



UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE LETRAS – IL
DEPARTAMENTO DE LÍNGUAS ESTRANGEIRAS E TRADUÇÃO – LET
LETRAS TRADUÇÃO – ESPANHOL

MOISÉS MONTEIRO PACHECO WANZELLER

TECNOLOGIAS DE APOIO À TRADUÇÃO: TUTORIAL DE OCR

Brasília – DF
2018

UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE LETRAS – IL
DEPARTAMENTO DE LÍNGUAS ESTRANGEIRAS E TRADUÇÃO
LETRAS TRADUÇÃO – ESPANHOL

Moisés Monteiro Pacheco Wanzeller

TECNOLOGIAS DE APOIO À TRADUÇÃO: TUTORIAL DE OCR

Projeto Final do Curso de Tradução, apresentado como requisito parcial à obtenção do grau de Bacharel em Letras Tradução – Espanhol pela Universidade de Brasília (UnB).

Orientadora: Prof^a. Magali de Lourdes Pedro

Wanzeller, Moisés Monteiro Pacheco

Tutorial para auxiliar a assimilação de conteúdos de uma oficina sobre OCR - Optical Character Recognition

Projeto Final de Curso (bacharelado) – Universidade de Brasília, Instituto de Letras, 2018.

Orientadora: Prof^a. Magali de Lourdes Pedro

1. Estudos da Tradução. 2. Tradução. 3. Tecnologia. 4. Subcompetência instrumental

Folha de aprovação

TECNOLOGIAS DE APOIO À TRADUÇÃO: TUTORIAL DE OCR

Projeto Final do Curso de Tradução julgado como requisito parcial para a obtenção do grau de Bacharel em Letras Tradução – Espanhol.

Área de Concentração: Tradução e Tecnologia

Moisés Monteiro Pacheco Wanzeller

Projeto Final aprovado em: _____ / _____ / _____

Prof^a. M.Sc. Magali de Lourdes Pedro
(Orientadora – LET/UnB)

Banca Examinadora: _____
Prof^a. Dr^a. Sandra María Pérez López

Banca Examinadora: _____
Prof^a. M.Sc. Marta Molina

AGRADECIMENTOS

Em primeiro lugar, agradeço a Deus e a minha família por me darem forças para chegar até aqui. Aos meus pais, Miriam e Reginaldo, por me inspirarem com toda a sua luta e dedicação para me fornecer o melhor, pelo amor e por todo o cuidado que fizeram com que esse momento se tornasse realidade. Vocês são mais do que meus pais, são a minha mais forte e segura certeza da existência de um Deus.

Certa vez, Vinicius de Moraes disse: “a gente não faz amigos, reconhece-os”. Certamente, durante essa jornada, os reconheci. A Beatriz Siqueira, Karina Brito e Nathany Araújo por todas as risadas, conversas e conselhos que foram parte vital não somente desta monografia, mas de uma vida inteira. A vocês, meu muito obrigado.

Ao meu amor, Eduarda Marques, por toda a paciência e amor. Sem a sua presença, seu companheirismo e seu cuidado nada disso seria possível. Sua importância é imensurável.

A todos os meus amigos do curso de Tradução Espanhol pelo companheirismo durante essa jornada, sem vocês tudo se tornaria mais difícil.

Ao corpo docente da Universidade de Brasília, registro o meu muito obrigado pela generosidade em compartilhar conhecimento de forma ímpar, em especial a orientadora desse projeto, Magali Pedro, por seu exemplo, paciência (muita!) e dedicação. Você é uma figura que inspira!

A banca examinadora dessa monografia, obrigado pela disposição em aceitar o convite para fazer parte dessa etapa, certamente todas são figuras essenciais na minha formação.

A todos que de alguma forma contribuíram para a conclusão deste projeto.

“Estamos na situação de uma criancinha que entra em uma imensa biblioteca, repleta de livros em muitas línguas. A criança sabe que alguém deve ter escrito aqueles livros, mas não sabe como. Não compreende as línguas em que foram escritos. Tem uma pálida suspeita de que a disposição dos livros obedece a uma ordem misteriosa, mas não sabe qual ela é”.

(Albert Einstein)

RESUMO

O presente trabalho, que constitui o Projeto Final exigido como requisito parcial para a obtenção do grau de Bacharel em Letras/Tradução Espanhol, da Universidade de Brasília (UnB), consiste na apresentação de um tutorial sobre o uso de um *software* de OCR, para auxiliar na assimilação dos conteúdos de uma oficina dedicada ao assunto. São apresentadas suas origens, funções e importância, interligando as tecnologias de apoio à tradução com o ofício em seu último estágio, isto é, o traduzir.

Palavras-chave: Estudos da Tradução. Tradução e Tecnologia. Subcompetência instrumental. OCR.

RESUMEN

El presente trabajo, que constituye el Proyecto Final exigido como requisito parcial para la obtención del grado de Bachiller en Letras / Traducción Español, de la Universidad de Brasilia (UNB), consiste en la presentación de un tutorial sobre el uso de un software de OCR, para auxiliar en la asimilación de los contenidos de un taller dedicado al tema. Se presentan sus orígenes, funciones e importancia, interconectando las tecnologías de apoyo a la traducción con el oficio en su última etapa, es decir, el traducir.

Palabras clave: Estudios de la Traducción. Traducción y Tecnología. Subcompetencia instrumental. OCR.

LISTA DE ABREVIATURAS E SIGLAS

ALT - Alternativa

CTRL – Control

IMR - Intelligent Machines Research Corporation

OCR – Optical Character Recognition (Reconhecimento Ótico de Caracteres)

PAT – Programa de Apoio a Tradução

MT – Template Memory (Memórias de Tradução)

TO – Texto Original

TS – Texto Simples

UnB – Universidade de Brasília

LISTA DE ILUSTRAÇÕES

Figura 1 – Interface do *Wordfast*

Figura 2 – Caminho para Identificar o arquivo para o OCR

Figura 3 - Atalho para abrir o documento direto no software *Abbyy Finereader* versão 12

Figura 4 - Editor de OCR *Abbyy Finereader* versão 12 *Profesional*

Figura 5 – Caminhos para exportar o arquivo para um documento em *word*

Figura 6 – Documento recém enviado para o *word* pelo programa *Abbyy*

Figura 7 – Ícone para limpar todo o documento

Figura 8 – Ícone para limpar todos os indicadores do documento

Figura 9 – Margens personalizadas configuradas

Figura 10 – Caminho para personalizar as margens

Figura 11 – Caminho para definir a fonte, seu tamanho e a linguagem do documento

Figura 12 – Ícones para alterar o layout do texto

Figura 13 – Ícones para adicionar negrito, itálico e sublinhado, respectivamente.

Figura 14 - Ícone para mostrar espaçamento e imagens que demonstram a quantidade utilizada, respectivamente.

Figura 15 – Caminho para localizar, substituir um ou vários arquivos de uma só vez

Figura 16 – Passo a passo para remover hiperlinks do texto

Figura 17 – Ícone para selecionar bordas

Figura 18 – Caminho para salvar e renomear o arquivo

Figura 19 – Documento recém enviado ao *Word* pelo *Abbyy*

Figura 20 – Documento com o OCR finalizado

SUMÁRIO

| | |
|--|----|
| RESUMO..... | 7 |
| RESUMEN..... | 8 |
| LISTA DE ABREVIATURAS E SIGLAS..... | 9 |
| LISTA DE ILUSTRAÇÕES..... | 10 |
| INTRODUÇÃO | 12 |
| CAPÍTULO I:..... | 14 |
| 1.1. A competência tradutória e o uso de tecnologias de apoio à tradução | 14 |
| 1.2. OCR | 15 |
| 1.3. Os softwares e a tradução | 15 |
| 1.4. As memórias de tradução e o OCR: tecnologias interligadas | 18 |
| CAPÍTULO II:..... | 21 |
| PASSO A PASSO – TUTORIAL PARA OCR..... | 21 |
| CONSIDERAÇÕES FINAIS | 37 |
| REFERÊNCIAS BIBLIOGRÁFICAS | 38 |

INTRODUÇÃO

Na história da civilização humana, a prática tradutória sempre esteve presente. Como em praticamente todas as áreas de estudos, o entorno e o núcleo da tradução evoluíram e se desenvolveram com o passar dos anos. James Holmes (VASCONCELLOS; BARTHOLAMEI JUNIOR, 2009) foi um dos pioneiros a perceber que as complexidades do mundo tradutório exigiam uma oferta específica. Por meio de uma comunicação oral num congresso em Estocolmo, em 1972, os estudos da tradução passaram a ter o caráter de disciplina, isto é, área de conhecimento. Dentro dessa perspectiva, a tradução se desenvolveu com o passar dos anos como área de conhecimento específico, e conquistou espaço no mundo universitário, com disciplinas que compreendem tanto aspectos teóricos como práticos dessa atividade.

Ao mesmo tempo, os avanços tecnológicos que se deram ao longo do século XX, impactaram a atividade profissional da tradução, a tal ponto que é praticamente impossível ignorar sua presença nesse âmbito. É o que corrobora Vales, ao afirmar que:

As tecnologias de apoio à tradução são frequentemente associadas a uma ideia de modernidade da profissão, mas também integram as competências necessárias para o sucesso profissional, constituindo uma exigência de muitas instituições multilíngues relevantes como as integrantes da União Europeia, as Nações Unidas e outras similares. (VALES, 2016, p. 1).

Contudo, segundo Danilo e Vera Nogueira (2004) há uma certa resistência na aceitabilidade do uso de programas de apoio à tradução por parte dos profissionais:

Quem se recusa a usar programas de apoio à tradução (PAT) costuma alegar que esses programas podem ser muito úteis para os outros, mas, para o “seu” tipo de tradução, tem pouca serventia. Entretanto, os PATs são úteis, para não dizer indispensáveis, para todos os tradutores e tipos de tradução. (NOGUEIRA, DAN; NOGUEIRA, VER; 2004, p. 18).

A universidade, ao mesmo tempo em que preserva sua vocação acadêmica de reflexão sobre o traduzir, não ignora as necessidades dos tradutores em formação com relação ao mercado de trabalho. Nas palavras de Souza (2017, p. 25), “(...) existe, entre a preparação universitária e a prática do tradutor, um descompasso marcado por movimentos sinuosos de aproximações e distanciamentos e não por um distanciamento sem limites.”

Nesse cenário de diálogo crítico, o que motivou o presente trabalho foi um curso relativamente novo: o Bacharelado em Letras – Tradução Espanhol da Universidade de Brasília – UnB. Ao longo do curso, na condição de estudante, fui convidado a colaborar, como instrutor, durante jornadas acadêmicas e apresentações em semanas universitárias. Isto deveu-se à minha atividade profissional: editor de OCR (*Optical Character Recognition*) para tradução juramentada e revisor de tradução em empresa privada.

As oficinas levadas a cabo durante essas jornadas tinham seis horas de duração e um número máximo de 15 participantes. Apesar de o tempo ser adequado para a apresentação da funcionalidade, alguns alunos solicitavam os *slides* apresentados para lembrar dos procedimentos.

Dessa forma, o objetivo geral deste trabalho é auxiliar na aprendizagem e assimilação do uso de recursos tecnológicos de apoio à tradução e, como objetivos específicos:

- Propor a elaboração de um manual com passo a passo para a criação de OCR voltado para a tradução.
- Criar em arquivos de áudio e vídeo um tutorial que ensine de modo prático e voltado para as necessidades do mercado de trabalho e da academia como desenvolver um trabalho de OCR com a maior qualidade no menor prazo possível.
- Utilizar *softwares* de OCR, mais precisamente o programa *ABBYY FineReader*, para auxiliar na confecção dos arquivos.

Assim, o primeiro capítulo discorre brevemente sobre o uso de tecnologias na tradução relacionando-o à competência tradutória, assim como sobre o uso de *softwares* na tradução. O segundo capítulo contém o tutorial para o uso de OCR e, a seguir, apresentamos nossas considerações finais.

CAPÍTULO I:

1.1. A competência tradutória e o uso de tecnologias de apoio à tradução

Nos Estudos da Tradução, tem sido realizado um esforço no sentido de estabelecer quais seriam as “competências” do tradutor, noção que emerge da distância percebida entre a formação e a prática profissional (SOUZA, 2017, p. 12). No Brasil, o trabalho de Gonçalves e Machado (2008, pp. 53-56) apresenta um mapeamento das competências apresentadas pela literatura especializada, como vemos a seguir:

1. Competência linguística na língua materna.
2. Competência linguística prévia na(s) língua(s) estrangeira(s).
3. Competência linguística a ser desenvolvida na(s) língua(s) estrangeira(s).
4. Competência pragmática e sociolinguística na língua materna.
5. Competência pragmática e sociolinguística na(s) língua(s) estrangeira(s).
6. Conhecimento de ambas as culturas das línguas de trabalho.
7. Conhecimentos temáticos (referentes a áreas especializadas de conhecimento)
8. Terminologia.
9. Conhecimentos declarativos sobre tradução.
10. Conhecimento relacionado à prática profissional.
11. Conhecimentos relacionados ao uso de fontes de documentação.
12. Tecnologias que podem ser aplicadas à tradução.
13. Conhecimentos operativos/procedimentais sobre tradução.
14. Aspectos cognitivos.
15. Aspectos metacognitivos.
16. Conhecimentos contrastivos.
17. Aspectos emocionais/subjetivos.

Os autores, citando um estudo de Gonçalves (2003), incluem o item 12 – Tecnologias que podem ser aplicadas à tradução, como subcategoria do item 10 –

Conhecimento relacionado à prática profissional, sob a denominação de competência instrumental/profissional (GONÇALVES; MACHADO, 2008, p. 55).

Em trabalho posterior, Gonçalves (2015) reformulou sua proposta inicial, e o item referente ao uso de tecnologias surge nos seguintes termos:

Habilidade no uso de tecnologias aplicadas à tradução – habilidade de coordenar e otimizar a utilização de diversas ferramentas, especialmente as tecnológicas, a serviço do tradutor, tais como editores de texto, memórias de tradução, tradutores eletrônicos, glossários e dicionários (eletrônicos ou impressos), programas informáticos de busca e pesquisa, materiais de referência e textos paralelos de diversas fontes etc. Com relação às disciplinas mais comumente relacionadas a esta capacidade, disciplinas de tecnologias aplicadas à tradução, metodologias de tradução, tradumática etc. são as mais representativas. (GONÇALVES, 2015, p. 120).

No próximo tópico, trataremos brevemente das origens do OCR, uma funcionalidade que está presente em várias ferramentas de apoio à tradução, e cujo ensino do manejo é foco deste trabalho e, a seguir, trataremos do uso de *softwares* na tradução.

1.2. OCR

O OCR, segundo relatos, já existia e era praticado há vários anos; contudo, só se tornou objeto de pesquisas e registros em 1950, quando David Shepard e Louis Tordella pesquisavam uma forma de automatizar dados da Agência de Segurança das Forças Armadas que, em alguns anos, se tornaria a Agência de Segurança Nacional Americana (NSA) (HEITLINGER, 2007, p. 43). Auxiliados por Harvey Cook, Shepard e Tordella criaram o primeiro *software* de OCR, conhecido como Gismo e, após esse passo inicial, Shepard fundou a IMR, empresa que se tornou pioneira no desenvolvimento de *softwares* comerciais de OCR. Após 3 anos do primeiro *software*, uma gigante da área de tecnologia da época adquiriu uma licença da empresa de David e desenvolveu seu próprio *software* que ficou conhecido como *Optical Character Recognition* surgia, dessa forma, a sigla OCR.

1.3. Os softwares e a tradução

Segundo Lia Wyler (2003), o cenário profissional para tradução apresenta-se da seguinte forma:

São tradutores de obras literárias e técnicas para editoras; tradutores assalariados e autônomos que traduzem textos de circulação interna em empresas comerciais e públicas; intérpretes e tradutores de conferências; tradutores públicos e intérpretes comerciais; tradutores de peças teatrais; tradutores de letras de músicas; tradutores para legendas de filmes; tradutores para dublagem de filmes e vídeos; tradutores que transcrevem fitas gravadas; e, mais recentemente, tradutores de sites da Internet e tradutores especializados em localização. (WYLER *apud* MARTINS, 2006, p. 27).

Nos casos citados por Lia Wyler (2003) no tocante à questão da necessidade da leitura óptica de caracteres para a tradução, é possível inferir que há a clara necessidade de OCR para otimizar a tradução. Contudo, porque há necessidade de OCR em alguns casos e em outros não?

A tradução interpretação simultânea é um ramo da tradução que possui algumas particularidades se comparada à tradução tradicional, vista pelo senso comum como aquela que se utiliza o papel como principal suporte material do trabalho. No caso da interpretação simultânea, a dinâmica é muito maior e o tempo de resposta ao que está sendo apresentado como texto original (TO) é praticamente instantâneo, isto é, o que é dito é traduzido em tempo real, não abrindo margem para nenhum tipo de análise mais detalhada do processo. Outra singularidade dessa modalidade de tradução é o formato em que o TO é apresentado. Como, nesse caso específico, tratamos de uma língua de partida que está em formato oral, não há nenhuma necessidade de mudança de formato para otimizar a interpretação.

Se em interpretações simultâneas não há a necessidade de adequar um formato de texto, como as outras modalidades de tradução, em geral escritas, geram demanda para a atividade de OCR? Para estabelecer um parâmetro para essa pergunta, é necessário entender a ordem temporal e cronológica da civilização contemporânea e, por consequência, da tradução. Quando James Holmes (1972-1988) percebeu que a tradução alcançara o status e a demanda necessária para converter-se em uma disciplina autônoma a necessidade de uma preparação de originais (OCR) não estava em discussão naquele momento, apesar de a prática de preparação de originais acontecer desde meados da década de 50, com o Gismo, o primeiro *software* de tradução. O debate sobre a importância da tradução como ciência e disciplina naquela época não incorporava qualquer discussão acerca de OCR, sigla criada alguns anos antes. A preparação de originais como conhecemos hoje se tornou pauta de discussão com a ascensão dos chamados *softwares*, ou seja, programas de computador destinados a determinadas finalidades. Com a

aparência desses *softwares*, a tradução passou a um patamar mais elevado, deixando de lado o papel e a caneta e passando a utilizar-se de programas como, por exemplo, o Microsoft Office *Word*, *software* de edição de texto criado em 1983 por Richard Brodie, mas somente integrado ao pacote office como conhecemos hoje em 1997.

Com o desenvolvimento de novos programas para a novidade que era o computador à época, em 1993, outra empresa de tecnologia, *Adobe Systems*, cria o formato de arquivo PDF (*Portable Document Format*). O formato PDF permite que arquivos possam ser lidos em PDF independente do formato original daquela ferramenta, isto é, de modo prático, na tradução contemporânea não importa o formato em que o documento original chega às mãos do tradutor, visto que com a tecnologia da *Adobe Systems* é possível visualizá-lo sem maiores empecilhos.

A produtividade foi significativamente aumentada devido a esses programas, o que permitiu a criação de novas demandas e novas formas de traduzir. Apesar de o PDF facilitar o dia a dia do tradutor, havia um obstáculo: o formato foi desenvolvido para visualização e não permitia nenhuma forma de edição. Surge, dessa forma, uma combinação que dura até o presente momento com a junção das ferramentas *Word* e PDF. A facilidade em visualizar diferentes formatos de arquivos em apenas um, com o PDF, foi complementada pela função de editor de textos do Microsoft Office *Word*, ou seja, agora era possível além de visualizar os arquivos de maneira muito mais simples, editá-los a partir do mesmo computador, fazendo, assim, com que a velocidade de tradução, a tomada de decisões e a dinâmica para editar um trabalho se tornasse bastante mais simples.

Contudo, os dois novos *softwares*, tanto o Adobe PDF como o Microsoft Office *Word* não contavam com uma tecnologia de armazenamento de dados que pudessem ser usados mais adiante em futuros trabalhos, tornando, dessa maneira, a correspondência de segmentos textuais inseridos no mesmo contexto uma tarefa única e exclusiva da mente e capacidade de memória e associação de cada tradutor. Desse modo, o que garantia que a palavra “computador” seria traduzida por “computadora” em um serviço de tradução do português para o espanhol, onde todos os elementos textuais permitissem essa repetição, era exclusivamente a capacidade de análise do tradutor e de memória para lembrar que, em outra situação semelhante e/ou igual, o termo usado fora exatamente aquele.

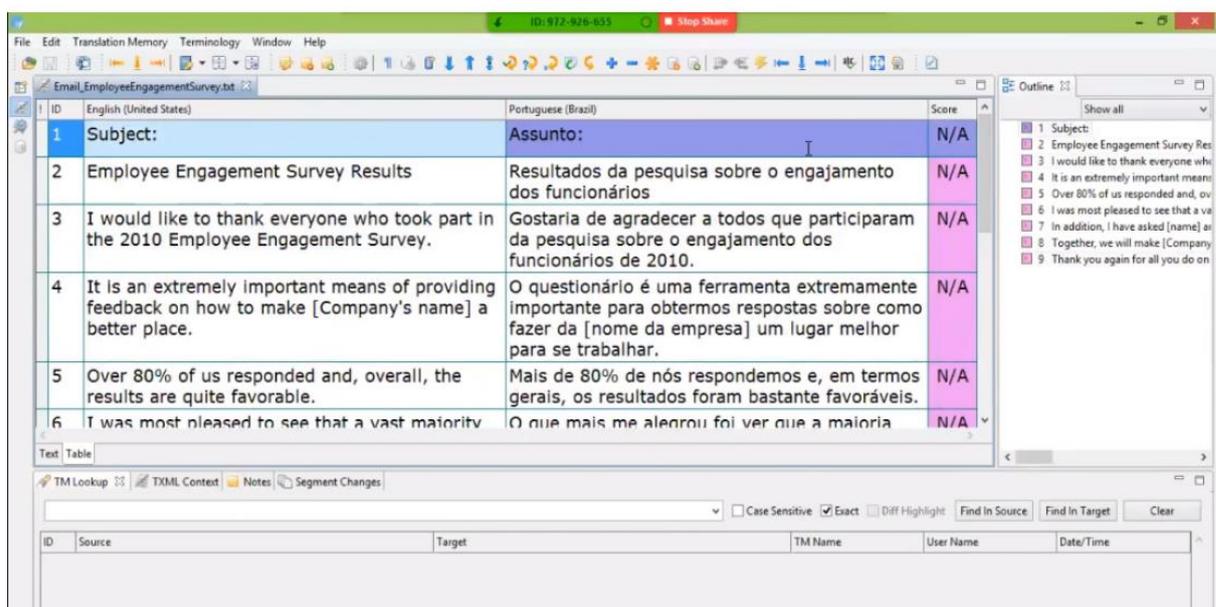
1.4. As memórias de tradução e o OCR: tecnologias interligadas

Diante do exposto anteriormente, uma nova nomenclatura emerge e se solidifica em meio à tradução: as ferramentas CAT – *Computer Aided Translation*. As CAT podem ser consideradas todo e qualquer programa de computador que auxilie de maneira mais direta a tradução. Com a explosão desse tipo de ferramenta no mundo moderno, tradutores e agências de tradução ganham um novo aliado para o dia a dia do ofício e surgem, então, as memórias de tradução. As memórias de tradução (MT) figuram como uma das principais funções de uma ferramenta CAT e têm como função principal salvar as unidades de tradução numa base de dados de maneira que sua reutilização seja possível em outros ou até no mesmo texto como forma de consulta, por exemplo: segundo RIECHE (2006, *apud* LIPARINI *et al.*, 2011) o uso dessas ferramentas tem aumentado cada vez mais no mercado de trabalho. Com avançadas funções de busca dentro dessas ferramentas, se torna possível encontrar segmentos com até 100% de similaridade, transformando a tradução em um material mais coerente, consistente e padronizado. Essas funções de busca e de armazenamento das MT pavimentam um caminho que nem o *Word* nem o Adobe PDF foram capazes de solucionar: a correspondência de determinados trechos e sentenças deixava de depender exclusivamente da memória individual de cada tradutor. Além da consulta imediata de como determinadas nomenclaturas e segmentos textuais haviam sido traduzidos em contextos parecidos anteriormente, as CAT permitem o armazenamento e criação de glossários, inclusive separando-os por temas, função bastante útil em vista da economia de tempo e precisão conquistadas. Mediante os benefícios evidentes trazidos por essas tecnologias, as MT tornaram-se indispensáveis no mundo contemporâneo da tradução, contudo a escassez de conhecimento a respeito dessas tecnologias não paira apenas sobre técnicas e conceitos a respeito do OCR, mas também sobre as CAT e MT.

Dentro das CAT, uma das ferramentas de auxílio à tradução que mais se popularizou foi o *Wordfast*. O *Wordfast* consiste em um programa de tradução projetado para possuir total sincronia e compatibilidade com o Microsoft Office *Word*, auxiliando o tradutor a trabalhar com documentos em formato editável, como, por exemplo, DOC e DOCX. Obviamente, por se tratar de uma CAT, a MT se faz

presente dentro da gama de opções que o *Wordfast* nos oferece como, também, funções para auxiliar na parte terminológica dos textos que são encontradas frequentemente nos chamados textos técnicos (TT). Além de permitir até três glossários simultâneos, a ferramenta ainda dá acesso a links, bancos de dados terminológicos e dicionários que se encontram armazenados online. Outra utilidade bastante importante das CAT, especialmente do *Wordfast*, são as funções de controle de qualidade (QA), funções essas que têm como carro chefe o corretor gramatical e o verificador de consistência terminológica, isto é, além da marcação das palavras e/ou trechos que o programa considere inadequadas após passar pelo crivo do corretor gramatical, ainda é possível verificar a consistência terminológica em lotes de tradução com o verificador terminológico.

Figura 1 – Interface do *Wordfast*



Fonte: Interface do Software *Wordfast*

Contudo, um fator determinante das CAT abre ainda mais espaço e amplia o mundo do OCR dentro da tradução: os arquivos aceitos por essas tecnologias, em sua grande maioria, necessitavam obrigatoriamente oferecer ao tradutor a possibilidade de editá-los, ou seja, para abrir arquivos dentro das CAT e utilizá-los posteriormente como MT, glossário ou dicionário, esses arquivos precisavam, em

um estágio inicial, ser editáveis. Como visto anteriormente, o PDF tornou-se a opção mais viável para a visualização de arquivos independente do seu formato, porém, mesmo com a grande utilidade dessa ferramenta, ela ainda não oferecia aos tradutores a opção de editar o conteúdo daquele documento. É a partir desse momento que surge uma necessidade bastante presente de converter os arquivos em PDF para um formato editável, passando todo o conteúdo daquele documento para esse novo arquivo que, nos dias atuais, está concentrado no Microsoft Office *Word*, devido à sua imensa gama de opções para editar qualquer tipo de material.

A necessidade de usar uma CAT atrelada a uma ferramenta que permita a edição de texto tem como principal razão a própria tradução em si, ou seja, somente com a certeza da possibilidade de edição irrestrita no documento será garantida ao tradutor a liberdade necessária para aplicar a sua capacidade técnica em para a tradução daquele documento.

Diante da vital importância em transformar documentos não editáveis em documentos editáveis, contextualiza-se o uso dos editores de OCR. Os editores de OCR são ferramentas que permitem a leitura de caracteres de um documento. Na prática, essa leitura economiza tempo, aumenta a produtividade e, principalmente, reduz altamente os níveis de erros de digitação que ocorreriam caso o mesmo trecho de texto, por exemplo, fosse produzido por um ser humano. Há vários editores no mercado, contudo um dos mais completos *softwares* e que oferece uma grande gama de soluções para os mais diversos tipos de documento é o *ABBYY FineReader*. Além de editor de texto, o *software* mencionado também é um editor de imagem, de modo que permite ao usuário não somente editar o texto, mas imagens, tabelas, marcas d'água e quaisquer outros elementos presentes na construção do OCR. O manejo desse tipo de *software* demanda do usuário um domínio sobre suas ferramentas e utilidades. Bem utilizado, esse tipo de ferramenta auxilia diretamente no preparo de OCR para uso nas CAT.

Diante do exposto, pode-se inferir que o OCR faz parte de uma competência instrumental que auxilia diretamente no preparo de documentos e no produto final que será a entrega ao cliente daquela tradução.

CAPÍTULO II:

PASSO A PASSO – TUTORIAL PARA OCR

Observando um cenário de déficit não apenas no que diz respeito a competência instrumental, mas também na falta de tutoriais que ensinem a aplicabilidade prática dessas competências, o presente trabalho propõe-se a oferecer, em formato de imagens com legendas visuais e textuais explicativas uma espécie de passo a passo básico para a criação de OCR voltado para a tradução. Visando uma maior dinâmica com relação ao ensino dessa prática e o desenvolvimento dessa competência, o manual também estará disponível em formato audiovisual com narração de todos os passos com o intuito de explicar o processo realizado até o produto final.

1º. - Abra a pasta onde se encontra o Arquivo em formato .pdf ou em qualquer formato que esteja diferente de *Word*;

Figura 2 – Caminho para identificar o arquivo para o OCR

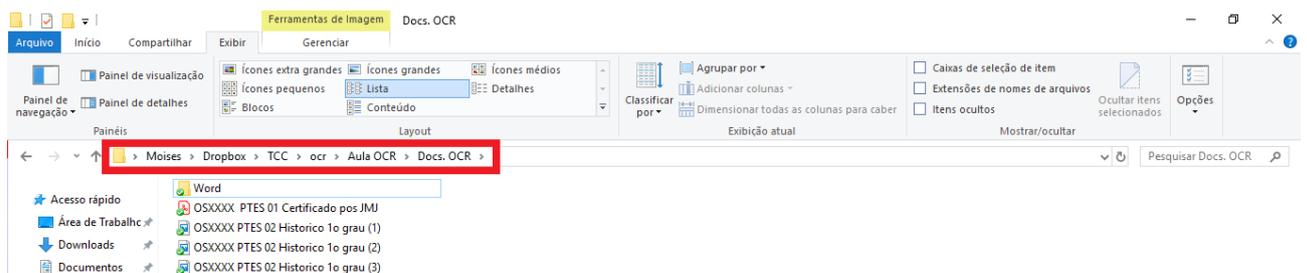


Figura 2 - Caminho para identificar o arquivo para o OCR

Em geral, a nomeação dos arquivos acontece da seguinte maneira: OS (Ordem de Serviço) juntamente com o número que ela representa, ou seja, OS500 significa que aquela ordem de serviço é a de número 500 que foi gerada. Logo em seguida aparece o idioma de origem e para qual língua o referido serviço será traduzido. Logo, se em um documento se encontra a extensão PTES, significa que a língua original do documento é o português e a tradução deverá ocorrer para o

espanhol. Por fim, é informado qual é o tipo de documento que será editado como históricos, certificados, certidões, carteiras de habilitação, entre outros e, em caso de mais de uma página onde os documentos não estão juntos em somente um arquivo, também encontramos entre parêntesis a numeração e a sequência de cada página.

2º - Caso o documento seja editável pelo *ABBYY*, clique com o botão direito do mouse em cima do arquivo e selecione a opção "*ABBYY FineReader > Abrir no ABBYY FineReader*".

Figura 3 - Atalho para abrir o documento direto no software *Abbyy Finereader* versão 12

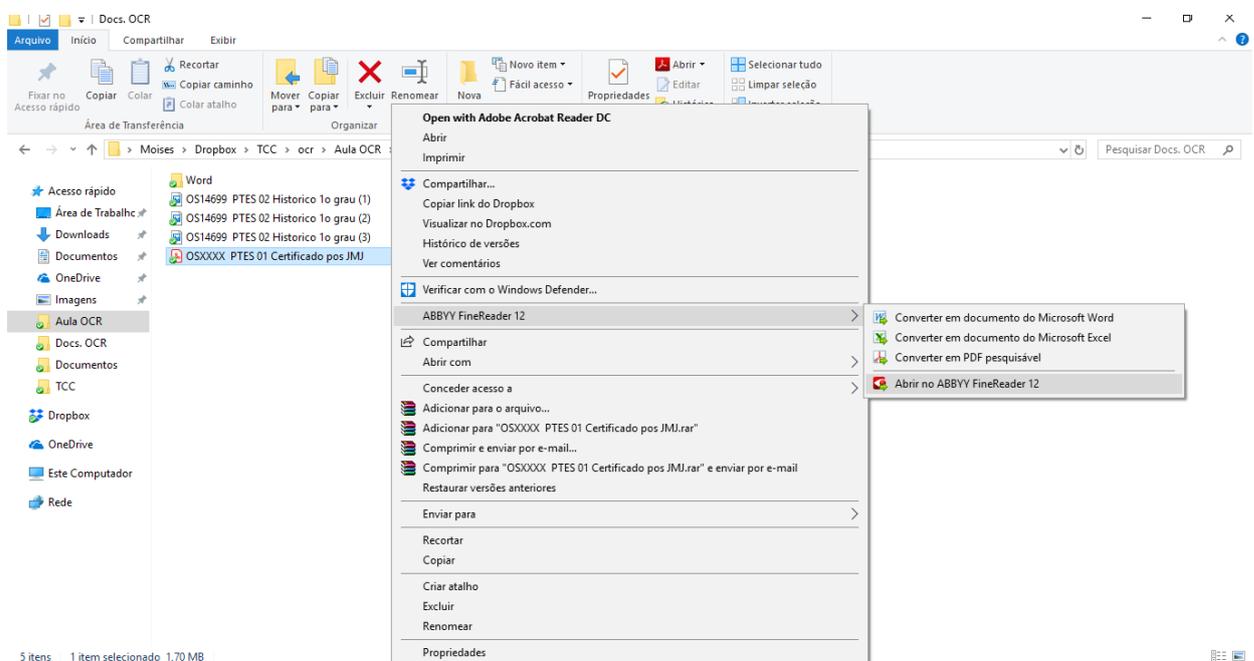


Figura 3 - Atalho para abrir o documento direto no software *Abbyy Finereader* versão 12

Pode-se considerar documentos não editáveis através do *ABBYY* aqueles que em sua composição possuem, majoritariamente, trechos em manuscrito, carimbos, selos ou informações sobrepostas ao texto. De modo geral, quaisquer elementos que dificultem a leitura de caracteres, seja por sobreposição ou por ilegibilidade.

3º - Para editar o documento do zero pressione CTRL + DEL, esse comando excluirá todas as marcações iniciais do *ABBYY*; em seguida, utilize a coluna

localizada do lado esquerdo do programa (destacada em vermelho na imagem) para selecionar área de texto (verde) ou tabela (azul) que, no caso, são as mais utilizadas na tarefa de OCR. A área marcada em vermelho destaca e envia imagens para o formato *Word*, contudo, não as torna editáveis, de modo que é apenas uma reprodução daquela imagem, sendo, por essa razão, pouco utilizada na tradução juramentada e constantemente mais aproveitada na tradução técnica. Clique com o mouse e arraste até a área que deseja marcar. Durante a marcação, ignore imagens e manuscritos apresentados no arquivo, essas partes do texto serão editadas diretamente no *Word*. Caso deseje, na função “Editar imagem” é possível utilizar uma borracha para selecionar e apagar imagens e fragmentos de texto não desejados. As marcações e cores citadas estão destacadas na imagem abaixo dentro do quadrado com bordas amarela.

Figura 4 - Editor de OCR ABBYY FineReader versão 12 Profissional

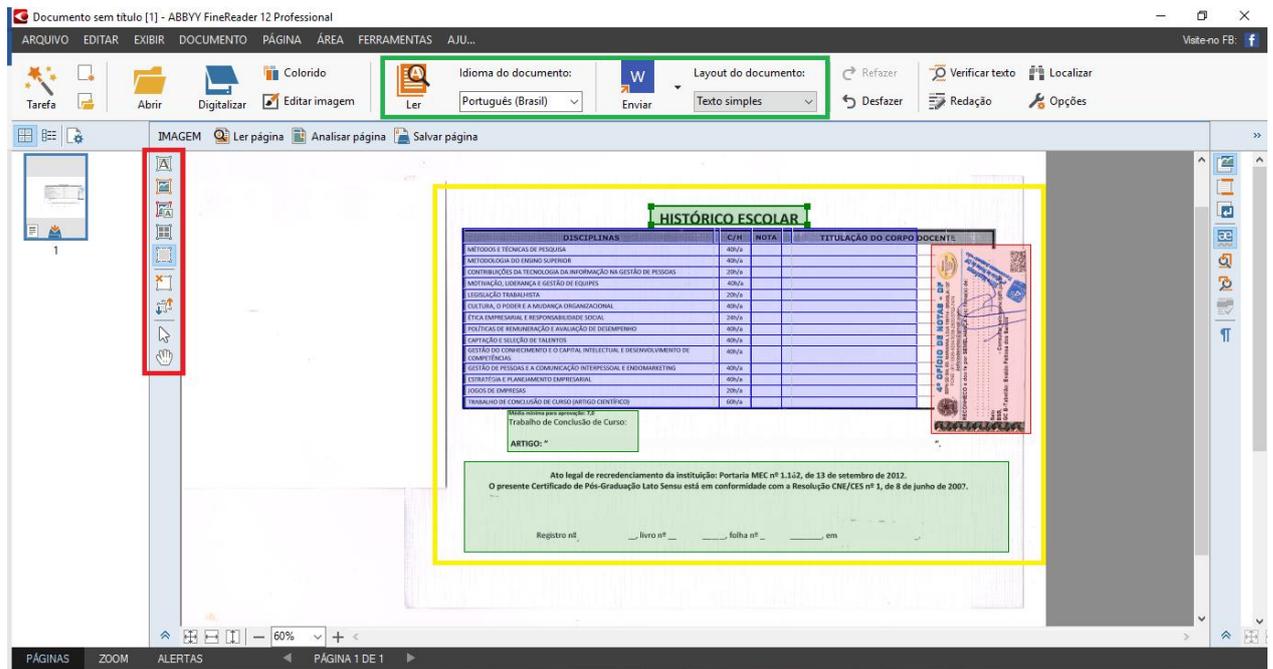


Figura 4 - Editor de OCR Abby Finereader versão 12 Profissional

Verifique se a seleção de idioma do documento está correta, tal como o modo de envio para o formato desejado (ícone do *Word* à direita) e o *layout* do documento, que deverá sempre ficar como "Texto Simples". A obrigatoriedade do formato "TS" ocorre devido ao caráter do documento, isto é, ele deverá ser editável e, em razão

dos outros formatos em “layout do documento” não permitirem ou dificultarem ao usuário a edição das informações, cria-se a obrigatoriedade em selecionar TS.

5º - Conferidos todos esses dados, clique em "Ler" e depois em "enviar" e aguarde que um documento em *Word* abra.

Figura 5 – Caminhos para exportar o arquivo para um documento em word



Figura 5 - Caminhos para exportar o arquivo para um documento em word

Figura 6 – Documento recém enviado para o word pelo programa Abbyy

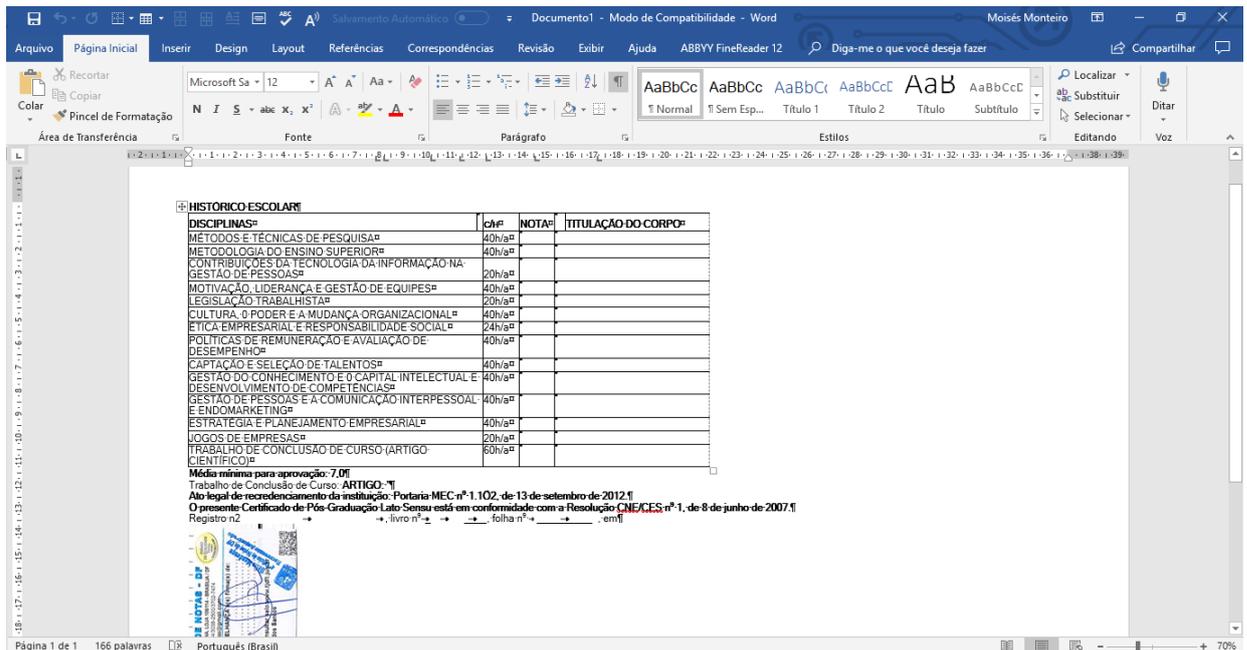


Figura 6 - Documento recém enviado para o word pelo programa Abbyy

6º - Com a tecla "Windows + Seta" ajuste seu monitor e divida a tela em duas, de modo que possa, de maneira constante, comparar o OCR ao original; esse passo é fundamental para evitar erros durante a edição.

7º - Já no documento enviado pelo ABBYY, selecione todo o texto e clique em "Estilos > Limpar Tudo". Essa ação excluirá todas as formatações estabelecidas previamente pelo programa, deixando o usuário com uma interface mais limpa para trabalhar.

Figura 7 – Ícone para limpar todo o documento

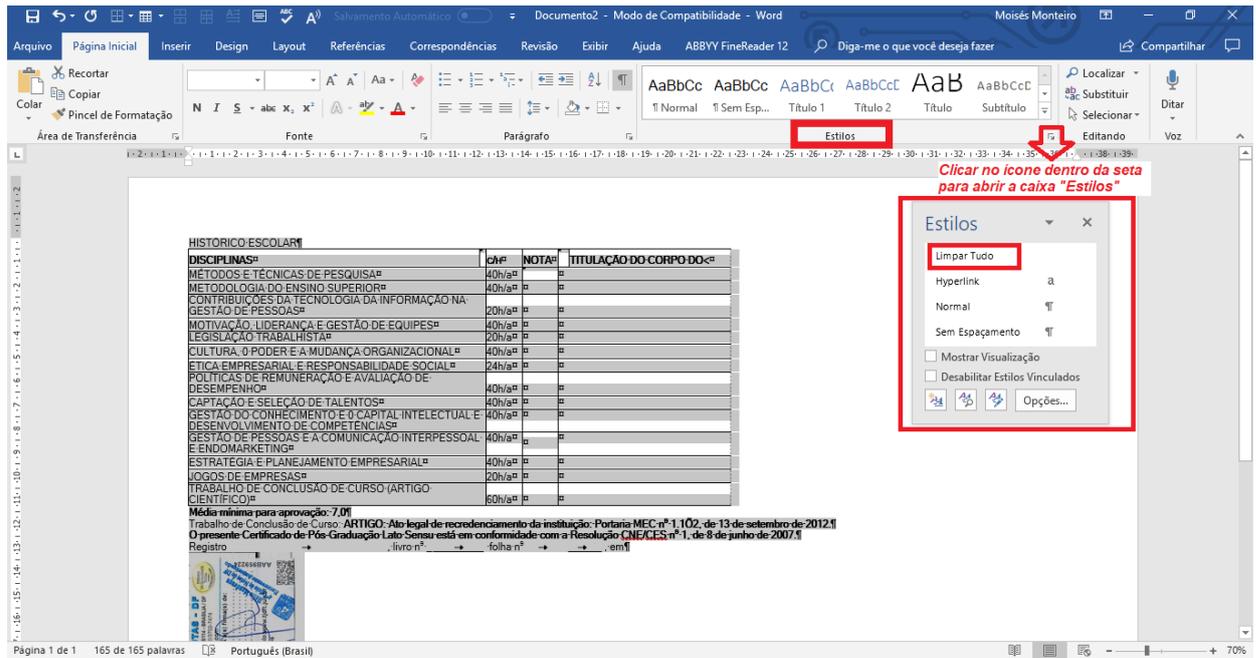


Figura 7 – Ícone para limpar todo o documento

Na aba acima, clique em "Inserir > Links > Indicador" e, se houver algum indicador, o selecione e clique em excluir. Após esse procedimento, o documento estará pronto para edição. Com a numeração de 1 a 4, a imagem indica qual é a ordem de cliques para chegar até a caixa de indicadores.

Figura 8 – Ícone para limpar todos os indicadores do documento

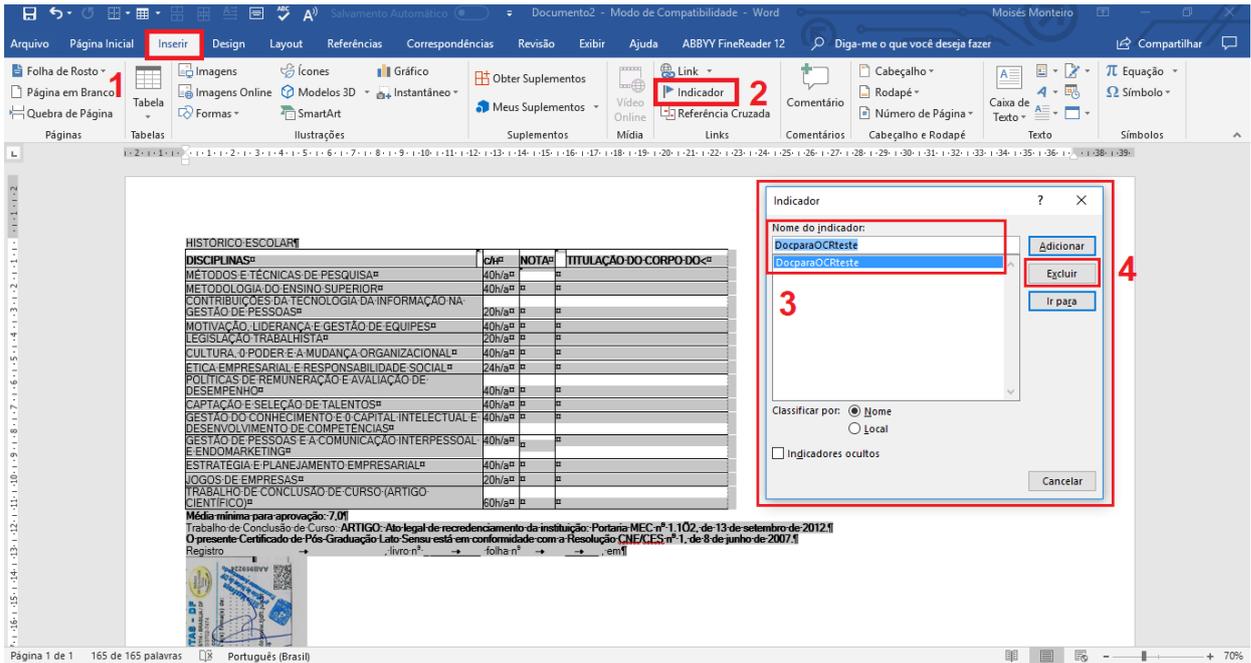


Figura 8 – Ícone para limpar todos os indicadores do documento

8º - Ainda com todo o texto selecionado, verifique o Layout do documento clicando em "Layout > Orientação > Tamanho > Margens. Como cada empresa ou pessoa física usa o formato de acordo com a sua necessidade, esse passo não é padrão para todos que venham a usar o *Word* para tradução. Em alguns casos, há formulários padronizados onde os tradutores enviam todas as suas traduções de modo que, por conta da existência desse padrão pré-estabelecido, as margens e o formato também seguem um padrão independentemente do tipo de documento que será feito, como no exemplo abaixo:

Figura 9 – Margens personalizadas configuradas

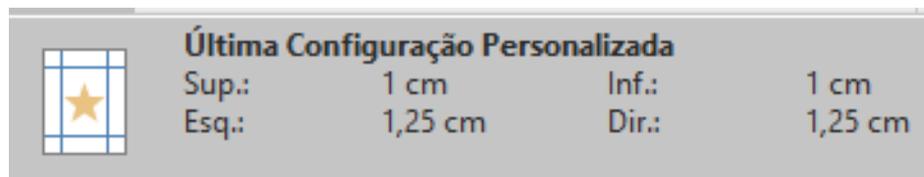


Figura 9 – Margens personalizadas configuradas

Caso não exista margens configuradas, vá até "margens personalizadas" no final da janela e configure as margens para a configuração desejada. Esse passo auxiliará na redução de tempo dedicada ao documento e diminuirá

significativamente a possibilidade de possíveis equívocos quanto ao formato de impressão, visto que com a pré-configuração de margens basta selecioná-las e todo o documento seguirá o padrão desejado.

Figura 10 – Caminho para personalizar as margens

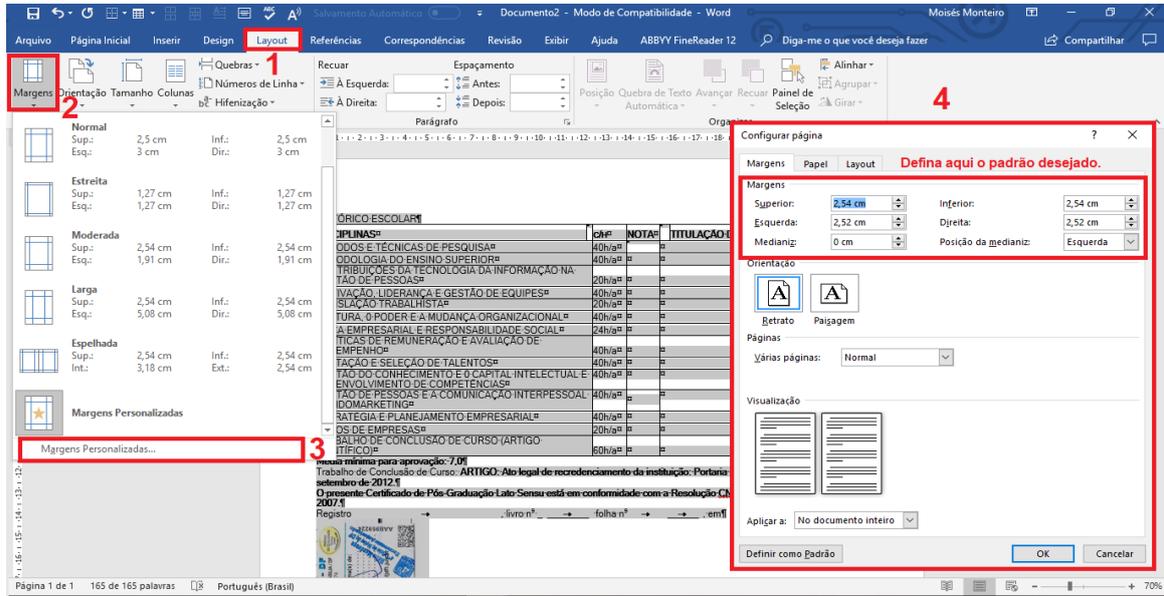


Figura 10 – Caminho para personalizar as margens

9º - Selecione a fonte desejada juntamente com o seu tamanho e, no rodapé da página, clique no idioma e verifique se ele está de acordo com a língua que o documento apresenta. As fontes utilizadas no OCR dependem da política interna de cada empresa, do tradutor ou da sensibilidade do editor que poderá verificar o melhor tipo de letra e tamanho para cada serviço. Contudo, a variação desses elementos transita entre no máximo tamanho 12 e, no mínimo, 8. Cabe ao editor verificar qual fonte e tamanho se adequa melhor à necessidade específica de cada serviço.

Figura 11 – Caminho para definir a fonte, seu tamanho e a linguagem do documento

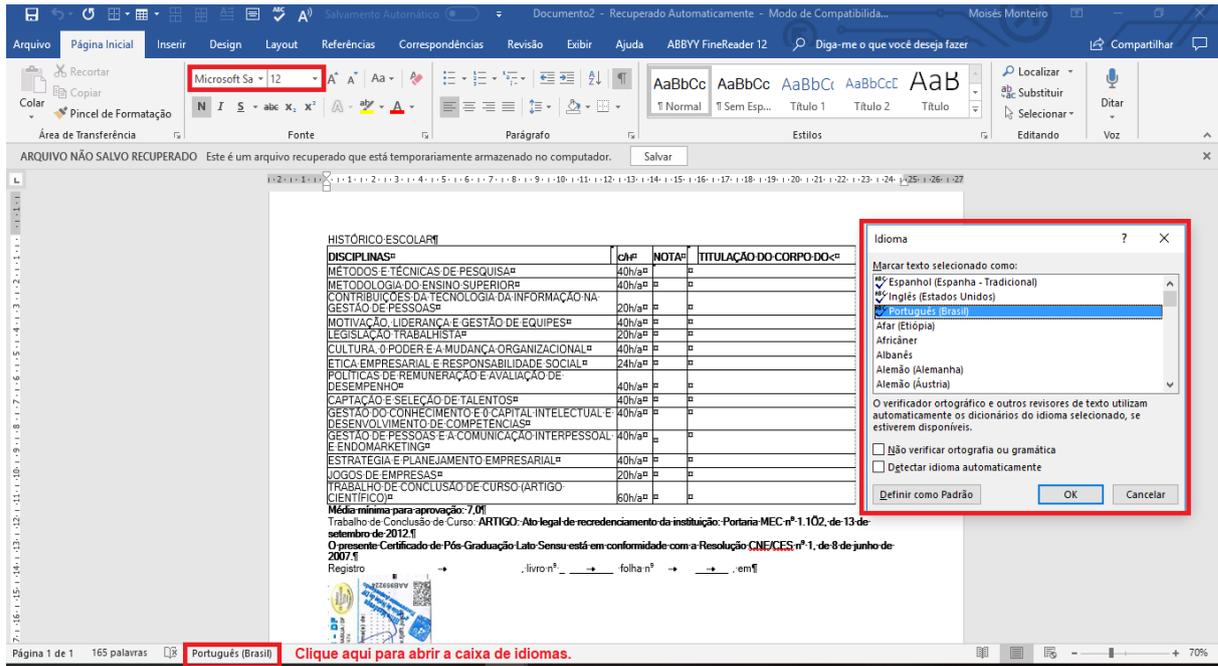
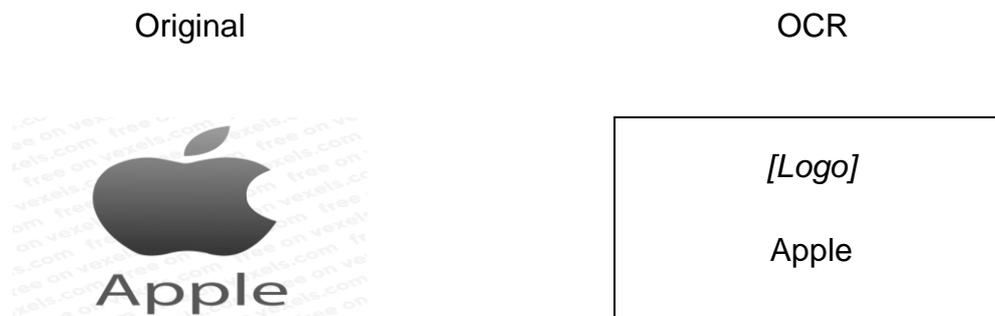


Figura 11 – Caminho para definir a fonte, seu tamanho e a linguagem do documento

10º – Caso se trate de traduções juramentadas, exigência de clientes ou qualquer situação que demande ao editor a retirada de imagens do documento, utilize o atalho "CTRL + I", abra colchetes e indique em itálico que tipo de imagens há naquele espaço, tais como brasões, logos, fotos, imagens, códigos QR e de barras, etc. Todos os elementos contidos no documento que não podem ser reproduzidos através de texto devem seguir essa regra, contendo, logo abaixo, os elementos textuais que fazem parte daquela parte específica. A imagem abaixo mostra, a esquerda, a imagem original e, a direita, como ela deverá ser retratada no OCR:



11º - Ajuste cada trecho do documento, se possível, a mesma forma do original. Exemplificando: se um trecho está centralizado, centralize-o no OCR; desse modo, o layout do documento ficará o mais próximo possível do documento original; todo trecho alinhado a margem esquerda do documento deve estar justificado; essa ação além de tornar o documento visualmente mais agradável, traz um ganho de espaço e uma organização ao documento que está sendo editado.

Figura 12 – Ícones para alterar o layout do texto

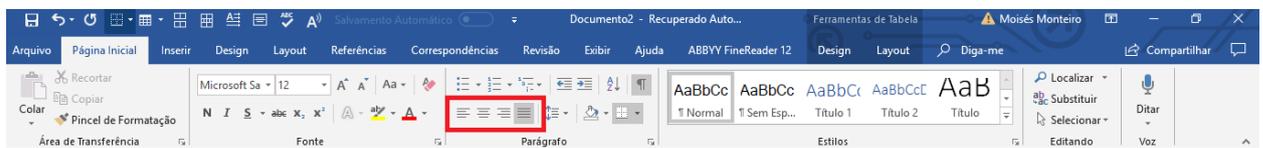


Figura 12 - Ícones para alterar o layout do texto

12º - Marque os negritos e sublinhados de acordo com o apresentado nos documentos originais.

Figura 13 – Ícones para adicionar negrito, itálico e sublinhado, respectivamente.

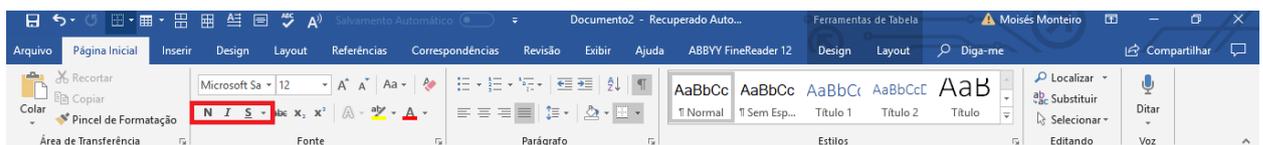


Figura 13 - Ícones para adicionar negrito, itálico e sublinhado, respectivamente.

13º – Como em outras situações o espaçamento padrão depende do executor daquele serviço, contudo na maioria dos casos é usado o espaçamento simples, utilizando somente um espaço entre as linhas; isto é, a tecla "Enter" não deve ser usada para criar mais de um espaço entre os parágrafos; Para verificar quantos espaços há entre os parágrafos, quantas vezes a tecla barra de espaço foi usada, entre outras funções é necessário utilizar uma das ferramentas de revisão do *Word*, localizada na parte superior do *software*, acima de "Parágrafo".

Figura 14 - Ícone para mostrar espaçamento e imagens que demonstram a quantidade utilizada, respectivamente.

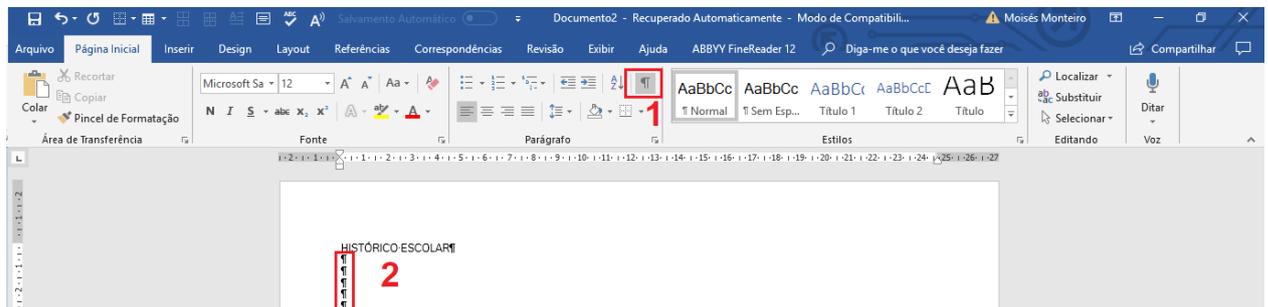


Figura 14 - Ícone para mostrar espaçamento e imagens que demonstram a quantidade utilizada, respectivamente.

Com o ícone demonstrado na imagem acima com a numeração 1 selecionado, todos os espaços do texto aparecerão conforme a imagem de número 2, possibilitando, dessa maneira, um maior controle do editor com relação aos espaçamentos contidos no texto. Vale ressaltar que essas imagens não aparecem no modo de impressão do documento e, conseqüentemente, não estarão presentes caso se faça necessário imprimir o documento. Trata-se, apenas, de uma ferramenta de revisão textual.

14º – Em alguns casos, as MT apresentam peculiaridades quanto a nomenclaturas existentes nos documentos. É, por exemplo, o caso do registro gramatical "nº". Algumas MT identificam o trecho em questão como o registro “no” e, conseqüentemente, propõe uma sugestão equivocada quanto a tradução daquele trecho. Por essa razão, a depender da MT usada para o serviço em questão, os trechos em que constam "nº" devem ser substituídos por "No.". Novamente, vale ressaltar que essa mudança depende das normas pré-estabelecidas entre as partes que tratam do serviço.

Conforme demonstrando na imagem abaixo, há um caminho dentro do Microsoft Office *Word* que permite a substituição de um segmento textual por outro conforme escolha do editor. Ao clicar em substituir, o programa dará acesso a uma caixa onde, o item sinalizado com a letra dois cumpre o papel de localizar todos os termos procurados dentro daquele documento. Exemplificando: ao digitar no campo a palavra “Primeira língua” e, em seguida, no campo número três a palavra “Língua

materna” as opções abaixo substituir, substituir tudo e localizar próxima serão desbloqueadas. Dessa forma, caso escolha a opção substituir, o *software* substituirá a palavra selecionada pela outra no trecho mais próximo identificado. No segundo caso, a substituição ocorrerá em todo o documento onde o segmento textual for identificado. Por último, a opção localizar próxima apenas nos dá a posição onde a palavra selecionada aparece dentro do texto, não ocorrendo, neste caso, nenhuma substituição. A função em questão é comumente usada para substituir erros que os *softwares* de leitura de OCR produzem. De modo mais claro, se em um documento verificou-se que, por alguma razão, um *software* de OCR identificou o número “3” e o leu como a letra “B”, é possível substituir todas de uma só vez sem a necessidade de revisá-las individualmente.

Figura 15 – Caminho para localizar, substituir um ou vários arquivos de uma só vez

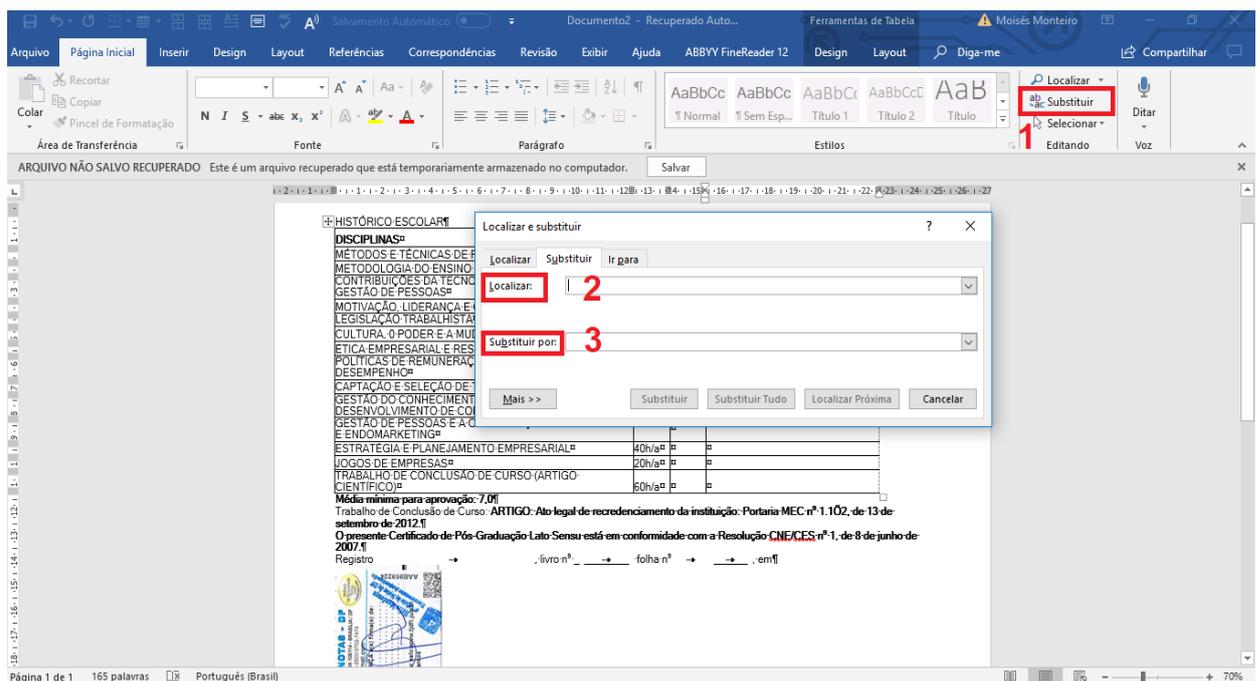


Figura 15 – Caminho para localizar, substituir um ou vários arquivos de uma só vez

15º – Para garantir o pleno funcionamento das MT, normalmente é solicitado ao editor de OCR a remoção de hiperlinks dentro do texto. Para tanto, basta clicar com o botão direito do mouse sobre o hiperlink em questão e selecionar a opção “remover hiperlink”.

Figura 16 – Passo a passo para remover hiperlinks do texto

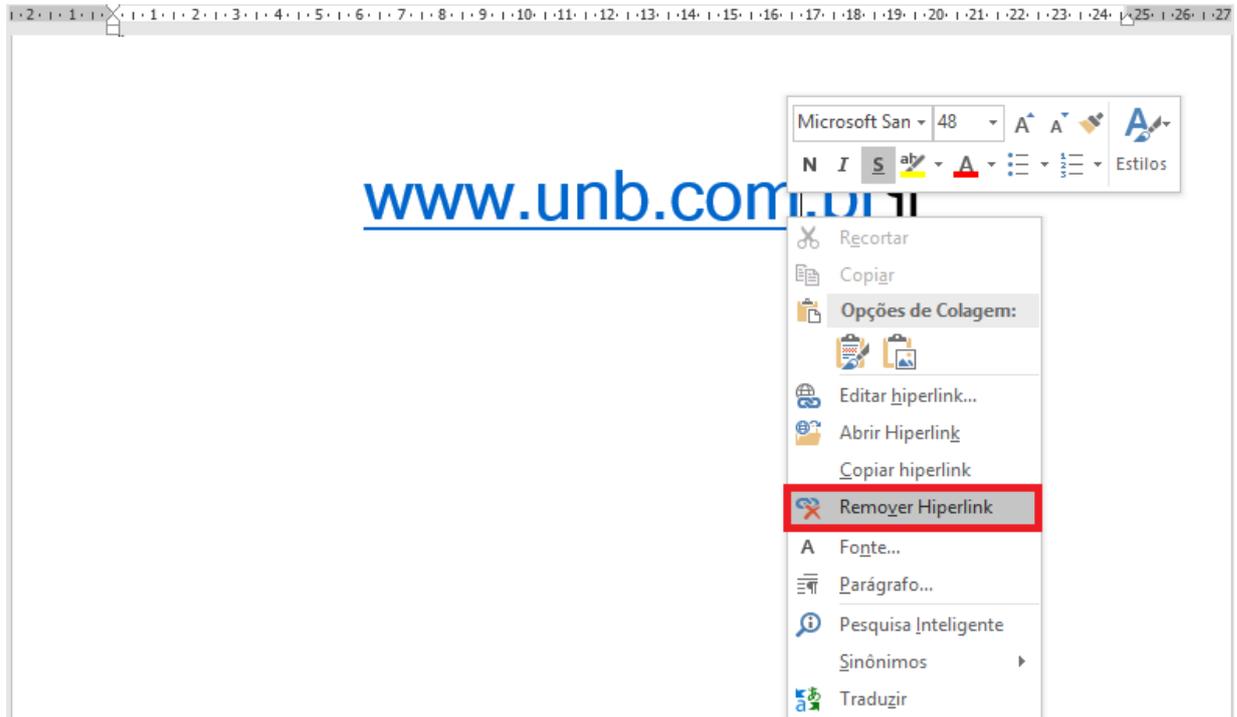


Figura 16 – Passo a passo para remover hiperlinks do texto

16º - Evite usar a tecla "Tab" nos documentos. Sempre que precisar colocar duas informações nos extremos da folha, opte por utilizar a tabela como exemplificado abaixo:

| | |
|---------------------|---------------------|
| INFORMAÇÃO 1 | INFORMAÇÃO 2 |
|---------------------|---------------------|

Caso deseje que as bordas fiquem invisíveis ou apareçam apenas em algumas partes, basta acessar no menu superior a aba "Bordas" e selecionar a que melhor atende à demanda do documento em questão, como no modelo abaixo:

INFORMAÇÃO 1

INFORMAÇÃO 2

| |
|---------------------|
| INFORMAÇÃO 1 |
|---------------------|

INFORMAÇÃO 2

INFORMAÇÃO 1

INFORMAÇÃO 2

Figura 17 – Ícone para selecionar bordas

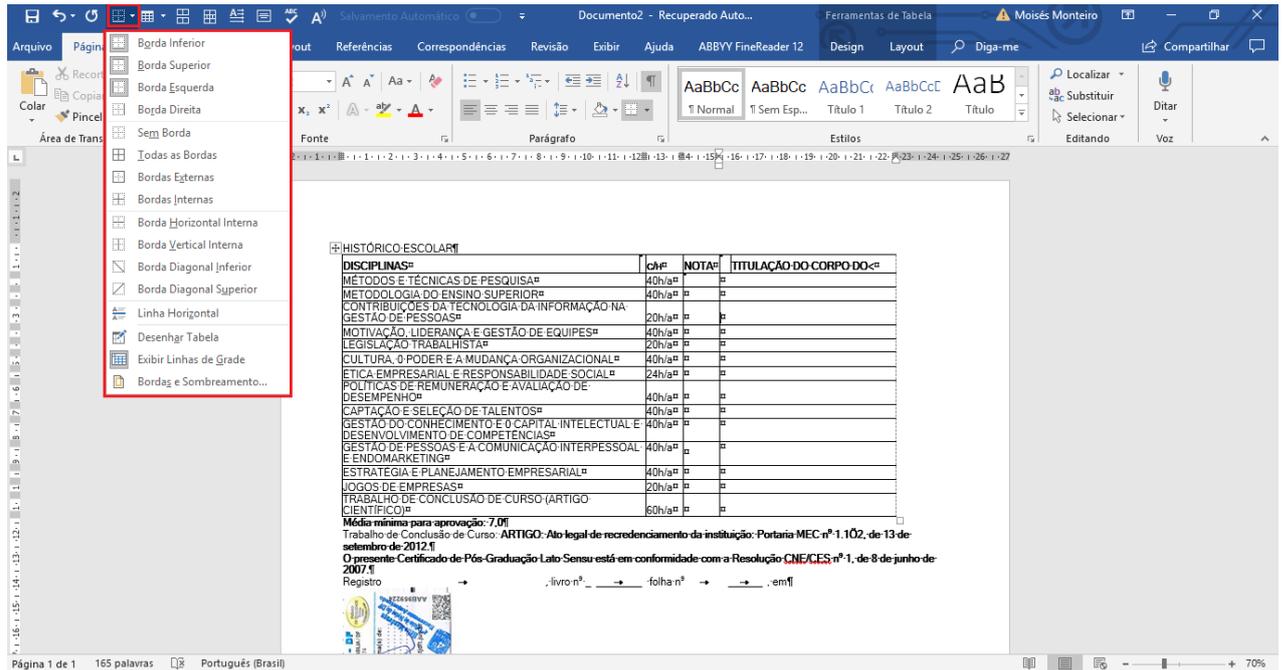
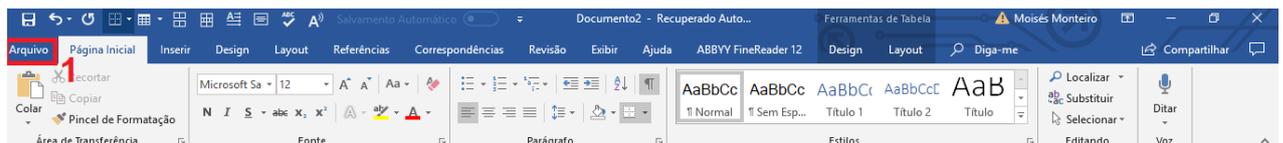


Figura 17 – Ícone para selecionar bordas

17º - Para renomear o documento, clique em "Salvar Como", selecione a pasta, renomeie o documento e clique em "salvar".

Figura 18 – Caminho para salvar e renomear o arquivo



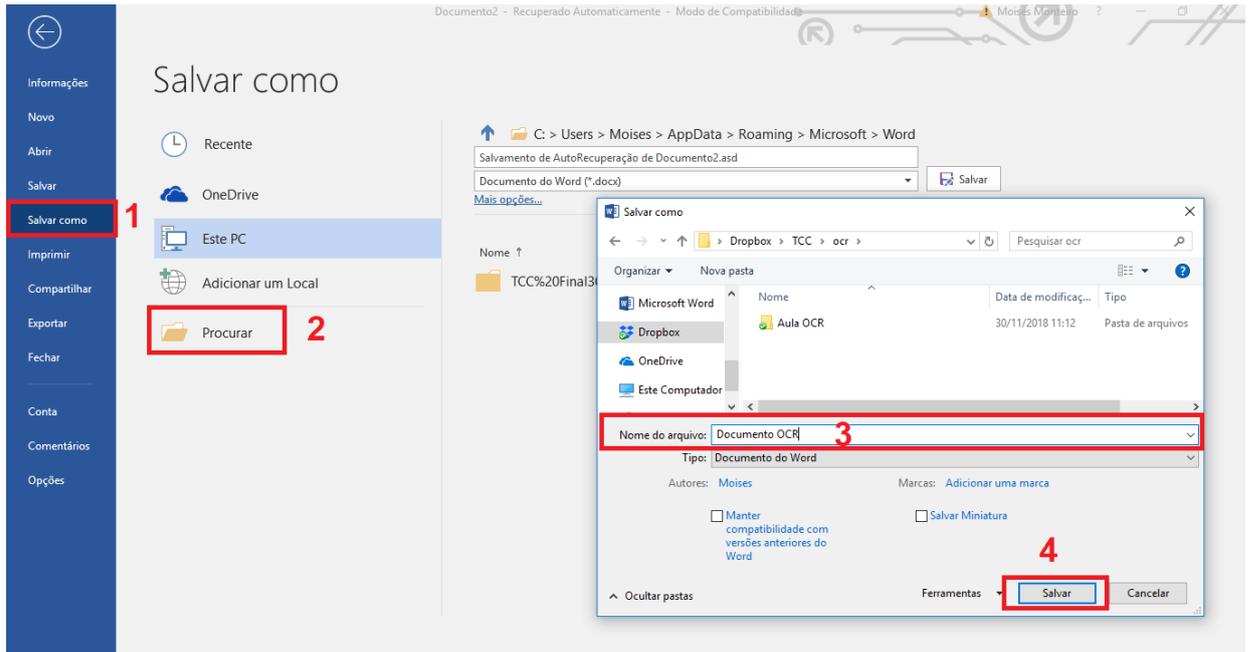


Figura 18 – Caminho para salvar e renomear o arquivo

18º. Seguidos todos esses passos, procure estruturar o documento o mais próximo possível do original. Caso esse arranjo não seja possível, adeque o documento de acordo com as margens e layout de trabalho fornecidos, criando, dessa forma, toda uma estrutura que faça sentido dentro da tradução. Após esses passos, o OCR deverá apresentar um documento com sequência lógica, coerente e legível.

Figura 19 – Documento recém enviado ao *Word* pelo *Abbyy*

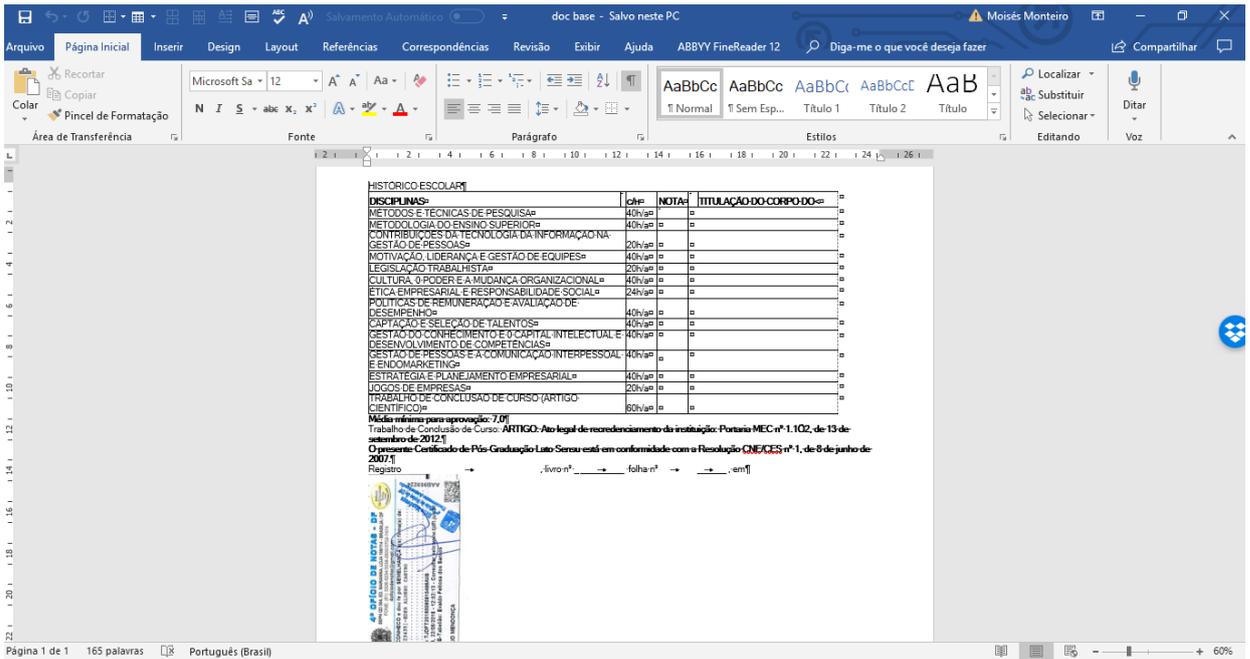


Figura 19 – Documento recém enviado ao Word pelo Abbyy

Figura 20 – Documento com o OCR finalizado

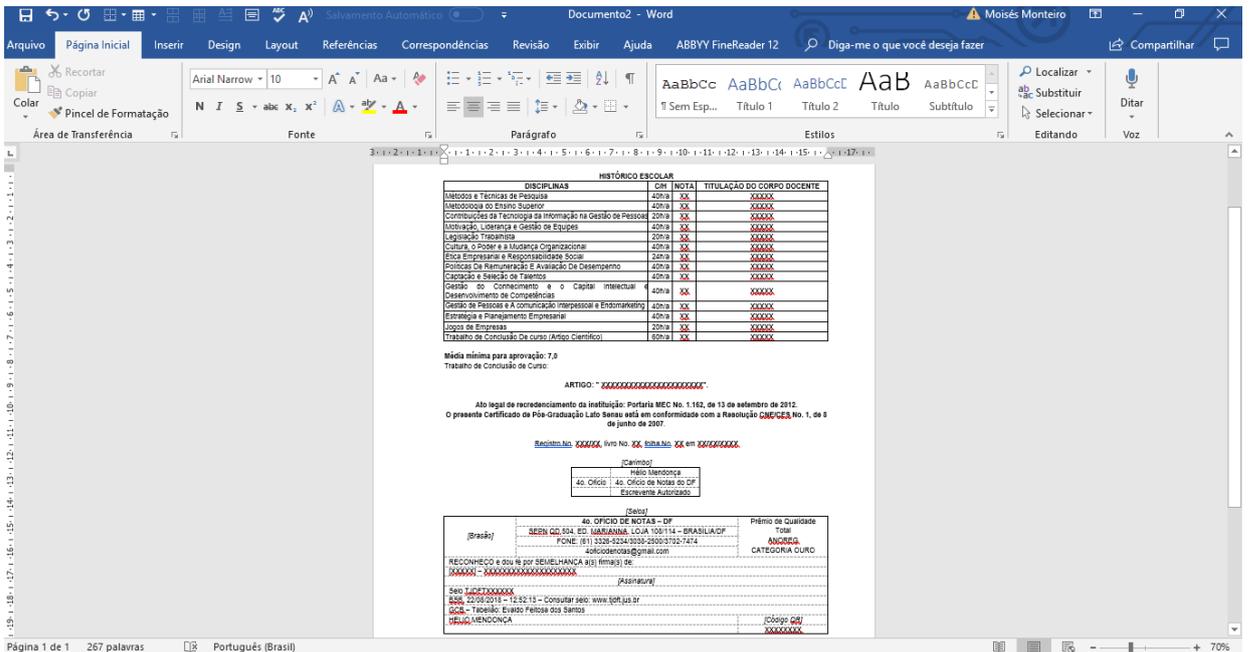


Figura 20 – Documento com o OCR finalizado

Teclas de atalho úteis do Word

- "CTRL + J"
- "CTRL + E"
- "CTRL + G"
- "CTRL + Q"

- Justifica o texto.
- Centraliza o texto.
- Alinha o texto a margem direita.
- Alinha o texto a margem esquerda.

| | |
|----------------|--|
| "CTRL + I" | Itálico. |
| "CTRL + N" | Negrito. |
| "CTRL + S" | Sublinhado |
| "CTRL + setas" | Seleciona uma palavra, trecho ou parágrafo dependendo de como usado. |
| "Tab" | Cria setas no texto ou muda de uma célula a outra dentro de uma tabela; se a célula da tabela for a última, cria mais uma célula. |
| "Shift + F3" | Altera trechos entre maiúsculos e minúsculos; muito útil em históricos. |
| "F4" | Repete o último comando; Exemplo: Ao digitar o número 5 e apertar "F4" o número irá aparecer novamente por quantas vezes a tecla for acionada. |
| "CTRL + A" | Salvar como |
| "CTRL + B" | Salvar |
| "CTRL + T" | Selecionar todo o texto |
| "CTRL + O" | Abre um novo documento em <i>Word</i> |
| "CTRL + W" | Fecha o documento em <i>Word</i> |
| "CTRL + P" | Visualizar a impressão |

CONSIDERAÇÕES FINAIS

Observando a evolução da tradução como ciência e como ofício, apresenta-se uma breve visão sobre a subcompetência instrumental e como ela se relaciona com a tradução. Através de uma análise histórica a respeito do OCR, este trabalho busca mostrar em que se baseia a necessidade de editar documentos para a tradução e a razão pela qual essa funcionalidade é útil e tende a desenvolver-se juntamente com a tradução.

Esse trabalho propôs a criação de um tutorial escrito e em formato audiovisual visando criar um passo a passo prático para a criação de OCR voltado para o mercado de trabalho e para a universidade. Foi possível observar o desenvolvimento das tecnologias de apoio à tradução, a importância dos *softwares* para a execução do trabalho e a dinâmica que envolve a tradução moderna. Seria de grande valia submeter o material em questão ao crivo de alunos e, dessa forma, implementar as melhorias que se façam necessárias.

REFERÊNCIAS BIBLIOGRÁFICAS

HEITLINGER, PAULO. A legibilidade das letras para leitura automática. **Cadernos de Tipografia** Nr 3. Disponível em < <http://www.tipografos.net/cadernos/cadernos-03.html> >. Acesso em 25 de agosto de 2018.

GONÇALVES, J. L. Repensando o desenvolvimento da competência Tradutória e suas implicações para a formação do tradutor. **Graphos**, João Pessoa, v. 17, n. 1, UFPB/PPGL, p. 114-130, 2015. Disponível em <<http://www.periodicos.ufpb.br/ojs/index.php/graphos/article/view/25053/13707>>. Acesso em: 18 de setembro de 2018.

GONÇALVES, J. L.; MACHADO, Ingrid Trioni Nunes. Um panorama do ensino de tradução e a busca da competência do tradutor. **Cadernos de Tradução**, Florianópolis, v. 1, n. 17, p. 45-69, abr. 2006. ISSN 2175-7968. Disponível em: <<https://periodicos.ufsc.br/index.php/traducao/article/view/6856/6408>>. Acesso em: 15 de setembro de 2018.

LIPARINI, T.; SOUSA, R. C. de; GOMES, M. H. P. Formação de tradutores: o desenvolvimento das subcompetências instrumental e estratégica. **Cultura & Tradução**, João Pessoa, v.1, n.1, 2011. Disponível em: <<http://www.periodicos.ufpb.br/index.php/ct/article/download/13048/7554>>. Acesso em: 04 de outubro de 2018.

VASCONCELLOS, M. L.; BARTHOLAMEI JR., L. A. J. **Estudos da Tradução I**. Curso de Bacharelado em Letras/Libras. Florianópolis: CCE/UFSC, 2008. Disponível em: <http://www.libras.ufsc.br/colecaoLetrasLibras/eixoFormacaoPedagogico/laboratorios/site/midias/laboratorio_3_tudo/textos_base/Texto_Base_Estudos_Traducao_I.pdf>. Acesso em: 16 de novembro de 2018.

METATEXIS; **Porquê traduzir com uma ferramenta CAT?** Disponível em: <<http://www.metatexis.net/portuguese/cat.htm>>. Acesso em: 18 de novembro de 2018.

CHAMPOLLION, YVES; **Wordfast**. Guia do Usuário. Disponível em: <<http://www.chadel.com/wordfast/start.htm>>. Acesso em: 18 de novembro de 2018.

NOGUEIRA, Danilo; NOGUEIRA, Vera Maria Conti. Porque usar programas de apoio à tradução? **Cadernos de Tradução**, [s.i.], p.18-35, 2004. Anual. Disponível em: <<https://periodicos.ufsc.br/index.php/traducao/article/view/6359>>. Acesso em: 17 nov. 2018.