos textos.

### **3. O Estágio Profissional**

O estágio profissional trata-se da aplicação prática de caráter pedagógico com o propósito de promover a aquisição de competências profissionais e o desenvolvimento de habilidades, hábitos e atitudes que contribuem com a formação profissional.

A escolha pela conclusão do curso de mestrado por meio da realização do estágio curricular surgiu antes mesmo da minha chegada a Portugal. Veio do desejo de adquirir o conhecimento de como é o funcionamento de uma empresa de prestação de serviços linguísticos e como se dá a sua atuação no mercado da tradução.

#### **3.1 A Entidade de Acolhimento**

O estágio curricular do mestrado em Tradução foi realizado na empresa Quicksilver Translate por meio de seu escritório situado em Sant Cugat del Vallès, Barcelona. O contrato, cuja duração inicial seria de três meses, foi ampliado para o total de cinco meses, com início em março e término em julho de 2020.

A Quicksilver Translate é uma empresa fundada no ano de 2004 por Colin e Stephen Whiteley, pai e filho, respectivamente. A empresa possui três escritórios, um em Espanha, Barcelona, um nos Estados Unidos da América, em Miami e o outro no Reino Unido, na cidade de Londres. A partir do ano de 2006 obteve a certificação ISO 9001 que garante a qualidade das traduções e procedimentos desenvolvidos pela empresa.

O escritório localiza-se na região metropolitana de Barcelona e emprega nove pessoas localmente, sendo duas funcionárias responsáveis pela gestão administrativa e financeira, as quais são diretamente supervisionadas por Stephen (CEO da empresa), além de cinco funcionários, sendo uma diretora de operações e quatro gestores de projetos, diretamente supervisionados por Colin (presidente da empresa). A empresa trabalha com um extenso banco de tradutores *freelancers* espalhados ao redor do mundo que são contactados de acordo com as demandas dos projetos. Além da gestão dos tradutores, os gestores de projeto são os responsáveis por todo o processo de atendimento aos clientes desde sua captação até o suporte após a realização dos

projetos, inclusive revisam os trabalhos concluídos pelos tradutores e até mesmo realizam traduções, conforme necessário.

Com o intuito de tornar os processos internos conhecidos por todos os integrantes da equipa, eram realizadas duas reuniões semanais, uma para o alinhamento e partilha de questões administrativas e financeiras, e outra reunião para a realização de formações rápidas com o objetivo de disseminar conhecimentos que pudessem otimizar o trabalho cotidiano, tais como conhecimentos avançados em Excel, por exemplo, os quais eram ministrados por membros da própria equipa.

Os trabalhos destinados a mim eram encaminhados por e-mail contendo informações preliminares sobre a realização dos mesmos e, geralmente, acompanhados por ficheiros em PDF para referência. Caso fosse necessário, o solicitante conversava diretamente comigo para sanar possíveis dúvidas ou prestar esclarecimentos. A ferramenta de tradução adotada pela empresa é o memoQ, portanto antes de enviar o projeto para os tradutores, os gestores preparam os documentos e os inserem na ferramenta a fim de que todos possam acessar os respectivos documentos e realizar a tradução nos pares linguísticos aos quais foram designados.

O estágio teve início no escritório da empresa, porém uma semana e meia depois, o trabalho local foi substituído totalmente pelo modelo a distância por conta do início do confinamento em toda a Espanha. Dessa forma, as reuniões e formações semanais, além das orientações sobre a realização dos trabalhos, migraram totalmente do presencial para o virtual.

### **3.2 As Tarefas Desenvolvidas no Estágio Profissional**

Acredito que a função mais importante do estágio é possibilitar a visão geral do mercado de trabalho e, dessa forma, estar à disposição para desenvolver todo o tipo de serviços linguísticos que a empresa se propõe a realizar. Além das traduções, desenvolvi trabalhos de pós-edição, digitação, formatação, elaboração/gestão de glossários, criação de memórias de tradução e legendagem, conforme a tabela seguinte:

<b>Atividades Desenvolvidas</b>	
Tradução	58 projetos
Glossário	26 atualizações
Memória de Tradução	16 bases terminológicas
Formatação	5 ficheiros
Digitação	4 ficheiros
Pós-edição	2 projetos
Legendagem	1 vídeo

Tabela 7 – Atividades desenvolvidas no estágio profissional

Importante ressaltar que não houve a assinatura do termo de confidencialidade, mas todos os e-mails continham a seguinte informação: “AVISO DE CONFIDENCIALIDADE - Esta mensagem e seus arquivos são destinados exclusivamente ao destinatário, e podem conter informações confidenciais sujeitas ao sigilo profissional. Ele não pode ser reproduzido ou distribuído sem a permissão expressa da QUICKSILVER TRANSLATE, SL. Se você não for o destinatário final, por favor, exclua-o e informe-nos”.

Portanto, as imagens relacionadas aos projetos de tradução não serão inseridas no trabalho visando a proteção da imagem e dos dados dos clientes, de acordo com o regulamento geral de proteção de dados em vigor.

### **3.2.1 Tradução**

Conforme dados acima, a atividade que realizei com mais frequência foram as traduções. As línguas de trabalho foram o EN>PT e o ES>PT, com o total de 39 887 palavras, sendo 76% em língua inglesa e 24% em língua espanhola. Cerca de 95% das traduções tiveram como língua de chegada o português de variante europeia, o que de certa forma foi considerado por mim como um desafio e um incentivo à pesquisa no sentido de refinar meus conhecimentos linguísticos da variante brasileira para a variante europeia. A contagem de palavras a traduzir era feita com base na “*low word count*”, ou seja, a contagem de palavras que não incluía as repetições e os segmentos que já estavam em 100% e 101% que ocorrem após a inserção dos projetos no memoQ por meio da

associação à memória de tradução do respectivo cliente. Os projetos de tradução foram desenvolvidos para empresas das seguintes áreas de atuação: indústria alimentícia para humanos e para pets, indústria cosmética, indústria farmacêutica, tecnologia da informação, documentos históricos (museu), hotelaria, estacionamento, jardinagem e paisagismo, material para trabalhos manuais, confecção de roupas e editora de livros.

A situação pandêmica impactou significativamente a demanda de traduções, pois cerca de 60% dos projetos realizados tinham como objetivo a divulgação de informações relativas aos cuidados sanitários que o momento exigia e os setores que demandaram mais trabalhos nesse sentido foram o setor hoteleiro, a indústria farmacêutica e o setor de estacionamento, respectivamente.

Tendo em conta as fases do processo de prestação de serviços de tradução, de acordo com Gouadec (2007), enquadro-me na fase da tradução, especificamente na etapa da transferência. Para a realização das traduções contava com a MT e base terminológica de cada cliente, o uso de dicionários especializados, a consulta a websites de empresas do mesmo ramo de atividade, bem como o uso dos motores de tradução DeepL e Google. O procedimento de pré-tradução era totalmente realizado pelos gestores de projetos. A fase de pós-tradução era realizada por mim, num primeiro momento, através da revisão do próprio trabalho e da aplicação das ferramentas de controle de qualidade do memoQ, nomeadamente QA e LQA, em seguida o texto era revisado por profissionais externos e a finalização do projeto era de responsabilidade dos gestores de projetos. Segue abaixo a descrição de cada trabalho de tradução realizado.

Tomando como base o tempo e a técnica aplicada, o grau de esforço cognitivo para a tradução dos projetos foi em sua maioria baixo, tendo em conta que se tratava de clientes regulares que solicitavam traduções com frequência e que já possuíam uma MT consolidada. As traduções que apresentaram grau de esforço cognitivo médio ou alto foram os projetos de clientes novos ou com pouca demanda de serviços.

Em caso de dúvida sobre o uso de novos termos, as consultas eram realizadas em websites de empresas do mesmo ramo de atuação dos clientes, em glossários disponibilizados na internet, além da consulta a dicionários online como o Infopédia (<http://www.infopedia.pt>) e o Priberam (<http://dicionario.priberam.org>), bem como os websites de dicionários monolíngues como o da Real Academia Española (<http://www.rae.es>) e o dicionário Collins (<http://www.collinsdictionary.com>). Na

ocorrência de alguma dúvida quanto aos termos relacionados à pandemia, além da pesquisa habitual em websites de referência como o da Direção Geral de Saúde de Portugal (<http://www.dgs.pt>), os gerentes de contas intermediavam o contato com os clientes a fim de que chegássemos a melhor solução no sentido de atendê-los em suas especificidades.

### **Setor Hoteleiro**

A maior demanda de projetos de tradução, vinte e oito ao todo, tratava-se da atualização de informações presentes nas fichas técnicas de uma empresa do setor hoteleiro. Os ficheiros incluíam basicamente informações sobre a localização do hotel, dados sobre o aeroporto mais próximo, distância dos principais pontos turísticos, coordenadas geográficas e mapa. O informativo também incluía a descrição dos arredores do complexo hoteleiro, bem como a quantidade e variedade de opções de quartos, restaurantes, bares e cafeterias, piscinas, quadras poliesportivas, serviços exclusivos de cuidados pessoais, atividades específicas para crianças, disponibilidade de acesso à internet e rede wi-fi, entre outros.

Além das informações descritas acima, as quais já fazem parte dos projetos regulares da empresa, a rede hoteleira solicitou a tradução de comunicados com o intuito de informar aos funcionários e aos clientes sobre as medidas sanitárias adotadas pelos hotéis, bem como das medidas quanto às alterações de datas ou cancelamentos de estadias devido ao impacto causado pela pandemia. Os pares linguísticos dos projetos desse cliente tinham como LP o espanhol e, além das traduções para a língua portuguesa, contemplava traduções para o inglês, o catalão, o italiano, o alemão, o francês e o russo.

Por se tratar de uma grande rede de hotéis, havia termos que não poderiam ser traduzidos, como o nome dos hotéis, endereços, nomes dos serviços oferecidos e nomes dos bares e restaurantes. Todos os projetos desse cliente tinham como destino a publicação no website da rede, bem como a impressão de cartazes para a afixação nas áreas comuns dos hotéis, além da disponibilização de folhetos no setor de recepção e nos quartos.

Quando havia a necessidade da pesquisa por algum termo específico, as buscas eram feitas em websites de empresas do ramo hoteleiro para que fosse feita uma comparação em termos de aproximação temática tais como: Nau Hotels & Resorts

(<http://www.nauhotels.com>) e Anantara Vilamoura Algarve Resorts (<http://www.anatara-hotels.com>), além de consultar os websites que disponibilizam dicionários online, tais como o da Real Academia Española, Infopédia e Priberam, quando se tratava da busca por informações genéricas.

### **Indústria Farmacêutica**

O setor que se segue na quantidade decrescente de projetos de tradução, foi a indústria farmacêutica, com o desenvolvimento de cinco projetos, sendo todos eles ligados à demanda gerada pela pandemia. Trata-se de um único cliente cujos projetos foram relacionados à tradução de textos instrutivos quanto ao uso dos produtos, nomeadamente, máscara para proteção do rosto infantil e adulto, gel higienizante para as mãos, toalhas higienizantes para as mãos e superfícies, além de advertências quanto ao uso indevido dos produtos, possíveis efeitos adversos e cuidados quanto ao uso do produto por parte das crianças pequenas.

Os pares linguísticos dos projetos desse cliente tinham como LP o espanhol e LC para a língua portuguesa, inglesa, italiana, norueguesa, francesa, polaca, finlandesa, dinamarquesa, árabe, alemã, grega, sueca e holandesa. Os termos utilizados eram de fácil compreensão, visto que tinham como objetivo a instrução para o cliente final da marca, dessa forma não poderia apresentar linguagem demasiado técnica, além disso a empresa já possuía uma MT consolidada com os principais termos utilizados para o tipo de tradução solicitada. As traduções foram inseridas na embalagem do produto.

### **Setor de Estacionamento**

Na sequência, o setor de estacionamento surge como o próximo na lista com o total de cinco projetos, além de solicitações avulsas contendo pequenos parágrafos ou frases. Todos os projetos e solicitações foram gerados devido a questões ligadas à pandemia, pois a empresa foi levada a modificar seus procedimentos de atendimento aos clientes em seus parques de estacionamento, visto que o contato direto com o cliente exigiu modificações por parte de funcionários e dos próprios clientes. A temática dos textos era sobre instruções quanto às novas normas de higiene e proteção no recebimento e entrega dos veículos, das novas diretrizes para a realização dos pagamentos, além da emissão de comunicados de agradecimento pela dedicação dos funcionários em um momento de

fragilidade e incertezas, e foram emitidos comunicados de pesar pela perda de funcionários e familiares por conta da COVID.

Os pares linguísticos dos projetos desse cliente tinham o espanhol como LP e, além das traduções para a língua portuguesa, contemplava traduções para o italiano e o alemão. A MT dessa empresa não possuía termos especificamente técnicos pois os textos gerados tinham por base informação generalista e quando havia a necessidade de busca por termos específicos, eram consultados os seguintes dicionários: Real Academia Española, Infopédia e Priberam. As traduções foram disponibilizadas no website da empresa por meio de janelas *pop-up*, através de cartazes afixados nas instalações da empresa e, no caso dos comunicados aos funcionários, foram enviados para os e-mails de todos os colaboradores.

### **Indústria Alimentícia para Pets**

A empresa que se segue desenvolve seus serviços no ramo da indústria alimentícia para pets (cães e gatos) e solicitou a abertura de quatro projetos de tradução, sendo divididos em folhetos informativos dos produtos sobre nutrição dos pets e uma apresentação do Powerpoint sobre uma reunião a ser realizada com os promotores de vendas da marca e colaboradores distribuídos em várias regiões de Portugal para a promoção de um novo produto.

O par linguístico dos projetos desse cliente tinham o espanhol como LP e somente o português como LC. O público-alvo dos folhetos de divulgação dos produtos era composto por tutores de pets, pessoas que geralmente não possuem conhecimentos técnicos sobre os cuidados com os animais, portanto as informações precisavam manter o caráter instrutivo, porém com uma linguagem apelativa que os fizesse comprar os produtos. Com relação à tradução do material para a apresentação aos promotores de vendas e colaboradores, o ficheiro tinha o conteúdo técnico que necessitava de uma pesquisa mais apurada no website da própria empresa e em websites de empresas que atuam na mesma área como Purina (<http://www.purina.pt>) ou Pétis (<http://www.petis.pt>), além da busca em glossários e dicionários especializados na área veterinária, tais como o Atlas de Anatomia Patológica e Veterinária da Universidade de Lisboa (<http://fmv.ulisboa.pt>) e o Dicionário de Termos Técnico Científicos de Medicina Veterinária - Editora Bio (<http://bibliotecaagptea.org.br>). A temática girava em torno da

nutrição dos pets com uma descrição detalhada da fabricação dos alimentos e da atuação de cada nutriente no organismo dos animais. Um dos temas abordados foi sobre a importância da oferta de uma alimentação equilibrada com o objetivo de prover a manutenção da saúde dos pets. A fabricação dos produtos é elaborada com ingredientes selecionados visando a ingestão de proteínas que proporcionam músculos fortes, de ácidos graxos que auxiliam na saúde da pele e do pelo dos animais, de vitaminas e minerais que proporcionam o fortalecimento de dentes e ossos, e da ingestão de fibras que auxiliam no processo de digestão dos animais. Também abordava questões de marketing e logística envolvidos no processo de distribuição dos produtos em todo o território português.

As traduções foram disponibilizadas no website do cliente, através da distribuição de folhetos em lojas especializadas em comercialização de produtos para pets e o ficheiro do Powerpoint foi apresentado nas reuniões regionais com os promotores de vendas e colaboradores.

### **Indústria Cosmética**

Os projetos desenvolvidos para a indústria cosmética foram solicitados por três empresas distintas, as quais chamarei de empresa A, B e C. A empresa A solicitou a tradução de informativos sobre os benefícios do uso de seus produtos de cuidado facial, nomeadamente um sérum e uma máscara faciais. A empresa B solicitou a tradução de componentes das fórmulas e modo de uso de produtos para maquilhagem. A empresa C solicitou a tradução da ficha técnica de um dos produtos da empresa, o óleo de semente de uva.

Os pares linguísticos dos projetos solicitados pela empresa A tinham como LP o espanhol e como LC o português, o inglês, o árabe, o italiano, o russo, o hebraico, o chinês, o francês, o coreano, o grego e o alemão. Por se tratar de um cliente novo a MT foi criada para a realização desse projeto. Os textos apresentavam informações sobre as instruções de uso, bem como informações técnicas sobre a composição dos produtos, o que impactou na produtividade por conta do tempo empregado na pesquisa dos termos e demandou a pesquisa dos termos em websites de empresas do mesmo ramo de atuação e em websites especializados, tais como o Clube da Química (<https://clubedaquimica.com/glossario-de-quimica/>) e Google Acadêmico com foco em artigos da área química e cosmética.

Os pares linguísticos dos projetos solicitados pela empresa B tinham como LP o espanhol e como LC o português, o inglês e o francês. A MT desse cliente já estava consolidada com os principais termos, o que impactou positivamente na questão da produtividade na realização dessa atividade.

O projeto solicitado pela empresa C, consistia na tradução de informações descritas na ficha da produção do óleo de semente de uva e eram de caráter estritamente técnico, tais como: identificação da substância/mistura e empresa, identificação de perigos no manuseio, composição química e ingredientes, primeiros socorros, medidas de combate a incêndios, medidas em caso de libertação acidental, manuseamento e armazenamento, controles de exposição/proteção individual, propriedades físicas e químicas, estabilidade e reatividade, informação toxicológica, informação ecológica, considerações sobre o descarte do produto, informações sobre cuidados com o transporte, informação regulatória, entre outras.

Os pares linguísticos do projeto da empresa C tinham como LP o inglês e o português como LC. Foi necessária a realização de pesquisa de terminologia especializada devido à gama de termos técnicos que necessitaram ser inseridos na MT, especialmente por se tratar de um cliente novo para a empresa. As buscas pelos termos foi realizada nos mesmos websites onde foram realizadas as consultas para a tradução dos textos das empresas A e B.

Os projetos das empresas A e B tinham como destino a divulgação dos produtos nos respectivos websites, bem como nas embalagens dos produtos. O projeto da empresa C tinha como objetivo a divulgação em uma área restrita do website destinada à venda de matéria prima para a empresas da área cosmética.

### **Jardinagem e Paisagismo**

Os projetos desenvolvidos para o setor de jardinagem e paisagismo foram solicitados por duas empresas distintas, as quais chamarei de D e E.

A empresa D solicitou a realização de um projeto para a tradução de um catálogo contendo a ficha técnica de toda a sua gama de produtos. Os pares linguísticos do projeto tinham como LP o espanhol e como LC o português e o francês. As traduções foram

solicitadas para a disponibilização no website da marca e para a impressão de catálogos para a distribuição aos revendedores.

A empresa E solicitou a realização de tradução de um folheto explicativo sobre o uso de sachês de etileno utilizados para a conservação de frutas delicadas no transporte de longas distâncias e mantas para a proteção no transporte de frutas mais resistentes por meio da acomodação em cestos e pallets. A empresa já possuía uma MT consolidada, o que permitiu que o processo de tradução fosse realizado sem a necessidade de pesquisa mais extensa. As traduções foram disponibilizadas no website da empresa.

### **Indústria Alimentícia**

O projeto desenvolvido para a indústria alimentícia foi solicitado somente por uma empresa e consistia na tradução de informações contidas na embalagem de bombons da marca, a qual tinha por objetivo a disponibilização de informação nutricional, ingredientes, nome e endereço da empresa, peso do produto, entre outros.

Os pares linguísticos do projeto tinham como LP o espanhol e LC o português, o inglês, o holandês, o polonês, o alemão e o grego. Embora esse fosse o primeiro projeto solicitado por esse cliente, o processo de tradução foi simples por se tratar de informação fartamente disponibilizada em websites de empresas alimentícias, tais como Imperial (<http://www.imperial.pt>) e Chocolates Regina (<http://www.chocolatesregina.pt>).

### **Documentos Históricos**

O projeto em questão foi desenvolvido por solicitação de uma instituição oficial espanhola com o intuito de enriquecer o acervo virtual do website do cliente. Foi desenvolvido um projeto que consistia na tradução de pequenos textos introdutórios que apresentavam o resumo de textos mais extensos com informação completa sobre as temáticas abordadas. Os pares linguísticos desse projeto tinham como LP o espanhol e como LC o português, o inglês, o italiano e o francês.

O texto histórico se caracteriza por uma leitura fluída e de fácil compreensão por se tratar da representação descritiva de factos pertencentes à cultura geral, porém foi realizado um trabalho de pesquisa em websites de instituições históricas tais como Sociedade Histórica da Independência de Portugal (<http://www.ship.pt>) e Rede

Portuguesa de Museus (<http://museusportugal.org>). Para o auxílio à tradução consultei os seguintes websites: Real Academia Española, Infopédia e Priberam.

### **Editora de Livros**

O projeto desenvolvido para a editora de livros consistiu na tradução de um post do Facebook que se tratava do lançamento de um livro sobre as aves brasileiras. Os pares linguísticos desse projeto tinham como LP o espanhol e LC o português com variante brasileira.

Por se tratar de um texto informal, característico da linguagem adotada nas redes sociais, a tradução não exigiu uma pesquisa mais detalhada, além do que, a tradução na variante brasileira foi um facilitador para a realização do trabalho por ser minha variante de domínio por ser originária do Brasil. Foi criada uma MT pois esse foi o primeiro trabalho realizado para o cliente. Os websites consultados para o apoio à tradução foram o da Real Academia Española (<http://www.rae.es>) e o Word Reference (<http://www.wordreference.com>).

### **Material para Trabalhos Manuais**

O projeto foi desenvolvido para uma empresa de material escolar para o desenvolvimento infantil por meio de trabalhos manuais e consistia na tradução do texto contido na embalagem do lápis de cera colorido. Os pares linguísticos desse projeto tinham o espanhol como LP e o português como LC.

A MT do cliente ainda não estava consolidada, visto que se tratava de um cliente novo. A tradução exigiu a realização de pesquisa em websites de empresas com atuação na indústria da fabricação de material para trabalhos manuais, tais como a Giotto (<http://www.fila.it/pt/pt>) e o dicionário Infopédia.

### **Confecção de Roupas**

O projeto desenvolvido para uma confecção de roupas e consistia na tradução de um texto promocional sobre a coleção primavera-verão 2020 para a divulgação no website do cliente. Os termos utilizados se tratava de informação generalista com relação à montagem de looks casuais para o público feminino. O texto apresentava informação sobre a origem e composição dos tecidos, sobre o respeito à natureza na confecção das peças, pois a empresa afirma sua responsabilidade social através do uso de tecidos

orgânicos como o algodão, o linho e a viscose ecológica e fala sobre como uma composição de peças equilibrada pode influenciar não só na questão estética como no bem-estar emocional das consumidoras. Os pares linguísticos desse projeto tinham como LP o espanhol e LC o português.

O texto privilegiava a linguagem apelativa, que tem como ponto principal o convencimento para a ação, nesse caso a ação de compra dos produtos por parte do público consumidor. As consultas para a tradução foram feitas nos websites dos dicionários Infopédia, além da consulta a websites de empresas do mesmo ramo de atuação como a Mango (<http://www.mango.pt>) e Vilanova (<http://www.vilanova.com>).

### **Tecnologia da Informação**

O projeto desenvolvido para uma empresa de tecnologia da informação consistia na tradução da política de privacidade para o website do cliente. Os pares linguísticos do projeto tinham como LP o inglês e LC o português. Este foi o maior projeto de tradução em termos de quantidade de palavras (7 988), visto que esse foi o primeiro trabalho solicitado pelo cliente, portanto a MT foi criada a partir da realização desse trabalho.

Houve a necessidade de pesquisa pormenorizada para a tradução de termos específicos da área da Tecnologia da Informação, portanto foram consultados o seguinte website: Glossário da Sociedade da Informação (<http://www.apdsi.pt>), empresas atuantes na mesma área como: Portugal Digital (<http://www.portugaldigital.gov.pt>) Made In Web (<http://www.madeinweb.pt>), além da consulta aos websites dos dicionários Infopédia e o Priberam.

### **3.2.2 Glossário**

A atualização de glossários foi a segunda atividade mais frequente que realizei no decorrer do estágio. A demanda partiu da necessidade de atualização de um glossário para a base terminológica dos projetos desenvolvidos para um cliente da indústria alimentícia voltada à nutrição de pets (cães e gatos). Os termos foram enviados em um ficheiro Excel dividido em colunas cujo título referia-se ao idioma em questão, conforme exemplo abaixo:

	Spanish	Portuguese	Portuguese	English	English	French	Italian
45	¿A partir de que edad recomienda Advance utilizar						A partire da quale età Advance consiglia di usare
46	¿Cuál de estas recetas contiene la textura						Quale di queste ricette presenta una textura
47	¿Cuál de estos packs es						Quale di questi pack è
48	¿Cuál de estos pouch es						Quale di queste pouch è
49	¿Cuál son los principales ingredientes del						Quali sono gli ingredienti principali di
50	¿Cuántos formatos existen en la gama						Quanti formati esistono nella gamma
51	¿En qué formatos encontramos						Quali sono i formati disponibili di
52	¿En qué gama de Natural Trainer Dog el ingrediente transversal es						In quale gamma di Natural Trainer Dog l'ingrediente trasversale è
53	¿Qué alimentación recomendarías para perros						Che alimentazione consiglieresti al proprietario di un cane
54	¿Qué dieta de Advance recomendarías para un gato						Quale dieta Advance consiglieresti per un gatto con
55	¿Qué especialidades encontramos						Qual è la gamma di specialità di

A atividade baseava-se na busca de termos na língua portuguesa, que fossem correspondentes na língua espanhola. A orientação do solicitante foi que a busca dos termos fosse realizada em dicionários especializados tais como os websites de empresas que atuam na mesma área como Purina (<http://www.purina.pt>) ou Pétis (<http://www.petis.pt>), além da busca em glossários e dicionários especializados na área veterinária, tais como o Atlas de Anatomia Patológica e Veterinária da Universidade de Lisboa (<http://fmv.ulisboa.pt>) e o Dicionário de Termos Técnico Científicos de Medicina Veterinária - Editora Bio (<http://bibliotecaagpatea.org.br>), além de utilizar a página multilíngue e o blog da empresa para consulta. A orientação era de que se o termo não fosse localizado a respectiva célula deveria ser mantida em branco.

Essa atividade se repetiu ao longo do estágio e foi realizada para a atualização dos glossários de cerca de vinte e cinco clientes com atuação nas mais diversas áreas do conhecimento. O processo de trabalho foi semelhante ao informado acima, com o recebimento do ficheiro em Excel e a busca de termos em dicionários especializados conforme a área de atuação das empresas, além do uso dos motores de tradução DeepL e Google.

### 3.2.3 Memória de Tradução

Foi solicitada a criação de dezesseis MT com o intuito de agrupar os pares linguísticos de maior demanda entre os clientes da Quicksilver, a fim de proporcionar uma base terminológica ampla para pesquisa e uso por parte dos gestores de projetos. As MT foram divididas em dois blocos, sendo oito pares linguísticos com base na língua

inglesa (EN-FR/IT/DE/PT e FR/IT/DE/PT-EN) e oito pares linguísticos com base na língua espanhola (ES-FR/IT/DE/PT e FR/IT/DE/PT-ES).

A figura abaixo exemplifica o processo de criação de uma memória de tradução, a qual tem como objetivo facilitar o processo de tradução ao favorecer a pesquisa de termos utilizados com frequência por parte dos clientes da empresa consoante a área de atuação.

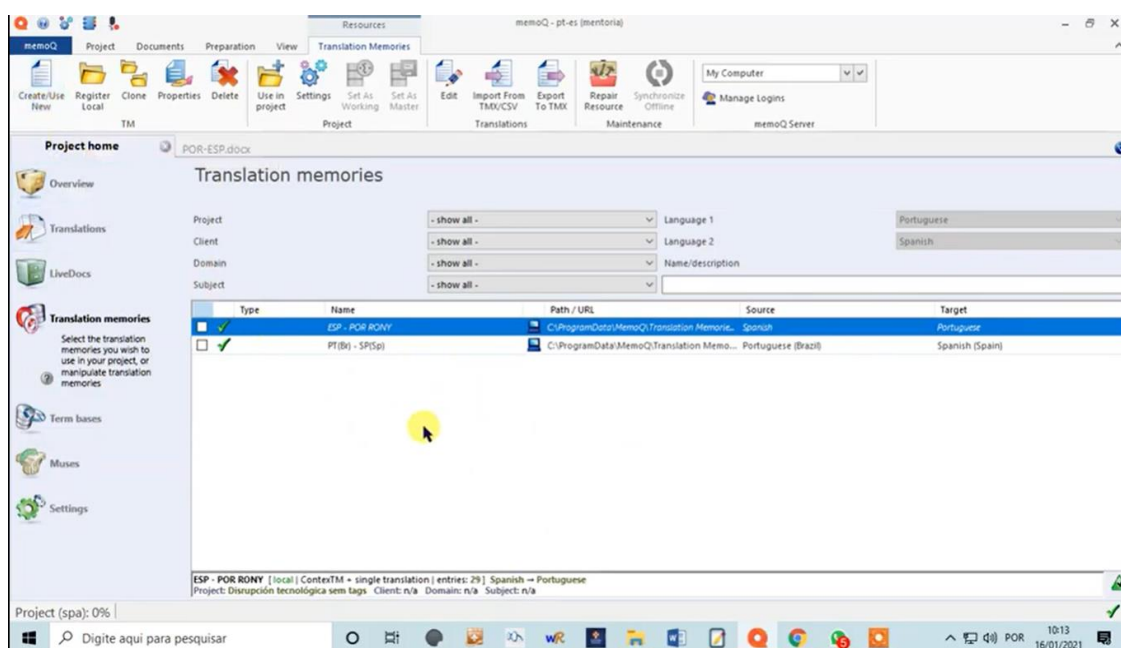


Figura 3 – Página de criação das MTs por pares linguísticos no memoQ

A criação das MTs foi um pedido que partiu dos gestores de projetos com o objetivo da consolidação de termos a fim de dar mais agilidade e consistência ao trabalho da equipe de tradutores dos projetos.

### 3.2.4 Formatação

O trabalho de formatação de ficheiros foi realizado para dois clientes distintos. O primeiro cliente era uma escola de negócios sediada em Barcelona, a qual enviou seis ficheiros do Excel juntamente com seis ficheiros do Powerpoint, onde cada planilha seria utilizada como script para a gravação de vídeos educativos. As planilhas em Excel (ver exemplo na figura 4) continham três colunas: a primeira já estava preenchida e tratava-se

das falas do narrador com o respectivo *time-code* para a inserção das falas gravadas, a segunda coluna continha a numeração dos slides que deveriam ser inseridos por mim no ficheiro para a orientação ao narrador a respeito de qual informação estava a ser transmitida no ecrã e a terceira coluna estava em branco para que o narrador pudesse fazer algum comentário que julgasse necessário.

CONTENIDO	Imagen de Slide	Comentarios
<b>VIDEO M3</b>		
[00:00:01] Module 3 and Module 4 belong to the SPDM content of What do I have.	1	
[00:00:08] Up to now what did we see? Last week we analyzed Knowing who am I, What do I offer and V/ho do I offer to. Remember? We saw Promise, Essence, Flame Red and then we also saw the Client Archetype and the Service Specifications and the Moments of Truth. So all that should be all set. Now we go into Part B. What do I have and What do I have, part of Module 3 and Module 4 has What do I know, What knowledge do I have? How to obtain resources from freeing up time and How to destroy waste. OK.	2	
[00:00:49] So now, in What do I have? Module 3 and Module 4. What we're going to see are the things that you have here in red and brown, How to obtain resources and What do I know in most basic operational components in Module 3, and How to destroy waste in Module 4.	2	
[00:01:11] So let's start with Module 3. Unlocking Capacity to Tackle Higher Value-Added Tasks.	4	
[00:01:20] What does it mean? It means that we are going to see how we free time. But before, first of all, we are going to analyze What do we know.		
[00:01:31] So let's start with that, because our trip is going to take us into two stops. From this port to the next port we're going to stop at the service mix that we have, Task equals Problem equals Knowledge, the Capacity Matrix, the Load Matrix, and the Task Benchmark.	5	
[00:01:54] You're going to have here two very, very nice Excel	6	

Figura 4 – Exemplo da planilha encaminhada para a realização da formatação

O segundo cliente era uma empresa multinacional especializada no desenvolvimento de equipamentos e substâncias químicas para a contenção e o combate a incêndios. O processo de formatação foi mesclado com o processo de revisão dos textos, pois se tratava das fichas técnicas de toda a gama de produtos fabricados pela empresa. O material já havia sido traduzido pelo próprio cliente, mas foi inserido no memoQ para que pudesse ser gerada a MT do cliente e para que possíveis correções pudessem ser feitas, sendo a LP o inglês e a LC o português.

A revisão textual foi feita a partir do memoQ e não houve mudanças significativas, somente adequações pontuais quanto aos termos técnicos consoante às nomenclaturas das substâncias químicas (gás carbônico, flúor, bicarbonato de potássio, bicarbonato de sódio) e às unidades de medidas (toneladas, libras, quilos, galões). As fichas técnicas em língua inglesa e em língua portuguesa, ambas em PDF, foram enviadas a mim em conjunto para que pudesse ser feita a comparação quanto à distribuição gráfica das

imagens e textos nas páginas para que o padrão adotado nos documentos originais fosse mantido nos documentos traduzidos. Os apontamentos quanto a possíveis alterações no documento traduzido foram feitos no Adobe Reader através da inserção de comentários em pontos específicos dos ficheiros. Essa atividade foi realizada em parceria com o gestor de projetos responsável pelas solicitações desse cliente, o qual tinha como encargo a conferência e a correção das alterações propostas.

### **3.2.5 Digitação**

Os trabalhos de digitação foram realizados a partir da digitalização de dois contratos empresariais que tinham como objetivo a venda de uma das propriedades de uma empresa do ramo jurídico, cujos documentos estavam escritos originalmente em alemão. A digitação dos ficheiros foi necessária pois a transformação do PDF para o Word não resultou conforme o esperado, especialmente com relação à formatação original e algumas partes do documento que estavam ilegíveis e desconfiguradas. Após a digitação os ficheiros foram inseridos no memoQ para que a tradução pudesse ser feita por um profissional designado pela gestora de projetos responsável pelo cliente.

A digitação é uma atividade que requer alto grau de atenção, pois mesmo parecendo uma atividade simples, a falta de atenção pode resultar na ausência de alguma parte do texto e impactar negativamente no resultado. No caso da digitação de um contrato os impactos podem gerar problemas maiores como a necessidade de aditamento de alguma cláusula ou até mesmo a sua anulação. Por esse motivo, além da atenção redobrada, é necessário que o trabalho seja revisado detalhadamente e com extremo cuidado.

### **3.2.6 Pós-Edição**

Antes de solicitarem que eu fizesse essa atividade, foi realizada uma reunião na qual um dos gestores de contas da Quicksilver partilhou dados a respeito da utilização cada vez mais crescente do uso dos motores de TA como suporte à realização dos serviços de tradução e falou a respeito dos avanços apresentados por essa ferramenta ao longo do tempo, bem como sobre as diretrizes fornecidas pela norma ISO 18587:2017 com relação

ao trabalho de pós-edição. Além disso apresentou à equipa o sistema de métricas TAUS DQF-MQM o qual imediatamente passou a integrar o processo de trabalho da agência em regime experimental.

O processo de pós-edição foi realizado a partir de textos de um cliente do ramo da nutrição de pets (cães e gatos) resultantes de TA a partir do Google, processo este conhecido por *Googlish*, que é o nome dado ao resultado da tradução de texto em língua estrangeira de qualquer um dos vários motores de tradução populares na web. Foi a partir da realização dessa atividade que decidi abordar o assunto da PE em meu relatório de estágio sobre a atuação do tradutor como pós-editor.

O trabalho foi realizado na ferramenta memoQ, no mesmo ecrã onde é realizado o processo de tradução, bastando para tal que o gestor de contas alterasse minha designação para revisora, como podemos ver na figura 5, e assim pude dar início à atividade.

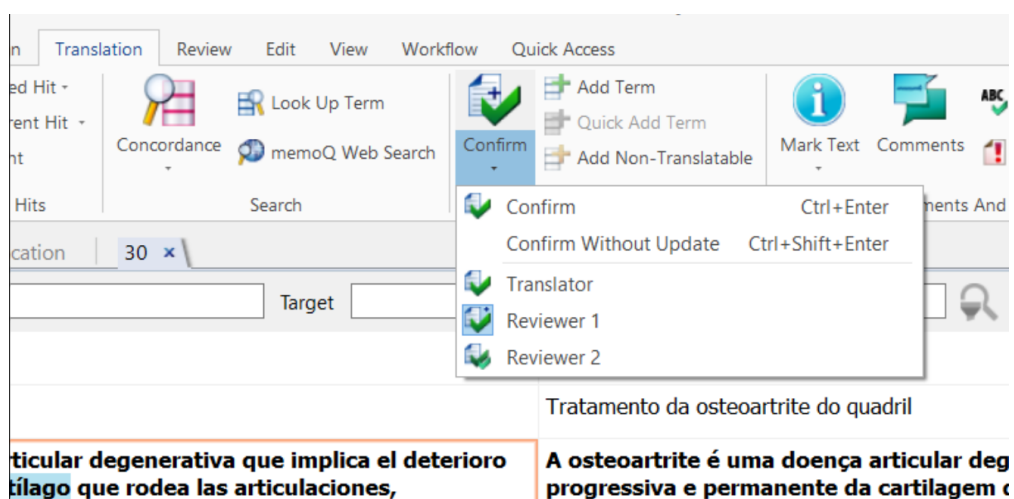


Figura 5 – Ecrã de designação de revisor no memoQ

O processo consiste na habilitação por parte dos gestores de contas da opção *Review* da guia *Confirm*, e a partir dessa liberação é permitido o acesso para a seleção de um trecho específico ou todo o segmento no intuito de classificá-lo de acordo com os parâmetros determinados pelo sistema de métricas TAUS DQF-MQM (anexo 2). Por exemplo, se um erro de fluência fosse detectado, era necessário definir em qual subcategoria o erro deveria ser classificado (pontuação, ortografia, registro gramatical, sintaxe, inconsistência, uso incorreto de ligações para acessos a páginas web relacionados

ao texto e codificação de caracteres) e, em seguida deveria ser definida qual a severidade do erro tendo em conta o impacto no texto como um todo.

Após a finalização da análise dos segmentos é emitido um relatório (figura 6) que apresenta o impacto da severidade dos erros consoante a análise global do texto, o qual demonstra se a quantidade de erros está acima ou abaixo do limite aceitável. A partir da análise desse relatório será definido se o processo tradutório foi satisfatório ou não de acordo com a necessidade do projeto/cliente em questão.

Severity level	PE 1	PE 2	PE 3	PE 4	PE 5	PE 6	PE 7	PE 8	PE 9	PE 10
major	6	29		2	39	21	24	100	7	14
minor		40	21	24	77	37	39	185	59	35
Total	6	69	21	26	116	58	63	285	66	49
	0.20%	2.28%	0.70%	0.86%	3.84%	1.92%	2.09%	9.43%	2.18%	1.62%

Figura 6 – Exemplo de relatório de severidade TAUS emitido pelo memoQ

A PE é um serviço de análise, revisão e correção de erros textuais que, além da experiência em tradução e revisão exige que o profissional busque aprimoramento constante de seus conhecimentos linguísticos.

O ponto mais desafiador no processo da pós-edição foi a definição do nível de severidade mais adequado, tendo em conta que, embora haja uma tabela explicativa dos fatores que devem ser observados para que tal definição seja adotada (ver anexo 1) a experiência profissional e as habilidades de leitura e revisão de textos são uma mais valia no sentido de agregar agilidade e precisão no desenvolvimento dessa atividade.

### 3.2.7 Legendagem

O trabalho de legendagem foi solicitado por uma empresa multinacional do ramo farmacêutico. O vídeo tinha cerca de 2'40" de duração e o conteúdo era sobre o desempenho da empresa no primeiro trimestre de 2020 e as estratégias para vencer os desafios impostos pela pandemia. Para a realização do processo de legendagem utilizei o software Subtitle Workshop e o auxílio do motor de TA DeepL e dos dicionários online Infopédia e Priberam, além do website do Glossário de Termos Econômicos – PUC/SP (<http://www.portalbrasil.net/economia-glossario.htm>). A LP da legenda era o inglês e a LC

era o português com variante brasileira. Assim como o trabalho realizado para a Editora de Livros, a tradução na variante brasileira foi um facilitador para a realização do trabalho por ser minha variante de domínio.

Os termos da área econômica exigiram a pesquisa no website do Portal Brasil ([http://portalbrasil.net/economia\\_glossario.htm](http://portalbrasil.net/economia_glossario.htm)) e no Word Reference.

#### 4. Reflexão crítica

A realização do estágio profissional decorreu com tranquilidade embora tenha acontecido em um dos momentos mais críticos dos últimos anos, o início da pandemia, é importante salientar que, apesar da convivência presencial ter sofrido uma quebra devido à situação de confinamento, senti-me parte da equipa como se estivesse no escritório. Um dos fatores que mais influenciaram no sentimento de pertença foi a questão da participação nas reuniões e mini formações semanais online, pois os encontros permitiram a oportunidade de conhecer melhor os colegas de trabalho, além do aprendizado proporcionado através da troca de experiências e de conhecimento.

Outro ponto importante foi que as solicitações para a realização dos trabalhos não se limitaram apenas a um dos gestores de projetos, portanto tive contato com todos os gestores da empresa e pude aprender com cada um o quanto o trabalho pode ser diverso e enriquecedor.

Desde o início do estágio fiz questão de saber o que era esperado na realização de cada projeto ou atividade e, ao término de cada trabalho, habituei-me a pedir o *feedback* de maneira a buscar o alinhamento das expectativas e a favorecer o diálogo em busca de externar minha disposição em absorver competências não só técnicas, como comportamentais e atitudinais frente aos desafios do mercado profissional.

O uso do memoQ para a realização dos projetos de tradução foi um facilitador pois já se tratava de uma ferramenta conhecida devido à apresentação na disciplina de Tecnologias de Apoio à Tradução. Embora o uso tenha sido limitado ao processo de tradução e pós-edição, pude ampliar meus conhecimentos e interagir com a gama de possibilidades oferecidas pelo software, como a ferramenta de controle de qualidade, por exemplo. O software permitiu com que pudesse organizar melhor meu ambiente de trabalho e tivesse clareza quanto à gestão dos projetos que estavam sob minha responsabilidade, de modo que pudesse estabelecer prioridades na realização das tarefas.

Antes do envio das atividades, os gestores entravam em contato para confirmar se já havia alguma tarefa em andamento e qual o prazo estipulado para a conclusão, para que não houvesse acúmulo ou conflito de atividades e para que os clientes fossem atendidos da melhor maneira. O mesmo procedimento era realizado caso surgissem trabalhos em regime de urgência. Quando o prazo estava próximo ao limite da entrega, geralmente um

dia ou algumas horas antes a depender do volume de trabalho, era enviado um e-mail automático gerado pelo próprio memoQ com um lembrete sobre o encerramento da atividade.

A abordagem da temática da PE permitiu com que tivesse acesso à uma área que não obteve relevância dentro do curso de mestrado, a qual chamou minha atenção justamente pela presença cada vez mais frequente no mercado de trabalho, conhecimento este que, provavelmente, não teria despertado o interesse caso não houvesse optado pelo estágio profissional.

É certo que a formação acadêmica privilegia o aperfeiçoamento do tradutor de maneira que o domínio dos pares linguísticos seja suficiente para que o profissional necessite minimamente de auxílio externo, afinal é importante ressaltar que o tradutor precisa ter conhecimento profundo sobre os idiomas de trabalho. Exatamente por esse motivo há tradutores que optam por trabalhar com áreas específicas do conhecimento, como a área do direito, da medicina ou da engenharia, por exemplo. Porém, com a demanda crescente do mercado da tradução e com o aperfeiçoamento das ferramentas TAC surgiu a necessidade de agregar mais habilidades à função do tradutor generalista, o que significa que o profissional precisa buscar a capacitação através de formações específicas para pós-editores ou formações que, além de privilegiar os estudos linguísticos, tenham como objetivo o estudo das ferramentas computacionais de apoio à tradução e, até mesmo, a linguística computacional.

Outro ponto importante, baseia-se no facto de que os estudos sobre a PE estão fortemente ligados ao desenvolvimento dos motores de TA, com o objetivo de trabalhar em favor do aperfeiçoamento dessas ferramentas por meio da avaliação criteriosa através do uso de métricas que possibilitam a leitura precisa dos pontos críticos da tradução realizada por máquina, visando sua evolução e integração com o trabalho humano, e essa é uma área relevante que impacta diretamente no cenário dos serviços de tradução na atualidade.

Embora o trabalho do tradutor seja realizado de maneira individual, as habilidades de comunicação e empatia são primordiais e foram essenciais no decorrer do estágio. O exercício da habilidade de comunicação facilitou o trabalho com os colegas e gestores, pois uma boa comunicação transmite confiança e coopera para a abertura ao diálogo que é capaz de resolver conflitos, evitar desentendimentos e expressar sentimentos. O

exercício da empatia representa a capacidade de compreensão mútua no ambiente profissional e foi primordial especialmente devido à situação em que todos estavam vivenciando momentos de tensão por conta da pandemia.

É importante salientar que as ferramentas de apoio continuam a ser apenas um auxílio para a realização do trabalho do tradutor. De facto, a principal e mais importante ferramenta de apoio ao tradutor é seu próprio corpo, suas capacidades cognitivas que são treinadas ao longo de anos de leitura, escrita e reescrita. O tradutor é forjado através do contato com vários tipos de textos, muita revisão, discussão e noites em claro. As ferramentas de tradução foram criadas e são aperfeiçoadas para ajudar o tradutor a realizar o seu trabalho com agilidade e rapidez, mas por mais avançadas que sejam necessitam da intervenção do tradutor para serem alimentadas e geridas.

O uso de motores de TA não é mais rejeitado como há poucos anos, pelo contrário, tem sido cada vez mais inserido no processo de tradução com o intuito de proporcionar, especialmente, o aumento da produtividade. Segundo a pesquisa ELIS 2022, as três ferramentas TAC mais utilizadas pelas empresas prestadoras de serviços de linguísticos são Trados, memoQ e Memsource e os três motores de tradução mais utilizados são DeepL, TA incorporada às ferramentas TAC e o Google Tradutor, o que demonstra que esses recursos se tornam cada vez mais integrados e multifuncionais. Nesse contexto, a formação especializada do tradutor se faz necessária não só para a preparação frente aos desafios do trabalho diário, mas também para atuar como integrante de equipas desenvolvedoras de novas ferramentas.

## Conclusão

A partir da experiência na realização do meu estágio profissional na empresa Quicksilver Translate tive a percepção de que o tradutor necessita de atualização constante frente aos avanços tecnológicos ocorridos especialmente nas últimas décadas, e por esse motivo, ficou clara a preocupação da empresa em inserir pequenas doses de partilha do conhecimento em sua agenda semanal. Estamos na era digital e já não cabe mais o questionamento sobre a pertinência do uso de ferramentas informáticas de apoio, pois essa tecnologia já faz parte do dia a dia do profissional da tradução.

Um dos pontos mais interessantes da realização do estágio foi a possibilidade de vivenciar o trabalho de um tradutor *freelancer* devido à situação de *home office* e o trabalho de um tradutor *in-house* devido à participação em todos os processos da agência de tradução como se estivesse presente, pois o contato era diário e a participação foi contínua desde o primeiro dia até ao término do estágio.

A análise do contexto histórico foi importante para perceber como surgiram as primeiras ideias e quais foram as impressões dos pioneiros dos estudos de tradução aplicada aos dispositivos de assistência à tradução, bem como de que maneira aconteceu o desenvolvimento dos equipamentos de tradução mecânica e qual a sua contribuição para a criação de dispositivos eletrônicos utilizados em larga escala atualmente.

É importante perceber o quanto o contexto histórico mundial impactou na realização de cada etapa do processo evolutivo para o desenvolvimento dos dispositivos de TA. O período compreendido entre a primeira e a segunda guerra mundial foi o ponto de partida para o início do processo, pois a necessidade da tradução de textos inimigos por parte dos norte-americanos originou a criação do dispositivo de segurança conhecido como criptografia, que posteriormente acabou por ser implementado ao processo de tradução e continua a ser utilizado ainda hoje.

É certo que o relatório ALPAC contribuiu com a estagnação das investigações apoiadas financeiramente pelo governo norte-americano na década de 1960, porém é interessante perceber que a falta do apoio governamental fez com que os estudos prosseguissem de forma consistente no âmbito educacional, o que despertou o interesse de empresários no desenvolvimento de dispositivos em larga escala.

Na sequência, a indústria apresentou um período de rápido crescimento devido à difusão de tecnologia para a tradução automática com o surgimento de inúmeras empresas especializadas no desenvolvimento de softwares de TAC, o que revolucionou o mercado das empresas prestadoras de serviços de tradução e dos utilizadores *freelancers* por começarem a oferecer produtos a preços mais competitivos. Esse também foi o período em que surgiram grandes empresas que acabaram por se tornar líderes na concepção e no fornecimento de softwares equipados com ferramentas desenvolvidas às necessidades dos clientes.

Com o uso cada vez mais próspero dos dispositivos de auxílio à tradução cenário mundial, a função do tradutor passa a ter outras características, como a necessidade da participação do tradutor como pós-editor. O pós-editor é o profissional que, por meio das ferramentas do controle de qualidade, nomeadamente as métricas, proporciona a supervisão técnica e aponta com precisão os erros em busca do aperfeiçoamento do processo como um todo.

Portanto, não há demérito na atuação do tradutor como pós-editor, pois essa faceta será cada vez mais necessária e, diria eu, imprescindível, pois trata-se de um profissional que agrupa, além das características fundamentais de um tradutor, a capacidade de atuar como o inspetor de qualidade da tradução, e essa função, só o olhar humano capaz de realizar.

O estágio profissional cumpriu o seu objetivo, pois além de possibilitar o envolvimento na realização de projetos de tradução, proporcionou o contato com os mais diversos tipos serviços oferecidos pelos prestadores de serviços linguísticos e, deste modo, fortaleceu o sentimento de capacitação mediante a gama de habilidades e técnicas desejadas para a realização de um trabalho qualificado que vai ao encontro das necessidades do mercado de trabalho cada vez mais exigente e competitivo.

## Bibliografia

- Allen, J. (2001). Postediting: An Integrated Part of a Translation Software Program. *Language International Magazine*, 26-29. Fonte: <http://www.oocities.org/mtpostediting/Allen-LI-article-Reverso.pdf>
- Allen, J. (2003). Post-editing. Em H. Somers, *Computers and Translation: A Translator's Guide* (pp. 297-317). Amsterdam: John Benjamins.
- ALPAC. (1996). *Language and Machines: Computers in Translation and Linguistics*. Washington, D.C.: The National Academies Press.
- Alves, F., Szpak, K. S., Gonçalves, J. L., Sekino, K., Aquino, M., Castro, R. A., . . . Mesa-Lao, B. (2016). Investigating Cognitive Effort in Postediting: A Relevance-Theoretical Approach. Em S. H.-S. Grucza, *Eyetracking and Applied Linguistics* (pp. 109-142). Berlin: Language Science Press.
- Arnold, D. (2003). Why Translation is Difficult for Computers. Em H. Somers, *Computers and Translation: A Translator's Guide* (pp. 119-142). Amsterdam: John Benjamins.
- Arnold, D., Balkan, L., Meijer, S., Humphreys, R. L., & Sadler, L. (1994). *Machine Translation: An Introductory Guide*. London: NCC Blackwell.
- Austermühl, F. (2001). *Electronic Tools for Translators*. Manchester: St. Jerome Publishing.
- Bielsa, E. (2022). *Routledge Handbook of Translation and Media*. Abingdon, New York: Routledge.
- Bowker, L., & Fisher, D. (2010). Computer-Aided Translation. Em Y. Gambier, & L. v. Doorslaer, *Handbook of Translation Studies* (pp. 60-65). Amsterdam: John Benjamins.
- Carl, M., & Way, A. (2003). *Recent Advances in Example-Based Machine Translation*. Dordrecht: Kluwer.
- Castells, M. (2000). *The Rise of Network Society*. Oxford: Blackwell.
- Depraetere, I. (2010). *What Counts as Useful Advice in a University Post-Editing Training Context?* St. Raphael, France.: 14th Annual Conference of European Association for Machine Translation.
- Dizdar, D. (2009). Translational Transitions: "Translation Proper" and Translation Studies in the Humanities". *Translation Studies*, 2, pp. 82-102. doi:10.1080/14781700802496274
- Fontes, H. L. (2013). Evaluating Machine Translation: Preliminary Findings from the First DGT-wide Translators' Survey. *Languages and Translations* 6, 10-11.

- Fonte:  
<http://ec.europa.eu/dgs/translation/publications/magazines/languagestranslation/>
- Forcada, M. L. (2010). Machine Translation Today. Em Y. Gambier, & L. v. Doorslaer, *Handbook of Translation Studies - Volume 1* (pp. 215-223). Amsterdam: John Benjamins.
- Forcada, M. L. (2017). Making Sense of Neural Machine Translation. Em D. Kenny, & J. Moorkens, *Translation Spaces - Volume 6* (pp. 291-309). Amsterdam: John Benjamins.
- Forcada, M. L., & Ñeco, R. P. (1997). Recursive Hetero-associative Memories for Translation. *International Work Conference on Artificial Neural Networks*, 453-462.
- Garcia, I., & Stevenson, V. (2005). Trados and the Evolution of Language Tools. *Multilingual Magazine*, 28-32.
- Garcia, I., & Stevenson, V. (2006). TRADOS and the Evolution of Language Tools. *Multilingual Magazine*, 24-27.
- Goaudec, D. (2007). *Translation as a Profession*. Amsterdam: John Benjamins.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. Cambridge: MIT Press Book.
- Google. (2016). *Google AI Blog*. Fonte: Googleblog: <https://ai.googleblog.com/2016/11/zero-shot-translation-with-googles.html>
- Görög, A. (2014). Quantifying and Benchmarking Quality: The TAUS Dynamic Quality Framework. *Revista Tradumàtica*, 443-454. Fonte: <https://revistes.uab.cat/tradumatica/article/view/n12-gorog2/pdf>
- Guzmán, R. (2007). Manual MT Post-editing: If It's Not Broken, Don't Fix It! *Translation Journal*. Fonte: <http://translationjournal.net/journal/42mt.htm>
- Hans P. Krings, G. S. (2001). *Repairing Texts: Empirical Investigations os Machine Translation Post-Editing Processes*. Ohio: Kent State University Press.
- Hutchins, W. J. (1986). *Machine Translation: Past, Present, Future*. New York: Halsted Press.
- Hutchins, W. J. (1996). ALPAC: The (In)Famous Report. *MT News Internacional*, 9-12. Fonte: <https://aclanthology.org/www.mt-archive.info/90/MTNI-1996-Hutchins.pdf>
- Hutchins, W. J. (1998). The Origins of The Translator's Workstation. *MT News Internacional*, 287-307.

- Hutchins, W. J., & Lovtskii, E. (2000). Petr Petrovich Troyanskii (1894–1950): A Forgotten Pioneer of Mechanical Translation. *Machine Translation*, 187-221. Fonte: <https://link.springer.com/article/10.1023/A:1011653602669>
- Hutchins, W. J., & Somers, H. (1992). *An Introduction to Machine Translation*. London: Academic Press.
- IBM. (2020). *IBM Cloud Education*. Fonte: <https://www.ibm.com/cloud/learn/natural-language-processing>
- ISO 17100. (2015). *Translation Services - Requirements for Translation Services*. Fonte: <https://www.iso.org/standard/59149.html>
- ISO 18587. (2017). *Translation Services - Requirements for Post-Editing of Machine Translation Output*. Fonte: International Organization for Standardization: <https://www.iso.org/standard/62970.html>
- Jakobsen, A. L., & Jensen, K. T. (2008). Eye Movement Behaviour Across Four Different Types of Reading Task. Em S. Göpferich, A. L. Jakobsen, & I. M. Mees, *Looking at eyes : Eye-Tracking Studies of Reading and Translation Processing* (pp. 103-124). Copenhagen: Copenhagen Studies in Language.
- Jakobson, R. (1976). *Linguística e Comunicação*. São Paulo: Cultrix.
- Kenny, D. (2022). *Machine Translation for Everyone - Empowering Users in the Age of Artificial Intelligence*. Berlin: Language Science Press.
- Koponen, M. (2016). Is Machine Translation Post-Editing Worth the Effort? A Survey of Research into Post-Editing and Effort. *The Journal of Specialised Translation*, Issue 25, pp. 131-148. Fonte: [https://www.jostrans.org/issue25/art\\_koponen.php](https://www.jostrans.org/issue25/art_koponen.php)
- MacLuhan, M. (1964). *Os Meios de Comunicação como Extensões do Homem*. São Paulo: Cultrix.
- Meer, J. V. (2017). Measuring Translation Quality: From Translation Quality Evaluation to Business Intelligence. *TAUS*, 12.
- Mirko Plitt, F. M. (2010). A Productivity Test of Statistical Machine Translation Post-Editing in a Typical Localisation Context. *The Prague Bulletin of Mathematical Linguistics* 93, 7-16. Fonte: <http://ufal.mff.cuni.cz/pbml/93/art-plitt-masselot.pdf>
- Mossop, B. (2020). *Revising and Editing for Translators*. Manchester: St. Jerome Publishing.
- O'Brien, S. (2002). *Teaching Post-Editing, a Proposal for Course Contents*. Manchester: 6th Annual Conference of European Association for Machine Translation.
- O'Brien, S. (2006). *Machine-Translatability and Post-Editing Effort: An Empirical Study Using Translog and Choice Network Analysis*. Dublin: Dublin City University.

- O'Brien, S., Balling, L. W., Carl, M., Simard, M., & Specia, L. (2014). *Postediting of Machine Translation: Processes and Applications*. Newcastle: Cambridge Scholars Publishing.
- O'Brien, S., Choudhury, R., Meer, J. V., & Monasterio, N. A. (2011). TAUS Diamic Quality Evaluation Framework. *TAUS Labs Report*.
- Peris, Á., Domingo, M., & Casacuberta, F. (2017). Interactive Neural Machine Translation. *Computer Speech & Language*, pp. 201-220. doi:<https://doi.org/10.1016/j.csl.2016.12.003>
- Pym, A. (2011). Translation research terms: a tentative glossary for moments of perplexity and dispute. *Translation Research Projects* 3, pp. 75-110.
- Pym, A. (2016). Exploring Translations Theories. *Cadernos de Tradução*, pp. 214-317. doi:<https://doi.org/10.5007/2175-7968.2016v36n3p214>
- Sin-wai, C. (2015). *The Routledge Encyclopedia of Translation Technology*. London: Routledge.
- Somers, H. (2001). *Machine Translation Applications*. London: Routledge.
- Somers, H. (2003). Machine Translation: Latest Developments. Em R. Mitkov, *Handbook of Computational Linguistics* (pp. 513-528). Oxford: Oxford University Press.
- Specia, L., Raj, D., & Turchi, M. (2010). Machine Translation Evaluation Versus Quality Estimation. *Machine Translation*, pp. 39-50. doi:<https://doi.org/10.1007/s10590-010-9077-2>
- Thicke, L. (2013). The Industrial Process for Quality Machine Translation. *The Journal of Especialised Translation* 19, 8-18. Fonte: [http://www.jostrans.org/issue19/art\\_thicke.pdf](http://www.jostrans.org/issue19/art_thicke.pdf)
- Wagner, E. (1985). Post-Editing Systran - A Challenge for Commission Translators. *Terminologie et Traduction*, pp. 1-6. Fonte: <https://aclanthology.org/www.mt-archive.info/T&T-1985-Wagner.pdf>
- Wilson, D. (2011). The Conceptual-Procedural Distinction: Past, Present and Future. Em V. Escandell-Vidal, M. Leonetti, & A. Ahern, *Procedural Meaning: Problems and Perspectives* (pp. 3-31). Bingley: Emerald.
- Zhechev, V. (2014). Analysing the Post-Editing of Machine Translation at Autodesk. Em L. W. Sharon O'Brien, *Post-editing of Machine Translation: Processes and Applications* (pp. 2-13). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Zwischenberger, C. (2017). Translation as a Metaphoric Traveller Across Disciplines. *Translation and Translanguaging in Multilingual Contexts*, 3, pp. 388– 406. doi:[10.1075/ tmc.3.3.07zwi](https://doi.org/10.1075/tmc.3.3.07zwi)

Zwischenberger, C. (2019). From Inward to Outward: The Need for Translation Studies to Become. *The Translator*, 25, pp. 256– 268. doi:10.1080/13556509.2019.1654060

## Anexo 1 – Norma ISO 18587:2017 - Serviços de tradução - Pós-edição

ISO 18587:2017(E)

### iTeh STANDARD PREVIEW (standards.iteh.ai)

ISO 18587:2017  
<https://standards.iteh.ai/catalog/standards/sist/310c236fb40e-4027-a61e-e10c063df6c/iso-18587-2017>



#### **COPYRIGHT PROTECTED DOCUMENT**

© ISO 2017, Published in Switzerland

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester

ISO copyright office  
Ch. de Blandonnet 8 • CP 401  
CH-1214 Vernier, Geneva, Switzerland  
Tel. +41 22 749 01 11  
Fax +41 22 749 09 47  
copyright@iso.org  
www.iso.org

Contents	Page
Foreword	iv
Introduction	v
<b>1 Scope</b>	<b>1</b>
<b>2 Normative References</b>	<b>1</b>
<b>3 Terms and definitions</b>	<b>1</b>
3.1 Concepts related to machine translation	1
3.2 Concepts related to language and content	2
3.3 Concepts related to people or organizations	3
3.4 Concepts related to translation	4
<b>4 Post-editing process</b>	<b>5</b>
4.1 General	5
4.2 Pre-production processes	5
4.3 Production processes	6
4.3.1 Objectives in the post-editing process	6
4.3.2 Requirements of post-editing MT output	6
4.3.3 Post-editor's tasks	7
4.4 Post-production processes	7
4.4.1 Final verification and delivery	7
4.4.2 Feedback	7
<b>5 Competences and qualifications of post-editors</b>	<b>7</b>
5.1 Competences	7
5.2 Qualifications	8
5.3 Professionalism	8
<b>6 Requirements of full post-editing</b>	<b>8</b>
<b>Annex A (informative) Post-editor training</b>	<b>9</b>
<b>Annex B (informative) Light post-editing</b>	<b>10</b>
<b>Annex C (informative) Pre-editing</b>	<b>11</b>
<b>Annex D (informative) Client-TSP agreements and project specifications</b>	<b>12</b>
<b>Annex E (informative) Automatic post-editing</b>	<b>14</b>
<b>Bibliography</b>	<b>15</b>

## Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see [www.iso.org/directives](http://www.iso.org/directives)).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see [www.iso.org/patents](http://www.iso.org/patents)).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see the following URL: [www.iso.org/iso/foreword.html](http://www.iso.org/iso/foreword.html).

This document was prepared by Technical Committee ISO/TC 37, *Terminology and other language and content resources*, Subcommittee SC 5, *Translation, interpreting and related technology*.

ISO 18587:2017  
<https://standards.iteh.ai/catalog/standards/sist/310c236fb40e-4027-a61e-e10c063dfec/iso-18587-2017>

## Introduction

The use of machine translation (MT) systems to meet the needs of an increasingly demanding translation and localization industry has been gaining ground. Many translation service providers (TSPs) and clients have come to realize that the use of such systems is a viable solution for translating projects that need to be completed within a very tight time frame and/or with a reduced budget. When an MT system is used, clients can have material translated that can otherwise not be translated; translation costs can be decreased and the launch of products on specific markets, as well as the flow of information, can be accelerated. On the other hand, TSPs are able to:

- a) improve translation productivity;
- b) improve turn-around times;
- c) remain competitive in an environment where clients show an increasing demand for using MT in translation.

However, there is no MT system with an output which can be qualified as equal to the output of human translation and, therefore, the final quality of the translation output still depends on human translators and, for this purpose, their competence in post-editing.

The rate at which MT systems are changing renders it impractical to produce an overarching International Standard on these systems, which could stifle innovation or be ignored by the translation technology development industry.

This document therefore restricts its provisions to that part of the process that begins upon the delivery of the MT output and the beginning of the human process that is known as post-editing.

iTech STANDARD PREVIEW  
(standards.iteh.ai)  
ISO 18587:2017  
<https://standards.iteh.ai/catalog/standards/sist/310c236fb40e-4027-a61e-e10c063dffec/iso-18587-2017>

**iTeh STANDARD PREVIEW**  
**(standards.iteh.ai)**

ISO 18587:2017

<https://standards.iteh.ai/catalog/standards/sist/310c236fb40e-4027-a61e-e10c063dffec/iso-18587-2017>

# Translation services — Post-editing of machine translation output — Requirements

## 1 Scope

This document provides requirements for the process of full, human post-editing of machine translation output and post-editors' competences.

This document is intended to be used by TSPs, their clients, and post-editors.

It is only applicable to content processed by MT systems.

NOTE For translation services in general, see ISO 17100.

## 2 Normative References

There are no normative references in this document.

## 3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

- IEC Electropedia: available at <http://www.electropedia.org/>  
<https://standards.iteh.ai/catalog/standards/sist/510c236f-840e-4027-a61e-191851861189/iso-18587-2017>
- ISO Online browsing platform: available at <http://www.iso.org/obp>

### 3.1 Concepts related to machine translation

#### 3.1.1

##### machine translation

##### MT

automatic translation (3.4.2) of text (3.2.6) from one natural language to another using a computer application

[SOURCE: ISO 17100:2015, 2.2.2, modified – reference to translation of speech has been deleted as it is not relevant to this document; also “automated” has been replaced by “automatic” in order to avoid confusion with translation memory tools]

#### 3.1.2

##### machine translation output

##### MT output

result of machine translation (3.1.1)

[SOURCE: ISO 17100:2015, 2.2.3, modified – “outcome” has been changed to “result”]

#### 3.1.3

##### machine translation system

technology used to perform machine translation (3.1.1)

## ISO 18587:2017(E)

### 3.1.4

#### **post-edit**

edit and correct *machine translation output* (3.1.2)

[SOURCE: ISO 17100:2015, 2.2.4, modified – the note has been deleted]

### 3.1.5

#### **full post-editing**

process of *post-editing* (3.1.4) to obtain a product comparable to a product obtained by *human translation* (3.4.3)

### 3.1.6

#### **light post-editing**

process of *post-editing* (3.1.4) to obtain a merely comprehensible text without any attempt to produce a product comparable to a product obtained by *human translation* (3.4.3)

## 3.2 Concepts related to language and content

### 3.2.1

#### **content**

information in any form

EXAMPLE Text, audio, video, etc.

### 3.2.2

#### **source language**

language of the *content* (3.2.1) to be *translated* (3.4.1)

### 3.2.3

#### **source language content**

language *content* (3.2.1) to be *translated* (3.4.1) ISO 18587:2017  
<https://standards.iteh.ai/catalog/standards/sist/310c236fb40e-4027-a61e-e10e063df1ec/iso-18587-2017>

[SOURCE: ISO 17100:2015, 2.2.3]

### 3.2.4

#### **target language**

language into which *source language content* (3.2.3) is *translated* (3.4.1)

[SOURCE: ISO 17100:2015, 2.3.6]

### 3.2.5

#### **target language content**

language *content* (3.2.1) *translated* (3.4.1) from *source language content* (3.2.3)

[SOURCE: ISO 17100:2015, 2.3.3]

### 3.2.6

#### **text**

*content* (3.2.1) in written form

[SOURCE: ISO 17100:2015, 2.3.4]

### 3.2.7

#### **natural language**

##### **NL**

language with its origin unknown, but continuously developing sometimes in idiosyncratic ways as is used conventionally for human communication

[SOURCE: ISO/TS 24620, 2.12]

**3.2.8****controlled natural language**

controlled language

**CNL**

subset of *natural languages* (3.2.7) whose grammars and dictionaries have been restricted in order to reduce or eliminate both ambiguity and complexity

Note 1 to entry: As a generic, CNL is an uncountable noun that refers to the abstract properties of all controlled natural languages and not to a particular natural language or application for a specific purpose. It is engineered (i.e. constructed) with a view to reducing or eliminating ambiguity and complexity and aims both to make it easier for human readers (particularly non-native users, non-experts and people with limited comprehension) to read a *text* (3.2.6), and to improve the computational processing of a text.

Note 2 to entry: CNL is an engineered (i.e. constructed) language that is based on a particular natural language, but is more restrictive as regards lexicon, syntax, or semantics, while at the same time preserving most of its natural properties. Here, CNL is a countable noun.

[SOURCE: ISO/TS 24620, 2.6]

**3.2.9****segment**

unit of *text* (3.2.6) produced for a computer application to facilitate translation

Note 1 to entry: A segment can be a sentence, heading or other unit of text, such as phrase, word or a single character.

**3.2.10****locale**

set of characteristics, information or conventions specific to the linguistic, cultural, technical and geographical conventions of a target audience

**3.2.11****language register**

variety of language used for a particular purpose or in a particular social or industrial domain

**3.3 Concepts related to people or organizations****3.3.1****client**

customer

person or organization that commissions a service from a *TSP* (3.3.5) by formal agreement

Note 1 to entry: The client can be the person or organization requesting or purchasing the service and can be external or internal to the *TSP*'s (3.3.5) organization.

[SOURCE: ISO 17100:2015, 2.4.3]

**3.3.2****translator**

person who *translates* (3.4.1)

[SOURCE: ISO 17100:2015, 2.4.4]

**3.3.3****reviser**

person who revises *translation output* (3.4.4)

[SOURCE: ISO 17100:2015, 2.4.5, modified – no need to specify that it is against source language content as it is explained in the definition of “revision”]



**3.4.7****revision**

bilingual examination of *target language content* (3.2.5) against *source language content* (3.2.3) for its suitability for the agreed purpose

Note 1 to entry: The term "bilingual editing" is sometimes used as a synonym for revision.

[SOURCE: ISO 17100:2015, 2.2.6]

**3.4.8****verification**

confirmation by the *TSP's* (3.3.5) *project manager* (3.3.6) that specifications have been fulfilled

[SOURCE: ISO 17100:2015, 2.5.1, modified – TSP added for the sake of clarity]

**4 Post-editing process****4.1 General**

Post-editing is performed on MT output for the purpose of checking its accuracy and comprehensibility, improving the text, making the text more readable, and correcting errors. Post-editing differs from translation as it involves three texts: the source text, the MT output and the final target text.

There are two main levels of post-editing, light and full; their use depends on the purpose of the translation output and the client's requirements. This document deals with full post-editing. Light post-editing is described in [Annex B](#).

The post-editing process can be automatic or human. Automatic post-editing is described in [Annex E](#). Human post-editing refers to the MT output post-editing process performed by a human post-editor.

The post-editing process can take place in an environment where all the text to be processed has been machine translated or in an integrated environment, where MT and TM, along with consistent terminology management, are fully integrated to produce a multi-modal translation work environment.

Post-edited MT output can or cannot be fed into a translation memory (automatically or partially), depending on the organization's processes and its clients' quality assurance requirements.

**4.2 Pre-production processes**

The TSP shall have a process in place to determine, in consultation with the client when necessary, whether the source language content is suitable for MT and subsequent post-editing or not since combined MT/post-editing efficiency depends on the MT system, language combination, domain and style of the source language content.

The source language content can be pre-edited before machine translation to facilitate machine processing, improve raw MT output quality, and therefore reduce the post-editing workload. Pre-editing is described in [Annex C](#).

The TSP shall finalize an agreement with the client and retain a record of that agreement. If an agreement is made verbally or by telephone, the TSP shall confirm the agreement and its terms in writing (e.g. by letter, fax or email). The agreement, whether contractual or non-contractual, shall include or reference the commercial terms and the project specifications. The agreement may also call for conformity to this document. [Annex D](#) contains a list of the items which can be included in the agreement.

Any deviation from the original agreement shall be agreed upon by all parties before any action is taken that deviates from the original agreement, and the agreed amendment shall be recorded and stored with the original agreement documentation.

The TSP shall ensure that requirements for the post-editing project are identified, documented and accessible to the post-editor. The TSP shall also ensure that the relevant specifications (e.g. quality

## Anexo 2 - Categorias de erros de pós-edição – TAUS (DQF-MQM)

Post-editing error type categories (DQF-MQM template by TAUS <https://info.taus.net/dqf-mqm-error-typology-templ>)

Error type	Sub-type	Definition	Examples
Accuracy		The target text does not accurately reflect the source text, allowing for any differences authorized by specifications.	Translating the Italian word 'canals' into English as 'canals' instead of 'channels'.
	Addition	The target text includes text not present in the source.	A translation includes portions of another translation that were inadvertently pasted into the document.
	Omission	Content is missing from the translation that is present in the source.	A paragraph present in the source is missing in the translation.
	Mistranslation	The target content does not accurately represent the source content.	A source text states that a medicine should not be administered in doses greater than 200 mg, but the translation states that it should be administered in doses greater than 200 mg (i.e., negabon has been omitted).
	Over-translation	The target text is more specific than the source text.	The source text refers to a boy but is translated with a word that applies only to young boys rather than the more general term.
	Under-translation	The target text is less specific than the source text.	The source text uses words that refer to a specific type of military officer but the target text refers to military officers in general.
	Untranslated text	Content that should have been translated has been left untranslated.	A sentence in a Japanese document translated into English is left in Japanese.
	Improper exact TM match	A translation is provided as an exact match from a translation memory (TM) system but is actually incorrect.	A TM system returns Press the Start button as an exact (100%) match when the proper translation should be Press the Begin button.
	Punctuation	Issues related to the form or content of a text, irrespective as to whether it is a translation or not.	A text has errors in it that prevent it from being understood.
	Spelling	Issues incorrectly (for the locale or style).	An English text uses a semicolon where a comma should be used.
Fluency	Grammar	Issues related to spelling of words.	The German word Zustellung is spelled Zusetlun.
	Grammatical register	Issues related to the grammar or syntax of the text, other than spelling and orthography.	An English text reads The man was seeing the his wife.
	Inconsistency	The content uses the wrong grammatical register, such as using informal pronouns or verb forms when their formal counterparts are required.	A text used for a highly formal announcement uses the Norwegian du form instead of the expected De.
	Link/cross-reference	The text shows internal inconsistency.	A text uses both app. and approx. for approximately.
	Character encoding	Links are inconsistent in the text.	An HTML file contains numerous links to other HTML files; some have been updated to reflect the appropriate language version while some point to the source language version.
	Inconsistent with termbase	Characters are garbled due to incorrect application of an encoding.	A text document in UTF-8 encoding is opened as ISO Latin-1, resulting in all upper ASCII characters being garbled.
	Inconsistent use of terminology	A term (domain-specific word) is translated with a term other than the one expected for the domain or otherwise specified.	A French text translates English e-mail as e-mail but terminology guidelines mandated that courriel be used. The English musological term dog is translated (literally) into German as Hund instead of as Schnorre, as specified in a terminology database.
	Awkward	A term is used inconsistently with a specified termbase.	A termbase specifies that the term USB memory stick should be used, but the text uses USB flash drive.
	Company style	Terminology is used in an inconsistent manner within the text.	The text refers to a component as the brake release lever, brake disengagement lever, manual brake release, and manual disengagement release.
	Inconsistent style	The text has stylistic problems.	The translation of a light-hearted and humorous advertising campaign is in a serious and "heavy" style even though specifications said it should match the style of the source text.
Style	Third-party style	A text is written with an awkward style.	A text is written with many embedded clauses and an excessively wordy style. While the meaning can be understood, the text is very awkward and difficult to follow.
	Unidiomatic	The text violates company/organization-specific style guidelines.	Company style states that passive sentences may not be used but the text uses passive sentences.
		The text violates a third-party style guide.	One part of a text is written in a light and terse style while other sections are written in a more wordy style. Specifications stated that English text was to be formatted according to the Chicago Manual of Style, but the text delivered followed the American Psychological Association style guide.
		The content is grammatical, but not idiomatic.	The following text appears in an English translation of a German letter: "We thanked him with heart" where "with heart" is an understandable, but non-idiomatic rendering, better stated as "heartily".

					A document is formatted incorrectly.
					An English sentence is 253 characters long but its German translation is 51 characters long.
Design	Length	There is a problem relating to design aspects (vs. linguistic aspects) of the content.	There is a significant discrepancy between the source and the target text lengths.		A portion of the text displays a (non-systematic) formatting problem (e.g., a single heading is formatted incorrectly, even though other headings appear properly).
	Local formatting	Issues related to local formatting (rather than to overall layout concerns).			Markup is used incorrectly, resulting in incorrect formatting.
	Markup	Issues related to markup (codes used to represent structure or formatting of text, also known as tags).			A translation is complete, but during DTP a text box was inadvertently moved off the page and so the translated text does not appear in a rendered PDF version.
	Missing text	Existing text is missing in the final laid-out version.			The German translation of an English string in a user interface runs off the edge of a dialogue box and cannot be read.
	Truncation/text expansion	Truncation/text-expansion.			An incorrect format for currency is used for a German text, with a period (.) instead of a comma (,) as a thousands separator. A text translated into Japanese uses Western quote marks to indicate titles rather than the appropriate Japanese quote marks (「 and 」).
	Address format	The text does not adhere to locale-specific mechanical conventions and violates requirements for the presentation of content in the target locale.			An online form translated from English to Hindi requires a house number even though many addresses in India do not include a house number.
Locale convention	Date format	Content uses the wrong format for addresses.			An English text has 202-06-07 instead of the expected 06/07/2012.
	Currency format	A text uses a date format inappropriate for its locale.			A text dealing with business transactions from English into Hindi assumes that all currencies will be expressed in simple units, while the convention in India is to give such prices in lakh rupees (100,000 rupees).
	Measurement format	Content uses the wrong format for currency.			A text in France uses feet and inches and Fahrenheit temperatures.
	Shortcut key	A text uses a measurement format inappropriate for its locale.			A software product uses CTRL-S to save a file in Hungarian, rather than the appropriate CTRL-M (for mentem).
	Telephone format	A translated software product uses shortcuts that do not conform to locale expectations or that make no sense for the locale.			A German text presents a telephone number in the format (xxx) xxx - xxxx instead of the expected (xx) followed by a group of digits separated into groups by spaces.
Verity	Culture-specific reference	Content uses the wrong form for telephone numbers.			The text states that a feature is present on a certain model of automobile when in fact it is not available.
Other		The text makes statements that contradict the world of the text.			An English text refers to steps in a process as <i>First base</i> , <i>Second base</i> , and <i>Third base</i> , and to successful completion as a <i>Home run</i> and uses other metaphors from baseball. These prove difficult to translate and confuse the target audience in Germany.
		Content inappropriately uses a culture-specific reference that will not be understandable to the intended audience.			
		Any other issues.			

Severity Levels	
Critical	Errors that may carry health, safety, legal or financial implications, violate geopolitical usage guidelines, damage the company's reputation, cause the application to crash or negatively modify/misrepresent the functionality of a product or service, or which could be offensive.
Major	Errors that may confuse or mislead the user or hinder proper use of the product/service due to significant change in meaning or because errors appear in a visible or important part of the content.
Minor	Errors that don't lead to loss of meaning and wouldn't confuse or mislead the user but would be noticed, would decrease stylistic quality, fluency or clarity, or would make the content less appealing.
Neutral	To log additional information, problems or changes to be made that don't count as errors, e.g. they reflect a reviewer's choice or preferred style, they are repeated errors or instruction/glossary changes not yet implemented, a change to be made that the translator is not aware of.
Kudos	Used to praise for exceptional achievement.

**MemoQ LQA Model:** see *QS MTPE error typology* in *MemoQ LQA Settings*, under *Resource Console*.

<https://docs.memoq.com/current/en/Things/things-linguistic-quality-assurance.html>