

O USO DO MEMOQ NAS TAREFAS DE PRÉ-TRADUÇÃO: REVISÃO CRÍTICA DE FUNÇÕES E SUGESTÕES DE MELHORIA

Tatiana Limongi Chaves*

RESUMO: O presente artigo apresenta sugestões de melhorias e novas funcionalidades para o memoQ com o objetivo de conseguir um ambiente tecnológico integrado e abrangente de tradução, que inclua funções de pré-tradução. As melhorias sugeridas são baseadas em observações de campo feitas durante um estágio numa empresa de tradução. A análise das funções concentra-se nas tarefas de pré-tradução por ter sido a fase em que mais constrangimentos foram encontrados, enquanto as sugestões abrangem várias etapas do ciclo de tradução.

PALAVRAS-CHAVE: Memoq, Formatação, Online, Ambiente Integrado

1. Introdução

As CAT tools (“computer-aided translation tools”) são consideradas um campo de estudo emergente quando comparado a outros estudos envolvendo tecnologia e tradução. Cada ferramenta, por esse motivo, carece de melhorias e aperfeiçoamentos para que possa estar cada vez mais alinhada com os reais problemas do profissional de tradução (Peng, 2020, p. 1).

O objetivo deste artigo é contribuir para o avanço das ferramentas CAT, em especial o memoQ, tomando por base as observações e a experiência obtidas durante um estágio numa empresa de tradução. O artigo concentra-se nas tarefas de pré-tradução, por terem sido aquelas em que o software analisado tem maiores carências.

Este artigo está subdividido em três seções, detalhadas em seguida. Na seção 2, apresenta-se a evolução do uso de ferramentas tecnológicas no campo da tradução, a partir de um referencial teórico selecionado (Peng, H., Pietrzak, P., and Kornacki, M., Sinwai, C.). Inicialmente, faz-se um breve apanhado da evolução do uso de ferramentas tecnológicas no campo da tradução, procurando-se mostrar como os tradutores atualmente utilizam as ferramentas CAT de forma habitual. Prossegue-se com a apresentação de uma breve cronologia das ferramentas CAT, na qual se mostra a sua evolução em paralelo com a evolução tecnológica. A seção termina com uma análise da evolução da ferramenta memoQ em particular, uma vez que foi a ferramenta em que incidiu o meu projeto de investigação.

Na seção 3 descreve-se, de forma detalhada, alguns tipos de atividades práticas analisadas e os problemas encontrados nas ferramentas CAT, especificamente na ferramenta memoQ, durante o meu estágio e prática profissional. Procurei categorizar os tipos de atividades realizadas. Há uma maior incidência na atividade de formatação (pré-tradução) e grande parte das melhorias sugeridas teve como motivação o fato de pouco valer o uso de uma ferramenta CAT na tradução se o texto estiver mal formatado e com erros tipográficos (por exemplo).

* milaydis@med.up.pt

Na seção 4, a partir principalmente dos problemas descritos nas análises efetuadas (seção 3), elenca-se um conjunto de propostas de melhoria aplicáveis ao memoQ. Novamente, procurei categorizar as melhorias sugeridas, para facilitar sua explicação e compreensão. Conforme mencionado anteriormente, grande parte das melhorias concentra-se na atividade de formatação, exercida por muitos tradutores e outros profissionais de tradução.

2. Evolução do uso de ferramentas tecnológicas no campo da tradução

2.1 As ferramentas de auxílio à tradução

Nesta seção são abordadas algumas referências teóricas relevantes sobre o uso das ferramentas CAT (*computer-aided translation*) em ambiente profissional de tradução. Importa ressaltar aqui a definição deste tipo de ferramentas, os seus benefícios para o campo da tradução, e a evolução histórica de seu uso.

Hao Peng (2020) define uma ferramenta CAT como “Computer-aided Translation (CAT), a *software* using computer to help with translation, is working with the aid with the computer”, e prossegue afirmando que “CAT can help translators to translate better and help people to finish foreign language translation easily and effectively” (Peng, 2020, p. 1). Hoje, muitos tradutores estão dependentes da tecnologia para fazer o seu trabalho, pois as ferramentas CAT vêm expandindo as suas funcionalidades e abrangendo diversas tarefas do macroprocesso de tradução (Sin-wai, 2017).

Segundo Sin-wai (2017, p. 273), as tecnologias aplicadas à tradução são cada vez mais tema de reflexão no meio acadêmico:

For a very long time in the past, scholars in the field discussed translation from a literary and personal perspective. Nowadays, more and more attention has been given to translation technology.

Em relação ao fluxo de trabalho do macroprocesso de tradução, que compreende todas as etapas preparatórias para a tradução, as ferramentas CAT fizeram um avanço significativo nas suas funcionalidades. Segundo Pietrzak e Kornacki (2021, pp. 46-47), “Translation technology contributes to a significant change in the way in which translators approach and manage their tasks and broadly defined workflow.”

É inegável, segundo diversos autores, o ganho de produtividade e de qualidade que as ferramentas CAT trouxeram ao trabalho do tradutor. Esse ganho continua crescendo, à medida que os fabricantes lançam novas versões de suas ferramentas CAT. Conforme escrevem Pietrzak e Kornacki (2021, p. 47), “CAT technology has optimised the translation process to a large extent. First and foremost, it greatly enhanced the speed and efficiency of translation, which caused a lot of changes to the very process of translation”. Prosseguem ainda os autores, em referência ao impacto econômico decorrente desse aumento de produtividade e qualidade: “With various services that support cross-language communication, CAT technology has been developing continuously, and its innovations have had a huge economic impact.”

As atividades acessórias ao processo específico de tradução acabaram entrando no radar das ferramentas CAT. Diferentemente do processo de tradução propriamente dito, tem-se nesse caso uma miríade de tarefas que dão suporte à tradução. Conforme afirma Sin-wai (2017, p. 40):

Rather than simply translating the document, a computer-aided translator has to engage in other types of work, such as authoring, pre-editing, interactive editing, postediting, termbase management, translation memory database management, text alignment, and manual alignment verification.

Além das atividades mencionadas por esse autor, pode-se ainda acrescentar a essa lista duas outras: revisão de texto (*proofreading*) e reconhecimento ótico de caracteres (OCR).

Sin-wai (2017, p. 38) discorre sobre como, devido a esses avanços, as ferramentas CAT se tornaram indispensáveis para a atividade profissional do tradutor:

All computer-aided translation systems aim to increase translation productivity. In terms of the means of production, all translation nowadays can be considered computer-aided translation, as virtually no one would be able to translate without making use of a computer.

Pode-se ainda acrescentar que houve um grande avanço no processo de tradução auxiliado por ferramentas CAT em si mesmo, visto que está disponível atualmente ao tradutor um sofisticado ambiente integrado e interativo de suporte à tradução, disponibilizado por praticamente todos os fabricantes.

De acordo com Peng (2020, p. 4), os benefícios do uso de uma ferramenta CAT alargam-se a vários níveis:

(...) applying the CAT on the translation practice, can improve the operability, improve the translation practice quality, speed up translation, and meet the demands of translation practice, accelerating the communications between different countries and different languages.

As ferramentas CAT compartilham esses benefícios com todas as ferramentas do tipo CA (*computer-aided*), como, por exemplo, as ferramentas CAD (*computer-aided design*), CAE (*computer-aided engineering*), CASE (*computer-aided system engineering*), entre outras.

2.1 Breve cronologia das ferramentas CAT

Para acompanhar a breve história da tecnologia na tradução, optou-se por seguir a cronologia de Sin-Wai (2017, p. 1). O autor divide o desenvolvimento da CAT em 4 períodos:

1) 1967-1983: período da germinação;

- 2) 1984-1992: período de constante crescimento;
- 3) 1993 a 2002: período de rápido crescimento;
- 4) 2003-2014: período de desenvolvimento global.

A respeito do período pós-computador pessoal (PC) e início da disseminação da internet, Sin-wai (2017, p. 7) discorre sobre os grandes avanços da década a partir de então:

The decade that extends from 1993 to 2003 was a period of rapid growth, due largely to (1) the emergence of more commercial systems, (2) the development of more built-in functions, (3) the dominance of Windows operating systems, (4) the support of more document formats, (5) the support of more languages for translation, and (6) the dominance of Trados as a market leader.

Os principais avanços tecnológicos da época foram:

1) A consolidação do computador pessoal de 16 bits, que foi desenvolvido numa arquitetura padronizada por todos os fabricantes (o que não aconteceu com o antigo microcomputador de 8 bits) e que constituiu um importante marco. Dessa maneira, um programa desenvolvido para PC funcionava na máquina de qualquer fabricante sem precisar de adaptações.

2) O advento da internet, que também contribuiu desde o início para a expansão da oferta de ferramentas de computador. A grande contribuição da internet foi permitir a venda de programas pelo computador, bastando efetuar um *download*. Antes da internet, o cliente teria de encontrar uma loja que vendesse a caixa do programa com as mídias (disquete, CD, DVD) de instalação dentro. Evidentemente, nem todas as lojas tinham todos os programas à venda.

Já nas duas primeiras décadas do século XXI, registrou-se o surgimento de múltiplas ferramentas CAT, devido principalmente aos avanços tecnológicos da época. De acordo com Sin-wai (2017, p. 13):

Advances in technology have given added capabilities to computer-aided translation systems. During the last ten years [2004-2014], most old systems have been upgraded on a regular basis, and close to 30 new systems have been released to the market.

2.2 memoQ

Foi exatamente nesse período inicial do século XXI que surgiu o memoQ, conforme explica Sin-wai (2017, p. 14):

In 2004, Kilgray Translation Technologies was established by three Hungarian language technologists. The name of the company was made up of the founders' surnames: Kis Balázs (KI), Lengyel István (L), and Ugray Gábor (GRAY). Shortly after, in 2005, the company launched the first version of MemoQ, an Integrated Localization Environment (ILE).

Em 2004, esses três tradutores começaram a trabalhar em projetos referentes à tradução assistida e verificação ortográfica que deram origem ao memoQ. Esse movimento iniciou-se a partir de um encontro de profissionais da língua, que colaboraram na criação de um novo ambiente de tecnologia de tradução (memoQ ©, 2020).

Os três empreendedores começaram a desenvolver a tecnologia que imaginaram, correspondente ao núcleo do memoQ. Foi uma tarefa desafiadora, pois uns poucos intervenientes importantes da indústria dominavam o mercado naquela época. No entanto, os três fundadores da Kilgray tinham uma visão clara de como executar a transformação digital da indústria da tradução (memoQ ©, 2020).

O lado visionário dos profissionais que criaram o memoQ – que conheciam em profundidade os processos e problemas da tradução – foi tão amplo e popular que alastrou a outras empresas, passando a usar *interfaces* semelhantes para suas ferramentas CAT.

O memoQ conseguiu uma boa posição no mercado pois, ao tornar a tradução mais eficiente e agradável, fez sucesso entre os tradutores. Foi esse sucesso o fator determinante para que o memoQ pudesse enfrentar as grandes empresas já estabelecidas e conquistar uma fatia significativa do mercado.

3. A preparação de textos (etapa pré-tradução)

3.1 Definição de formatação e problemas habituais

Esta seção concentra-se nos processos pré-tradução e pós-tradução. Não será abordado, nesta seção, o processo de formatação pós-tradução, que não foi objeto do meu estágio. As indicações aqui incluídas são boas práticas com que me deparei durante o estágio e que podem ser integradas nas ferramentas, nalguns casos, como descrito na seção 4.

A formatação do documento é necessária para que a ferramenta CAT (memoQ neste caso) possa identificar corretamente o texto contido no arquivo. Não raro o cliente fornece o conteúdo a ser traduzido numa forma que o memoQ não consegue ler, como por exemplo uma foto de um documento. Nestes casos, torna-se necessária a formatação do documento, que também ajuda a solucionar o problema de “tags” excessivas. Ao limpar o texto de certas formatações, muitas dessas “tags” desaparecem.

As “tags” (*Text and Graphics*) são elementos de uma metalinguagem. Dentro do texto das páginas da internet há “tags” da metalinguagem HTML (*Hyper Text Markup Language*). Dentro do texto de um documento em Word há “tags” da metalinguagem RTF (*Rich Text Format*). Quando o memoQ processa um texto em Word, interpreta as “tags” do Word e gera “tags” suas.

Tipicamente, uma “tag” é composta por uma sequência não usual de caracteres, como por exemplo “~<”. Após a abertura da “tag”, surgem informações de formatação tais como início de texto em negrito ou em itálico, ou início de uma nova fonte, ou ainda fim de texto em negrito, etc. No final, vêm os caracteres de encerramento da “tag”. O Word, por exemplo, não mostra a imensa maioria de suas “tags”. Mostra apenas, quando configurado para tal, quebras de parágrafo, quebras de página, etc. Já o memoQ mostra algumas de suas “tags” através de um símbolo, tipicamente aquelas diferentes de

identificação de texto negrito, sublinhado ou itálico. A figura 1 abaixo ilustra um exemplo de “tag” no memoQ:

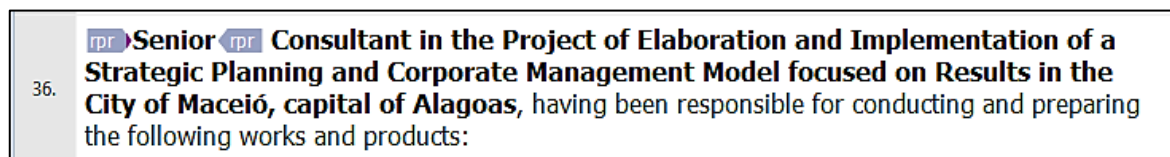


Figura 1. “tags”no memoQ.

As “tags” existem para permitir à ferramenta recriar a formatação inicial do documento na versão traduzida. Um documento graficamente complexo terá muitas “tags”, o que pode dificultar a tarefa do tradutor. A tarefa de formatação implica criar um documento com o menor número de “tags” possível, equilibrando tal objetivo com a recriação de um documento final graficamente bem-feito e pronto a entregar ao cliente.

3.2 Etapas da preparação de texto

3.2.1 A conversão

Para formatar um documento, primeiro deve-se convertê-lo para o Microsoft Word (Word). Dependendo do formato em que se recebe o documento, essa etapa será diferente. O ideal é pedir ao cliente uma versão do documento em Word, mas nem sempre será possível obtê-la. Aqui serão tratados apenas os arquivos do tipo PDF (*Portable Document Format*).

Se o documento for um PDF de texto – que possibilita selecionar e copiar o texto –, o primeiro passo é passar o texto para o Word. Existem duas formas de o fazer: a primeira é simplesmente selecionar tudo, copiar e colar no Word, mas é preciso ter em atenção que, se o documento tiver imagens, estas não serão copiadas. A segunda é, caso se disponha do Adobe Acrobat Pro (Adobe, 2021), ou de algum outro programa similar, selecionar a opção “Exportar para” e “Microsoft Word”.

Se o documento for um PDF de texto baseado em imagem – que não permite selecionar o texto –, o primeiro passo será utilizar uma ferramenta de OCR (*Optical Character Recognition*), que fará a conversão de imagens de textos digitados ou escritos para texto codificado por máquina. Este processo é muito utilizado quando o documento é um papel escaneado ou é uma foto de um texto. Quanto melhor a qualidade (resolução, iluminação, sem sombras) do texto escaneado, melhor o resultado do OCR. O Adobe Acrobat Pro possui um OCR embutido no programa capaz de converter a imagem em texto para o Word. Quando se seleciona o comando “Exportar para o Word”, o Adobe Acrobat Pro nota que o documento é texto baseado em imagem e automaticamente faz o OCR antes, sem precisar que tal seja especificado. Também é possível fazer o OCR sem exportar o documento, através dos comandos “Ferramentas” e “Digitalizar e OCR”. É recomendável não usar o “Converter para Texto”, e sim o “Aprimorar”, que fica do lado. É necessário, nesse momento, selecionar primeiro o idioma desejado antes de pressionar o “Aprimorar” (ver figura 2). A razão para tal é que ele ajusta a digitalização, ajusta páginas empenadas,

tira riscos, reconhece as palavras do idioma, deixa o PDF mais alinhado e reconhece então o texto com mais qualidade.

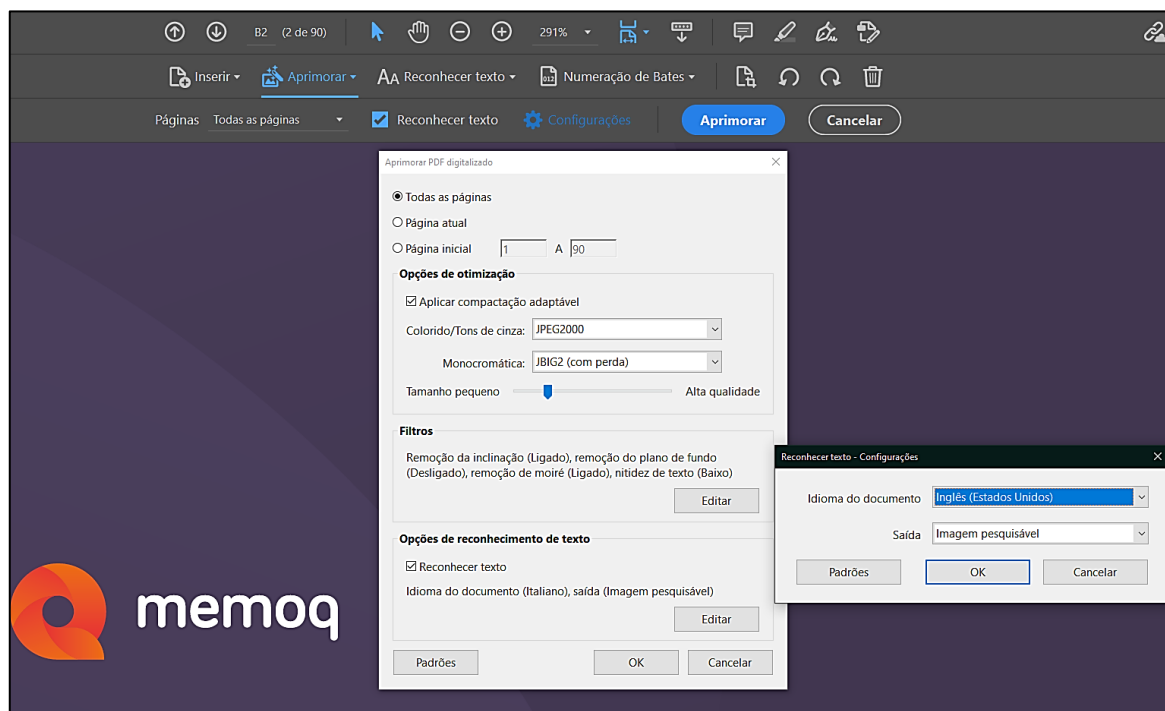


Figura 2. OCR no Adobe Acrobat Pro.

Em ambos os casos – textos que se podem copiar e colar para o Word e aqueles em que se deve utilizar o OCR –, deve-se verificar se existem imagens não editáveis, gráficos que possam ser convertidos usando o Word, para convertê-los antes da tradução, idealmente antes da orçamentação.

3.2.2 A formatação

A conversão do texto para o Microsoft Word é apenas a preparação para um longo e complexo processo de formatação de texto. Cada tradutor tem a sua preferência quanto aos métodos a utilizar. Alguns preferem copiar todo o texto, colar no Bloco de Notas (aplicativo do Windows) e então copiar o texto do Bloco de Notas para o Word (jblosser, 2011). Deste modo, grande parte da edição será apagada. Outros preferem utilizar apenas as ferramentas disponibilizadas pelo Microsoft Word. Existem, pois, vários métodos, mas aqui será abordado apenas aquele utilizado pela autora deste artigo, que se divide em três etapas.

Primeira etapa

Nesta primeira etapa, a ação inicial a ser realizada se refere a limpeza da formatação do texto, pois a exportação do texto PDF para Word frequentemente gera um excedente de formatações. Torna-se bastante comum o fato de, ao invés de um trecho de texto ser formatado como fonte azul, ser gerada uma formatação em cada palavra, todas iguais. Tal

provoca um excesso de “tags” no memoQ. Diante desta ocorrência, é melhor desformatar tudo e formatar de novo a fonte azul. Nessa situação, as “tags” reduzem-se ao mínimo, tal como deveria estar no documento original.

Antes de começar a formatação, é recomendado ativar o modo “Mostrar Tudo” (ilustrado pelo símbolo “¶”) na guia “Página Inicial” do Word. Este comando mostra caracteres ocultos, como espaços, marcadores de parágrafo ou marcas de tabulação. A importância de esta marcação ser ativada é que deixa visível os excessos de espaçamento, quebras de linha e inadequação de quebra de parágrafo. A figura 3 abaixo exemplifica melhor o que foi relatado:

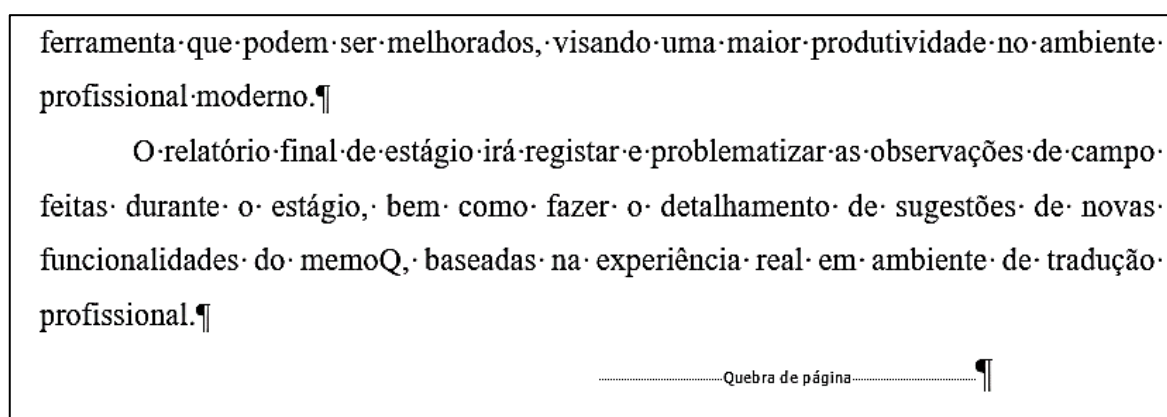


Figura 3. Caracteres ocultos.

Para fazer a limpeza das formatações deve-se primeiro selecionar todo o texto e, em seguida, na guia “Página Inicial” do Word, clicar no ícone “Limpar Toda a Formatação” (simbolizado por “A”). Depois – ainda com todo o texto selecionado – clica-se no símbolo “¶”, que está no canto inferior direito do grupo “Fonte” da barra de ferramentas, para mostrar a caixa de diálogo com todas as opções de formatação de fonte. Aí, na categoria “Efeitos”, desmarca-se todas as opções – tais como “Tachado”, “Tachado duplo”, “Sobrescrito”. Em seguida, muda-se a “Fonte” para “Times New Roman” – devido à facilidade em distinguir as letras, como o “l” (L) minúsculo do “I” maiúsculo (I) – o “Estilo da fonte” para “Regular” e o “Tamanho” para 12 (doze) (ver figura 4).

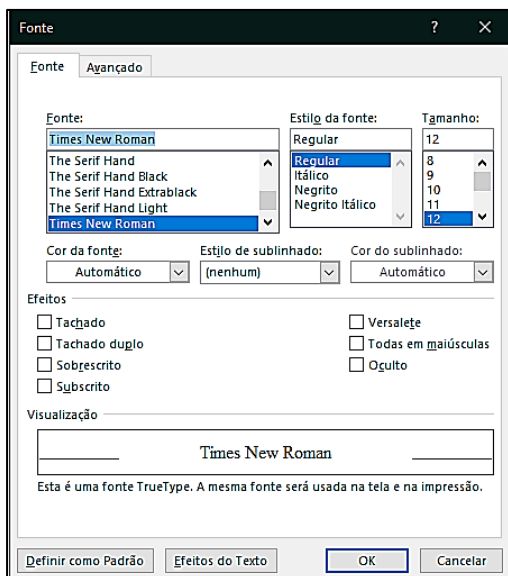



Figura 4. Caixa de diálogo Fonte.

Por fim, na guia “Layout”, seleciona-se a opção “Margens”, escolhe-se “Normal”, e depois clica-se no símbolo “”, que está no canto inferior direito do grupo “Configurar Página” da barra de ferramentas, para mostrar a caixa de diálogo com todas as opções de formatação de parágrafo. Aí, na aba “Papel”, na categoria “Tamanho do papel:”, escolhe-se a opção “A4”, conforme descrito na figura 5.

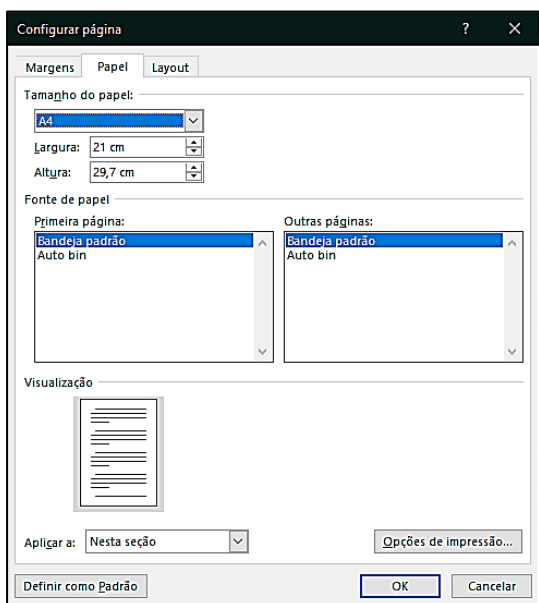


Figura 5. Caixa de diálogo “Configurar Página”.

Segunda etapa

Como na primeira etapa teve de ser retirada toda a formatação da conversão, pois estava repetitiva e excessiva, nesta etapa tem-se de colocar toda a formatação do documento de origem de volta. Esta ação tem o objetivo de otimizar a formatação.

A formatação padrão é a “Regular Normal”. Só são formatadas as palavras em negrito, itálico, sublinhado etc. Em algumas situações há que alterar também o tamanho da fonte de certas palavras em destaque.

Em relação às imagens, deve-se substituí-las por uma representação delas para que haja uma fidelidade segura da transposição de significados. Esta representação deve ficar no mesmo local da imagem substituída.

Serão apresentados, a seguir, alguns exemplos de várias situações a título de ilustração:

A) Independentemente de a figura vir com texto ou sem texto, ela deve ser excluída do documento;

B) No lugar da imagem deve ser incluído o nome dela entre colchetes – como, por exemplo [selo], [assinatura], [carimbo], [imagem], [foto], [marca de água], [logótipo], etc. Porém, deve-se colocar este texto exatamente no mesmo local onde estava a imagem. A tabela abaixo ilustra algumas possibilidades vistas nos exemplos citados neste parágrafo (tabela 2).





Imagem	Nomenclatura
	[selo]
	[carimbo]
	[imagem]
	[assinatura]

Tabela 2. Compilado de exemplos de imagens frequentes na segunda etapa da formatação

O mesmo deve ser feito quando as imagens contêm texto. No caso, o texto deve ser digitado junto com a nomenclatura da imagem – como [logótipo: Palitos Ltda.]. Esse procedimento também deve ser seguido quando apenas um pequeno texto aparece solto em diferentes posições na página – como avisos de direitos autorais, data da impressão, etc. Para isso, deve-se excluir as Caixas de Texto no documento, pois elas criam muitas “tags” no memoQ. Deve-se também arrumar a disposição dos elementos de outra forma, a fim de facilitar o trabalho do tradutor. A solução para esse problema é utilizar as tabelas e régua do Word. Compondo diversas linhas e colunas de tabela, forma-se uma grade na qual o texto da imagem pode ser disposto numa configuração que se assemelhe à da imagem. Utilizar a régua facilita o processo de visualização do texto e, conseqüentemente,

o processo de tradução, já que o uso da tecla “Enter” é proibido nesses casos, pois o memoQ iria separar a frase em segmentos, podendo causar confusão ao termos uma expressão dividida em dois segmentos, levando a uma má tradução. A tabela 3 abaixo apresenta as possibilidades de utilização de régua e de imagem com texto no processo de visualização:



Imagem	Nomenclatura
	[selo: 100% GARANTIA DE SATISFAÇÃO]
	[carimbo: A-0087 27.04.2018·16 DÜSSELDORF· IMMIGRATION]

Tabela 3. Exemplos de imagens com texto.

Um tipo de imagem que costuma causar muitos problemas é o gráfico. A melhor forma de o reproduzir, de modo a evitar a criação de “tags” no memoQ, é utilizando uma tabela do Word. As figuras 6 e 7, a seguir, ilustram um exemplo de gráfico criado utilizando tabelas no Word:

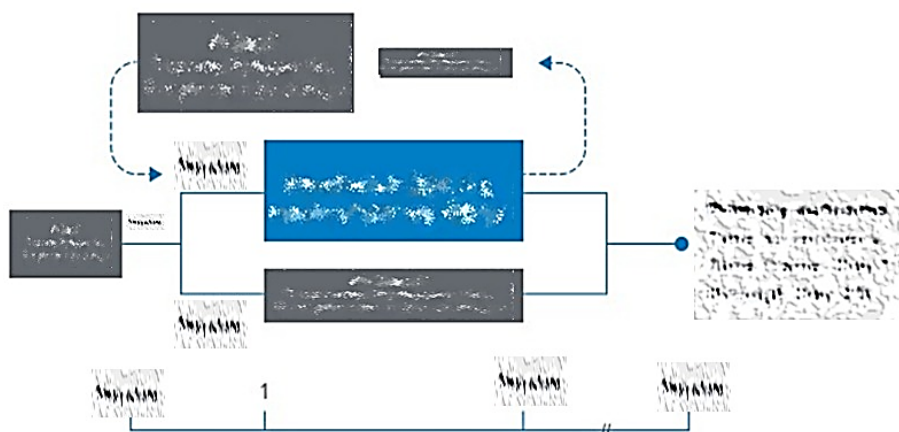


Figura 6. Gráfico (com texto protegido) do documento de origem.

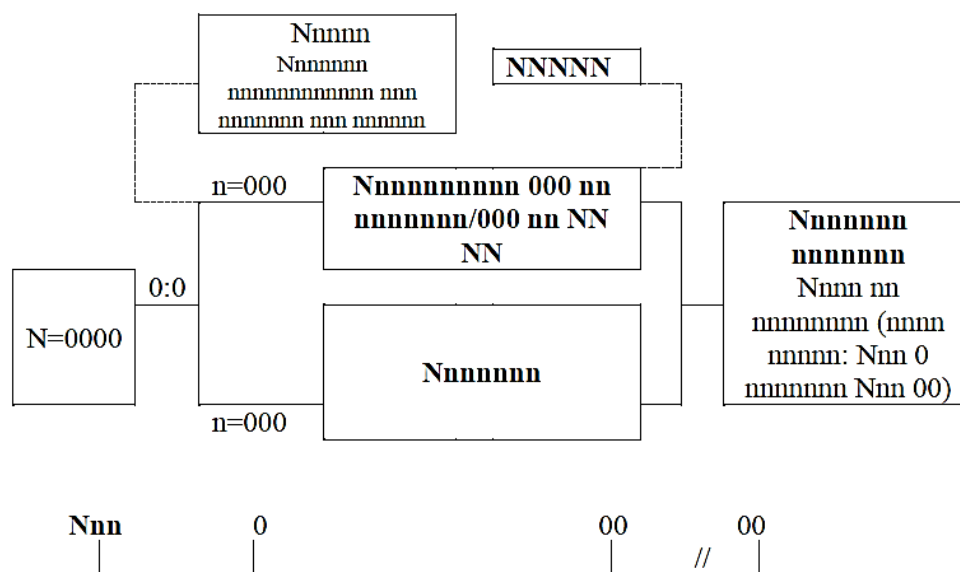


Figura 7. Gráfico reproduzido (com texto protegido) utilizando a tabela do Word.

A Figura 6 foi substituída pela tabela ilustrada na Figura 7. A tabela contém todos os textos da figura original. Buscou-se uma disposição espacial dos textos na tabela semelhante à dos textos na figura de origem.

Terceira etapa

Nesta última etapa, deve-se fazer a revisão da ortografia, uma vez que é muito comum que o OCR cometa erros de troca de caracteres, como a letra “i” maiúscula (I) pela letra “l” (L) minúscula, ou pelo número um (1), ou até mesmo pelo ponto de exclamação (!), que são muito comuns. Ver exemplo na tabela 4:

Texto no PDF	Formato no Word
9-514 Lazarim 46102 - E-mail: secretaria@	-514 Lazarim ;102 - E-mail: <u>so</u> cretaria@

Tabela 4. Exemplos de erros do OCR com a letra “i” trocada pela letra “l”, e a letra “e” trocada pela letra “o”.

Outras trocas comuns que resultam de erros do OCR são referentes à letra “o”, que, por vezes, se transforma no número zero (0), a letra “g”, que pode ser trocada pelo número oito (8), e a letra “m”, que pode aparecer como a combinação das letras “rn”, e vice versa. A tabela 5, abaixo mostra um exemplo:

Formato no PDF	Formato no Word
G ^a represents -CONR ^{7a} -, -NR ^{7a} CO-, -SO ₂ NR ^{7a} -, -NR ^{7a} SC	G ^a <u>represents</u> -CONR ^{7a} -, -NR ^{7a} CO-, -SO ₂ NR ^{7a} -, -NR ^{7a} SI

Tabela 5. Exemplos de erros do OCR com algumas letras “O” trocadas pelo número “0” (zero).

Por fim, algumas partes do texto podem parecer ilegíveis – seja devido a alguma mancha no documento, à má qualidade do escaneamento, ou até mesmo, nos textos manuscritos, à caligrafia da pessoa que escreveu ser difícil de decifrar. Nesses casos, se não houver meio de obter o documento numa qualidade melhor, deve-se pôr a marcação [ilegível] no local da palavra ou frase.

Todos estes procedimentos de ajuste acima descritos são necessários para que a tradução transcorra de maneira mais eficiente e com menor risco de erros. Fica claro que é fundamental conhecer em profundidade estas ferramentas de TI, tais como o Word e o Adobe Acrobat. Como diz Zélia Neves (comunicação pessoal, 18 de setembro de 2020):

Convém perdermos cada vez menos tempo com esta tarefa – para isso, é preciso perder muito tempo nos primeiros projetos em que estamos a preparar documentos. O melhor truque é conhecer muito bem o Word e, assim, saber utilizar todas as suas funções, para pouparmos tempo e criarmos documentos o mais limpos possível. Propostas de melhoria para o memoQ para a etapa pré-tradução.

4. Propostas de melhoria para o memoQ para a etapa pré-tradução

Esta seção apresenta ideias de melhorias do memoQ, com base na experiência adquirida no meu estágio – e problematizada na seção 3 –, e consolida o objetivo central deste trabalho acadêmico. Apesar de me concentrar nas tarefas de pré-tradução, as propostas aqui descritas incidem no aperfeiçoamento do serviço como um todo – tradução, formatação e revisão – e são referentes ao uso prático da ferramenta memoQ em ambiente de trabalho profissional. Optei por fazer sugestões abrangentes e não apenas para a fase de pré-tradução por serem sugestões que melhoram a utilização do software em todas as fases, incluindo a pré-tradução. Espero que o presente artigo possa contribuir para preencher uma lacuna existente na literatura sobre formatação e suas ferramentas especializadas.

4.1 Múltiplos aparelhos

Como o memoQ é um *software* executável, precisa ser instalado em um computador, ficando limitado ao uso apenas nesta mesma máquina. As funções de tradução ficam reduzidas a um único local. O ideal seria que o tradutor pudesse acessar o memoQ em qualquer aparelho e não apenas num único computador. Nesse sentido, uma melhoria fundamental seria poder acessar via internet (*browser* ou aplicativo) o memoQ, com todas as configurações do perfil do usuário. A empresa já está caminhando nesse sentido, tendo em conta a existência de versões online para os clientes que adquiram um servidor (memoQ WebTrans) e ainda a oferta de uma versão online reduzida. No entanto, claramente, este programa está ainda baseado na versão *desktop*.

Sabe-se que uma das alegações apresentadas que justificam o modelo de acesso restrito a apenas um computador repousa na ideia central de ser uma estratégia de segurança dos dados. No entanto, há várias modalidades de uso de *browser* sem quebra

de segurança, completamente online e bastante utilizada no mundo todo. Diversos autores, entre eles Sin-wai (2017, p. 14), analisam os serviços baseados em nuvem:

Due to the rapid development of new information technologies, most computer-aided translation systems during this period are server-based, web-based, and even cloud-based computer-aided translation systems, which have a huge storage of data. By the end of 2012, there were fifteen cloud-based computer-aided translation systems available on the market aimed at individuals or enterprises, such as Lingotek Collaborative Translation Platform, SDL World Server, and XTM Cloud.

Como contraponto ao alegado problema de segurança, observe-se que praticamente não existe organização com maior exigência de sigilo e segurança do que as forças armadas americanas. No entanto, o Ministério da Defesa dos Estados Unidos (Pentágono) decidiu colocar todos os seus dados em nuvens privadas. A Amazon, o Google, a Microsoft e a Oracle foram solicitadas a apresentar propostas para o projeto JWCC (Joint Warfighting Cloud Capability), para este projeto (Swartz, 2021).

Nesta linha de sugestões de melhoria, acrescenta-se a proposta de acesso nativo via aplicativo em *tablets*, *smartphones* e *notebooks touchscreen* com todas as funcionalidades existentes na versão *desktop*.

4.2 O memoQ e os sistemas operacionais da Apple

O memoQ não funciona nativamente na plataforma Apple. Isto é um problema para vários usuários do memoQ que possuem apenas equipamentos Apple.

A solução adotada até hoje por esses usuários é criar uma máquina virtual Windows dentro do equipamento Apple. Porém, essa solução, além de trabalhosa, não é tecnicamente fácil para a grande maioria das pessoas leigas em TI.

Uma solução muito mais simples para esse problema seria o memoQ disponibilizar uma versão completa e totalmente funcional na nuvem, visto que “all cloud-based systems can be run on Mac OS and OS X” (Sin-wai, 2017, p. 46). De fato, se o memoQ fosse baseado em nuvem, poderia funcionar com tranquilidade nos computadores da Apple – incluindo iPad e iPhone – via *browser*, sem precisar de máquina virtual. Porém, se utilizado via aplicativo, então teria de ser reprogramado para a plataforma Apple.

4.3 Assistente de voz

Dando seguimento aos aspectos de incremento do memoQ, apresento outra proposta de melhoria: a implementação de assistente de voz tanto para o comando das funcionalidades do memoQ como para a redação da tradução de documentos.

Conforme diz Sin-wai (2017, p. 75):

The use of speech as an inputting method is important in several ways. First, speech input is natural because it requires neither special equipment, but a microphone, nor special training or abilities. Second, speech input is convenient, as it allows fast and comfortable communication at a rate of 120 to 210 English words per minute without encumbering the

hands or eyes. Third, speech input is universal because nearly everybody can speak, except those who are vocally handicapped.

Essa implementação resolveria o problema de teclados estrangeiros ou de tamanho reduzido. Um outro importante problema que essa funcionalidade resolveria seria o da acessibilidade: o público que apresenta deficiência visual ou mobilidade reduzida poderia acessar o memoQ com muito mais facilidade. Portanto, tal melhoramento promoveria a inclusão social de tradutores e poderia também ajudar a evitar lesões, como por exemplo a síndrome da lesão do esforço repetitivo (LER).

Diversas plataformas digitais já possuem seu próprio assistente de voz (ver figura 8):

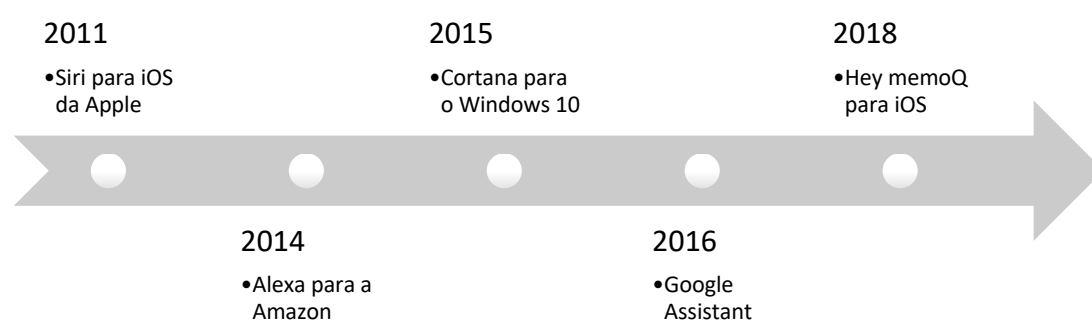


Figura 8. Linha do tempo dos assistentes de voz.¹

Em 2018, foi lançado o memoQ 8.7, cuja maior novidade foi a funcionalidade chamada de Hey memoQ (memoQ ©, 2021) para suporte a comando e ditado por voz. Este recurso, porém, é baseado no uso do assistente Siri da Apple (Apple Inc. ©, 2021), portanto só está disponível na plataforma iOS. A empresa declarou que não desenvolveu esta funcionalidade nas demais plataformas pois, segundo ela, os demais assistentes de voz não têm a mesma qualidade e funcionalidades do Siri.

Essas plataformas e seus assistentes de voz vêm, contudo, evoluindo significativamente nos anos recentes e, portanto, espera-se que o memoQ não só reveja essa sua decisão como também expanda a utilização do assistente de voz.

4.4 Precisão da funcionalidade OCR

A funcionalidade OCR do memoQ apresenta diversos problemas, como foi visto na seção 2. Um dos principais entraves desta funcionalidade é a sua falta de precisão.

O OCR tem as seguintes finalidades:

- a) reconhecer o texto da imagem;
- b) conseguir repassar de forma fiel os elementos da imagem para o Word.

O que acontece, porém, com muita frequência no OCR do memoQ são defeitos de precisão nessas duas finalidades. Percebe-se que o memoQ visivelmente não vem investindo no seu OCR como deveria, mantendo-se numa versão ultrapassada e atrasada, com sérios problemas de precisão que chegam a inviabilizar o seu uso.

¹ Apple Inc. ©, 2021, Amazon.com, Inc. ©, 2021, Microsoft ©, 2021, Google, s.d. e memoQ ©, 2021.

Nesta perspectiva, muitos tradutores utilizam outros programas como forma de melhorar a precisão do OCR, tais como o Adobe Acrobat Pro (Adobe, 2021) e o ABBYY FineReader PDF (ABBYY, 2021).

Portanto, percebe-se que o memoQ deve investir na melhoria contínua de seu OCR e/ou permitir a utilização de um *plugin* de OCR desenvolvido por terceiros.

4.5 Melhoria do tratamento e exibição de “tags”

Como foi discutido na seção anterior, alguns tradutores têm utilizado o Adobe Acrobat Pro para suprir a insuficiência da funcionalidade OCR do memoQ. Porém, existem também lacunas na formatação oriunda do OCR do Adobe.

Normalmente, gera-se um arquivo Word a partir desse documento PDF para se trabalhar com ele no memoQ. Quando o memoQ interpreta e exibe o texto do Word, deveria ter um procedimento de agrupar formatações de fonte de palavras sucessivas em uma única formatação do grupo de palavras. Por exemplo, se um trecho de diversas palavras sucessivas estiver com a fonte azul, o memoQ deveria considerar que os espaços entre essas palavras também têm fonte azul. Dessa maneira bastaria uma “tag” no início e outra no final do trecho. O mesmo procedimento se aplica à cor de fundo, cor de realce etc. Isso reduziria significativamente o número de “tags” no memoQ, bem como facilitaria e agilizaria o trabalho de quem estivesse formatando os documentos para serem traduzidos.

4.6 Expansão das funcionalidades de edição do memoQ

As funcionalidades do memoQ concentram-se apenas nas funções de tradução propriamente dita. Porém, o memoQ poderia e deveria permitir que o tradutor trabalhasse nativamente com formatações (pré- e pós-tradução) de textos, de maneira inteligente e integrada com a ação de tradução, transformando-se num ITE (*integrated translation environment*), do mesmo modo, por exemplo, que já existe o IDE (*integrated development environment*) para a comunidade da área de TI, de desenvolvimento de programas.

O ideal seria que o próprio memoQ fosse mais livre do Microsoft Word e que a edição fosse feita dentro da própria ferramenta CAT, inclusive o tratamento de documentos PDF, de modo a, por exemplo, poder interativamente:

- orientar o memoQ a transformar imagens no PDF em tabelas temporárias para evitar “tags”;
- fazer um melhor OCR em arquivos PDF baseados em imagem;
- reconstruir o documento PDF de origem com a tradução, permitindo fazer os ajustes necessários em imagens.

Várias das funcionalidades propostas acima poderiam também ser implementadas via *plugin*. Com a integração das funções de formatação pré- e pós-tradução no memoQ, ganha-se não apenas um ambiente integrado de tradução, mas, principalmente, um ambiente interativo especializado nos problemas específicos do tradutor.

4.7 Plugins do memoQ em outros programas

Várias funcionalidades de edição já propostas anteriormente podem ser implementadas via a utilização de *plugins*; porém, nesta seção, serão discutidos possíveis *plugins* do memoQ para outros programas, especificamente o Microsoft Word e o Adobe Acrobat.

O memoQ poderia criar um *plugin* para o Adobe Acrobat (figura 9), tal que o OCR já fosse adaptado a documentos que vão ser traduzidos em ferramentas CAT. Esse *plugin* poderia incluir um OCR personalizado em que, por exemplo, não fossem criadas caixas de texto, mas tabelas para poder distribuir o texto espacialmente de forma precisa.

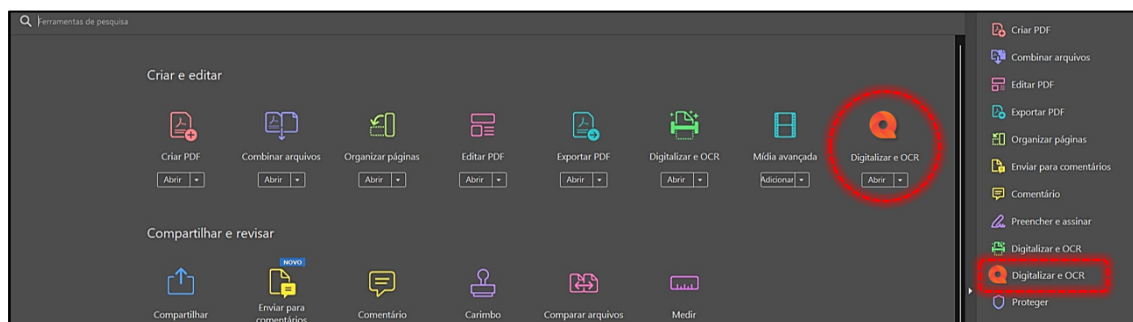


Figura 9. Ilustração de plugin do memoQ no Adobe Acrobat.

O memoQ também poderia criar um *plugin* para o Microsoft Word (figura 10) para auxiliar na formatação dos documentos. O *plugin* poderia conter funções como:

- Otimizar formatações;
- Substituir figuras por *placeholders*, como [FOTO], [SELO], [RÚBRICA], etc.;
- Transformar caixas de texto em tabelas;
- Eliminar quebras incorretas de linha;
- Identificar erros de OCR, como troca de caracteres (“rn” ter virado “m”, ou “O” ter virado “0”, etc.).

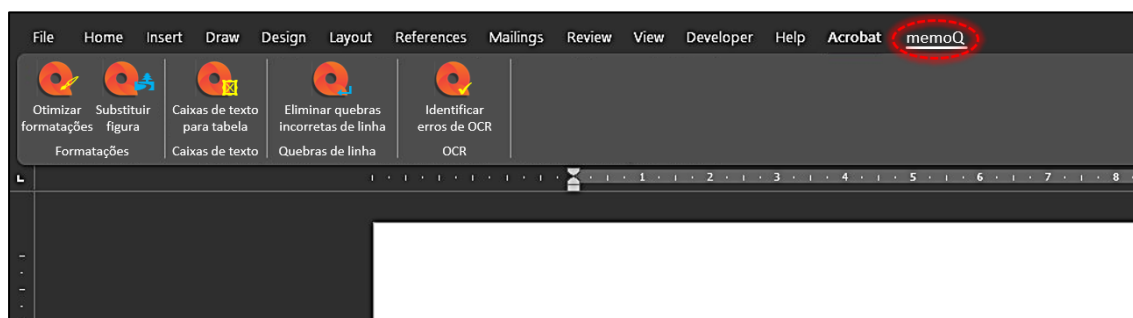


Figura 10. Ilustração de *plugin* do memoQ no Microsoft Word.

4.8 Implementação de diferentes revisores ortográficos no memoQ

Observa-se esta lacuna no memoQ: não há uma grande variedade de revisores ortográficos. Atualmente só existem duas opções, o Microsoft Word (Microsoft ©, 2021) e o Hunspell (Hunspell, s.d.), o que leva muitos tradutores a abrirem mão do revisor nativo

do Word, preferindo o auxílio de outros revisores via *plugins*, ou de outros programas como o FLiP (Priberam Informática, S.A., 2021) e o Grammarly (Grammarly Inc. ©, 2021).

Uma sugestão de melhoria seria a existência de um *plugin* no memoQ que daria acesso aos revisores ortográficos da escolha do tradutor.

5. Conclusão

As ferramentas CAT são relativamente recentes e, portanto, apresentam inúmeras oportunidades de melhorias. A seção 3 concentrou-se precisamente em oito sugestões de melhorias motivadas pelo meu estágio.

As ferramentas CA (*computer-assisted*), noutras áreas, oferecem um nível inicial de qualidade já alto. A partir de toda a comodidade que essas ferramentas oferecem, o usuário pode se concentrar melhor no seu trabalho e, daí, aumentar ainda mais o nível de qualidade do resultado. Logo, quanto mais funcionalidades o memoQ disponibilizar para automatizar, de forma integrada (ITE), tarefas que hoje são efetuadas fora dele, mais o usuário da ferramenta poderá concentrar-se na qualidade da tradução em si.

Assim como as ferramentas CA em geral percorreram uma trilha de constantes progressos, no sentido de integrar num único ambiente de trabalho (*integrated environment*) tanto as funcionalidades principais como as acessórias, sugere-se neste artigo que o memoQ caminhe também nessa direção. Foi neste sentido que o grupo de oito melhorias propostas neste artigo considerou a incorporação no memoQ de funcionalidades que hoje ele não possui. Algumas das melhorias baseiam-se em funcionalidades já existentes no memoQ mas implementadas de forma incipiente, provavelmente devido ao fato de automatizarem atividades acessórias à tradução.

A experiência durante o meu estágio e os trabalhos *freelance*, que se concentraram em atividades acessórias (formatação e revisão de textos), mostrou-me que, de forma geral, o memoQ ainda se destina, acima de tudo, a dar suporte à atividade de tradução em si, não se preocupando em integrar as atividades pré-tradução de forma nativa e eficiente dentro da ferramenta.

Fica então a sugestão de que futuros trabalhos possam não só aprofundar as oportunidades aqui elencadas, como também apontar outras oportunidades de melhorias que possam tornar o memoQ numa ferramenta CAT mais abrangente e integrada (ITE), dando, portanto, um maior e melhor suporte ao processo de tradução em todas suas etapas.

REFERÊNCIAS

- ABBYY (2021) *ABBYY finereader pdf* [online]. Disponível em: <https://pdf.abbyy.com/> (Acesso: 27 maio 2021).
- Adobe (2021) *Adobe* [online]. Disponível em <https://acrobat.adobe.com/> (Acesso: 27 maio 2021).
- Amazon.com, Inc. (2021) *Alexa* [online]. Disponível em: <https://www.amazon.com/b?ie=UTF8&node=21576558011> (Acesso: 7 jun. 2021).

- Chaves, T.L. – O uso do MemoQ nas tarefas de pré-tradução
Translation Matters, 4(1), 2022, pp. 104-124, DOI: https://doi.org/10.21747/21844585/tm4_1a7
- Apple Inc. (2021). *Siri* [online]. Disponível em: <https://www.apple.com/siri> (Acesso: 7 jun. 2021).
- Google (s.d.) (2021) *Hey google* [online]. Disponível em: <https://assistant.google.com/> (Acesso: 7 jun. 2021).
- Grammarly Inc. (2021). *Grammarly* [online]. Disponível em <https://www.grammarly.com/> (Acesso: 27 maio 2021).
- Hunspell (s.d.). (2021) *Hunspell* [online]. Disponível em: <http://hunspell.github.io/> (Acesso: 27 maio 2021).
- jblosser (2011) *Excessive tags when opening Word document in SDL Trados Studio 2009* [online]. Disponível em: https://www.proz.com/forum/software_applications/205493-excessive_tags_when_opening_word_document_in_sdl_trados_studio_2009.html (Acesso: 1 out. 2020).
- memoQ (2020) *memoQ* [online]. Disponível em: <https://www.memoq.com/> (Acesso: 28 Set. 2020).
- memoQ (2021) *Hey memoQ* [online]. Disponível em: <https://www.memoq.com/products/hey-memoq> (Acesso: 7 de jun. 2021).
- Microsoft (2021) *Cortana* [online]. Disponível em: <https://www.microsoft.com/en-us/cortana> (Acesso 7 jun. 2021).
- Microsoft (2021) *Microsoft Word* [online]. Disponível em: <https://www.microsoft.com/en-ww/microsoft-365/word> (Acesso: 7 jun. 2021).
- Mossop, B. (2019) *Revising and editing for translators* [online]. London: Routledge. Disponível em: <https://doi.org/10.4324/9781315158990> (Acesso: 27 Maio 2021)
- Peng, H. (2020) On the operability of computer-aided translation software. *IOP Conference Series: Materials Science and Engineering*, 768, 072033 [online]. Disponível em: <https://doi.org/10.1088/1757-899x/768/7/072033> (Acesso: 27 Maio 2021)
- Pietrzak, P. & Kornacki, M. (2021) *Using CAT tools in freelance translation*. New York: Routledge.
- Priberam Informática, S.A. (2021) *FLiP* [online]. Disponível em: <https://www.flip.pt/> (Acesso: 27 maio 2021).
- Redação do Olhar Digital (2020) *Olhar digital* [online]. Disponível em: <https://olhardigital.com.br/2020/09/09/pro/pentagono-confirma-contrato-com-a-microsoft-para-computacao-em-nuvem/> (Acesso em: 29 maio 2021).
- Sin-wai, C. (2017) *The future of translation technology*. Abingdon, United Kingdom: Routledge.
- Swartz, J. (2021) *The return of JEDI: why the sequel to military's cloud contract could cost much more than the \$10 billion original* [online]. Disponível em: <https://www.marketwatch.com/story/the-return-of-jedi-sequel-to-military-cloud-contract-will-be-much-more-complex-11633640142> (Acesso: 3 fev. 2022).
- Wikipedia (2020) *Joint enterprise defense infrastructure* [online]. Disponível em: https://en.wikipedia.org/wiki/Joint_Enterprise_Defense_Infrastructure (Acesso: 7 jun. 2021).
- Wikipedia (2021) *Integrated development environment* [online]. Disponível em: https://en.wikipedia.org/wiki/Integrated_development_environment (Acesso em: 3 de fev. 2022).

Sobre a autora: Tatiana Limongi Chaves é Mestre em Tradução pela NOVA FCSH e graduada em Licenciatura em Letras pela Universidade Católica de Pernambuco. Com base no seu relatório de estágio, está a iniciar uma investigação sobre preparação de textos para tradução.