



PORTUGUÊS COMO LÍNGUA ADICIONAL: LINGUÍSTICA E TRADUÇÃO

## A utilização de corpus paralelo como um recurso na tradução assistida por computador: um estudo de caso de tradução de sentenças judiciais português-chinês

*The use of parallel corpus as a resource in computer aided translation: a case study of translation of Portuguese – Chinese court verdicts*

*El uso de corpus paralelo como recurso en la traducción asistida por computadora: un estudio de caso de traducción de sentencias judiciales portugués-chino*

**Wang Tianlong<sup>1</sup>**

[orcid.org/0000-0002-1845-6263](https://orcid.org/0000-0002-1845-6263)  
[wtljames@gmail.com](mailto:wtljames@gmail.com)

**Ana Luisa Varani Leal<sup>1</sup>**

[orcid.org/0000-0003-1764-8535](https://orcid.org/0000-0003-1764-8535)  
[analeal@um.edu.mo](mailto:analeal@um.edu.mo)

**Recebido em:** 2/6/2020

**Aceito em:** 2/12/2020

**Publicado em:** 09/02/2021

**Resumo:** O presente artigo, mediante a compilação das sentenças judiciais proferidas pelo Tribunal de Última Instância da Região Administrativa Especial de Macau, investiga, a partir de um estudo introdutório, a necessidade e a viabilidade da construção de um *corpus* paralelo português-chinês na área jurídica, bem como o desenvolvimento de estratégias a serem aplicadas como recurso metodológico nas ferramentas de tradução assistida por computador, com o intuito de buscar formas para aumentar não só a eficiência, mas também a qualidade da tradução.

**Palavras-chave:** corpus paralelo. Tradução assistida por computador. Sentenças judiciais. Línguas portuguesa e chinesa.

**Abstract:** Through the compilation of the verdicts issued by the Court of Final Appeal of the Macao Special Administrative Region, this article investigates, from an introductory study, the need and feasibility of building a Portuguese-Chinese parallel corpus in the legal area, as well as the development of strategies to be applied as a methodological resource in computer aided translation tools, with the aim of seeking ways to increase not only the efficiency but also the quality of the translation.

**Keywords:** parallel corpus. Computer assisted translation. Verdicts. Portuguese-Chinese.

**Resumen:** Mediante la compilación de las sentencias judiciales dictadas por la Corte Final de Apelaciones de la Región Administrativa Especial de Macao, este artículo investiga, a partir de un estudio introductorio, la necesidad y viabilidad de establecer un corpus paralelo luso-chino en el ámbito jurídico, así como el desarrollo de estrategias aplicables como recurso metodológico en herramientas de traducción asistida por computadora, con el propósito de buscar formas de aumentar no sólo la eficiencia sino también la calidad de la traducción.

**Palabras clave:** corpus paralelo. Traducción asistida por ordenador. Sentencias judiciales. Portugués-chino.

### 1 Introdução

Atualmente, é reconhecido pelos tradutores profissionais que a utilização de *corpus* paralelo está a desempenhar um papel cada vez mais relevante na prática de tradução. O conceito não é recente e, segundo Baker (1995), o *corpus* paralelo refere-se a dois conjuntos de textos: um



<sup>1</sup> Universidade de Macau, FAH, Departamento de Português

conjunto de textos em uma determinada língua de origem e um outro conjunto composto por versões traduzidas destes mesmos textos para um outro idioma. Todavia, a sua utilização como um recurso no processo tradutório tem sido mais explorada na atualidade, especialmente, relacionado com a tecnologia de tradução assistida por computador (CAT)<sup>2</sup>. Assim, observa-se que um *corpus* paralelo de alta qualidade e em grande escala, constituído por textos bilíngues com linguagem especializada, confiável e padronizada, já se tornou um dos melhores meios para enriquecer a memória de tradução na CAT e dinamizar o processo de tradução manual.

Enquanto região administrativa especial da China, Macau possui duas línguas oficiais<sup>3</sup>: chinês e português. Não obstante, a existência de uma grande quantidade de dados bilíngues, a construção de *corpus* paralelo em Macau ainda se encontra em fase inicial, uma vez que, raramente, existem corpora bilíngues disponíveis aos tradutores, sobretudo em áreas específicas, tais como as jurídica, legislativa, médica e tecnológica. Em comparação com outras regiões onde a construção do *corpus* paralelo tem vindo a desenvolver-se a um ritmo mais acelerado, quer em termos de dimensão, quer em termos de qualidade, ainda há muito trabalho a ser realizado nesse processo em Macau.

Assim, a partir de observações empíricas, verificou-se que os tradutores profissionais, quando se deparam com termos e expressões especializadas e técnicas desconhecidas, algumas vezes, ou não conseguem estabelecer uma relação correspondente, ou necessitam de demasiado tempo para consultá-las, a fim de obter a informação para continuar o processo tradutório. Esse tipo de atitude vem sendo verificado diariamente por tradutores profissionais e equipes de tradução.

Nesses termos, para os tradutores da área de Direito em Macau, a criação de um *corpus* paralelo de especialidade jurídica pode ser considerada um projeto importantíssimo, em relação não só ao

aumento da eficiência de trabalho, mas também à melhoria da qualidade de tradução. Entretanto, para que os dados de uma memória de tradução possam ser utilizados como de referência para a tradução, é necessário garantir uma quantidade significativa de *corpus* e a qualidade dos textos originais e das respetivas traduções.

A verdade é que, quanto mais especializada for a área, mais difíceis de serem alcançadas as duas condições (CHEN, 2012). A criação de um *corpus* paralelo com quantidade adequada e suficiente pode ser limitada pelo tempo e pelo investimento de recursos, quer humanos, quer financeiros. Ou seja, sem um grande esforço, seria quase impossível construir uma base de dados em grande escala em um curto espaço de tempo. Assim, este estudo, de caráter preliminar, não tem por objetivo construir um *corpus* com grande quantidade de recursos linguísticos, mas, sim, discutir a possibilidade de se elaborar uma metodologia viável para a compilação de um *corpus* paralelo português-chinês.

O significado da construção de um *corpus* paralelo português-chinês na área jurídica justifica-se, em primeiro lugar, pelo reforço do intercâmbio e da cooperação económica e comercial entre a China e os países de língua portuguesa, e, também, pela procura pela tradução jurídica do par linguístico português-chinês encontrar-se em contínuo destaque. A utilização de *corpus* paralelo como um recurso em uma CAT pode promover, de maneira eficaz, o nível geral da qualidade da tradução entre as línguas portuguesa e chinesa em Macau e atender às necessidades atuais tais como a falta de tradutores jurídicos qualificados; além disso, em articulação com os meios tradicionais de estudos linguísticos, a utilização de *corpus* paralelo pode ainda impulsionar o desenvolvimento da investigação e do ensino da tradução, assim como a formação de tradutores qualificados. Por fim, mas não menos importante, os métodos adotados podem servir como referência para a criação de novas metodologias para a elaboração

<sup>2</sup> CAT: sigla de "Computer Aided Translation" ou "Computer Assisted Translations".

<sup>3</sup> Segundo o artigo 9.º da Lei Básica de Macau, "Além da língua chinesa, pode usar-se também a língua portuguesa nos órgãos executivo, legislativo e judiciais da Região Administrativa Especial de Macau, sendo também o português língua oficial."

de *corpus* paralelo em outras áreas, tais como a tecnologia, a medicina, a economia *etc.*

Assim, para abordar o tema em questão, isto é, a utilização de *corpus* paralelo como um recurso na tradução assistida por computador, este trabalho está organizado da seguinte forma: na seção 2, trata-se do conceito teórico que diz respeito ao *corpus* paralelo, CAT e sentenças judiciais; na seção 3, expõem-se os procedimentos metodológicos empregados na presente investigação, que abrangem principalmente a criação de *corpus* paralelo e a utilização de *corpus* paralelo na CAT; na seção 4, apresentam-se as análises e discussões realizadas neste estudo; a seção 5 aborda as considerações finais obtidas em conformidade com os resultados evidenciados, assim como os desdobramentos futuros.

## 2 Referencial Teórico e Estudos Correlacionados

O fundamento desta investigação está direcionado à intersecção de dois campos distintos, ambos relacionados com a atividade de tradução na área jurídica entre par linguístico português-chinês, isto é, a constituição de um *corpus* paralelo e a sua aplicação em uma ferramenta computacional do tipo CAT, para apoiar a tradução humana.

### 2.1 *Corpus* paralelo

*Corpus* paralelo é o elemento central desta investigação. Conforme definido por Baker (1995), o *corpus* paralelo consiste em textos originais na língua A e em suas versões traduzidas na língua B. De forma geral, a maior contribuição de um *corpus* paralelo consiste em "the support a shift of emphasis, from prescription to description" (BAKER, 1995, p. 231), o que nos possibilita observar de forma objetiva como os tradutores ultrapassam dificuldades durante o processo tradutório e usam as evidências oferecidas para fornecer modelos realistas para tradutores em formação (HU, 2019).

Em outras palavras, com um *corpus* paralelo, um tradutor pode ter acesso a exemplos reais e,

mediante a comparação entre o texto-fonte (TF) e o texto-alvo (TA), analisar as tendências tradutórias, assim como as características de tradução entre linguagens específicas (HU, 2006, p. 29). Nesse sentido, o *corpus* paralelo pode desempenhar um papel relevante durante o processo de tradução, nomeadamente em relação ao esclarecimento de ambiguidade linguística e à tomada de decisões. Na visão de Hu, Wu e Shu (2011), a maior vantagem do *corpus* paralelo consiste na apresentação da relação de correspondência lexical, frasal e textual entre duas línguas, por meio da qual é possível realizar análises quanto às orientações de conversão lexical. Todavia, essa relação não se aplica na totalidade, pois a língua chinesa é uma língua isolante<sup>4</sup>, ao contrário do português, o que requer um domínio linguístico mais profundo entre o par em questão.

No caso deste estudo, mediante a coleta de dados linguísticos da área jurídica em Macau, conseguimos criar um *corpus* paralelo português-chinês e aplicá-lo em uma ferramenta CAT, com vista a observar, analisar e avaliar, de maneira objetiva, como o tradutor humano pode utilizar-se deste *corpus* paralelo enquanto um recurso linguístico no processo tradutório.

### 2.2 Tradução assistida por computador – CAT

CAT é uma tecnologia de tradução em que tradutor humano pode utilizar os recursos de um computador para facilitar o processo tradutório, fazendo com que este seja interativo entre ser humano e computador. Segundo Bowker (2002), CAT refere-se ao processo por meio do qual os tradutores humanos utilizam instrumentos informatizados para os ajudar nas tarefas de tradução. O uso adequado de ferramentas CAT facilita não só a manutenção da uniformidade dos termos como também a boa gestão dos projetos de tradução e dos recursos linguísticos.

A mesma opinião foi compartilhada por Beira e Ferreira (2017, p. 136-137):

Trata-se de um conjunto de ferramentas informatizadas, objetivando apoiar o tradutor tanto no plano da coerência do texto quanto

<sup>4</sup> "Chinês é uma língua isolante em que cada carácter em si, normalmente, constitui um morfema" (MAI, 2012, p. 61).

na rapidez de execução do trabalho. As ferramentas de tradução assistida por computador permitem uma produtividade no processo de tradução graças à automatização dos processos repetitivos. Esse tipo de tradução permite uma produtividade maior devido à reciclagem (reutilização) da informação obtida mediante traduções prévias (armazenadas na memória), pelo aproveitamento da estrutura e do formato dos documentos originais gerados automaticamente.

Apesar de existirem diferentes ferramentas CAT com distintas formas de funcionamento, sem exceção, todas partilham da mesma concepção de reutilizar os recursos de tradução anteriormente produzidos ou preparados (STUPIELLO, 2014) e possuem duas funções básicas: memória de tradução e base de terminologia. Além disso, uma ferramenta CAT consegue ainda segmentar o texto-fonte e disponibilizar ao tradutor um ambiente de trabalho muito confortável, facilitando, assim, de forma muito positiva, o processo de tradução.

Considerando-se os recursos e as funções disponíveis, a tecnologia CAT pode ser pensada como uma espécie de sistema de gestão dos dados tradutórios para apoiar o processo de tradução dos tradutores humanos. A ferramenta CAT permite ao usuário criar, enriquecer e gerir uma base de estruturas linguísticas por meio de memórias de tradução e base terminológica, de modo a facilitar a reutilização no futuro. Em outras palavras, essa tecnologia auxilia e permite que o fluxo do trabalho de tradução seja totalmente informatizado, bem como o resultado da sua respectiva execução apresente estruturas coerentes e padronizadas.

Dessa forma, tendo em conta que a tecnologia CAT pode trazer vantagens para as tarefas de tradução, é de grande importância desenvolver um estudo para identificar tecnicamente o papel desempenhado na tradução de diferentes tipologias de textos como, por exemplo, texto jurídico. Além disso, considerando que Macau é um ambiente bilingue, em que português e chinês são línguas oficiais, selecionamos para análise as sentenças judiciais proferidas pelo Tribunal da Última Instância (TUI) como elementos essenciais.

### 2.3 Corpus - Sentenças Judiciais

Segundo Capez (2012, p. 527), a sentença judicial é considerada uma

[...] manifestação intelectual lógica e formal emitida pelo Estado, por meio dos seus órgãos jurisdicionais, com a finalidade de encerrar um conflito de interesses, qualificado por uma pretensão resistida, mediante a aplicação do ordenamento legal ao caso concreto.

Na linguagem processual, de acordo com o Glossário dos Serviços Digitais da Justiça de Portugal<sup>5</sup>, entende-se por "Sentença" o ato pelo qual o juiz decide a causa principal. Também é denominada dessa forma a decisão do incidente que apresente a configuração de uma causa. Normalmente, uma sentença judicial engloba três partes principais:

- a. Relatório – apresentação, de forma sucinta, do conteúdo do caso desde o momento da propositura da ação até o encerramento da discussão oral na audiência final.
- b. Fundamentos – apreciação jurídica da causa.
- c. Decisão – exposição das conclusões da parte fundamentadora da sentença.

Assim, pode-se ver que uma sentença judicial dispõe de um conjunto de tipologias textuais diferentes. No ponto de vista de Lopes (2016, p. 27), "Considerando sua estrutura composicional, a sentença judicial é composta pelo cruzamento dos tipos textuais, descritivo, narrativo, injuntivo e explicativo, apresentando uma predominância argumentativa."

No caso deste estudo, relativamente às sentenças judiciais analisadas, observamos que, na parte que consta do "relatório", há uma predominância de estruturas narrativas e descritivas, já que se trata de uma parte em que os fatos são detalhados de forma clara e com rigor; na parte relativa aos "fundamentos", observa-se uma predominância explicativa e argumentativa, uma vez que, nesta parte da sentença, o juiz necessita apresentar argumentos consistentes

<sup>5</sup> Disponível em: <https://justica.gov.pt/sdj/Normas-dos-servicos-digitais/Criacao-de-Conteudos/Glossario>. Acesso em: maio 2020.

e fundamentados em conformidade com suas intenções; na parte de “decisão”, verifica-se uma predominância injuntiva, pois esta consiste nas determinações do juiz, as quais se caracterizam por um caráter imperativo.

Contudo, no que diz respeito à estrutura, não é difícil verificar que as sentenças judiciais são relativamente padronizadas, tendo especialmente em conta que algumas frases e termos específicos são sempre utilizados por juízes relatores.

Vale a pena ressaltar que, em Macau, a tradução das sentenças é feita por uma equipe de profissionais especializados, cabendo ao juiz a sua última verificação<sup>6</sup>. Embora as versões traduzidas sejam consideradas de grande valor de referência, continuam a ser uma tradução em relação ao texto original. Dessa forma, considerando que raramente se encontram estudos que dizem respeito à criação de base de dados na área jurídica de Macau, isso justifica a necessidade de criação de um *corpus* paralelo, com vista a garantir a qualidade das traduções do campo em questão.

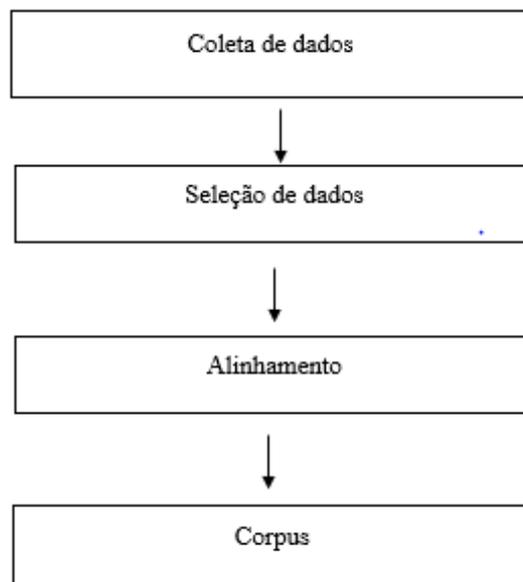
Assim sendo, para lidar com a tradução de textos jurídicos em Macau, é relevante perceber que não só a criação de um *corpus* paralelo entre o par linguístico chinês e português se faz necessária como também a utilização desse *corpus* associada à tecnologia CAT. Visando uma melhor compreensão do estudo realizado, propôs-se uma metodologia para análise, a qual descrevemos a seguir.

### 3 Metodologia

Conforme apresentamos na introdução, Macau é um território em que existe uma grande quantidade de textos bilíngues, incluindo documentos oficiais na área jurídica. Todavia, não se verificam muitos *corpora* paralelos disponíveis aos tradutores, sobretudo em relação aos tradutores utilizadores de ferramentas de CAT. Tendo em consideração essa situação, a metodologia da presente investigação consiste na apresentação do processo de criação de um *corpus* paralelo português-chinês e de como este foi utilizado na tecnologia CAT.

No que se refere ao presente estudo, realizamos as seguintes etapas, conforme a figura 1:

**Figura 1** – Fluxograma da criação do *corpus* paralelo



**Fonte:** elaborado pelos autores (2020)

#### 3.1 Criação de *corpus* paralelo

##### 3.1.1 Coleta de dados

Segundo Sardinha (2004, p. 29), “[...] o *corpus* deve ser adequado aos interesses do pesquisador, que deve ter uma questão a investigar para a qual necessita de um *corpus* específico”. O autor chama atenção à “adequação do *corpus*”, isto é, o *corpus* deve ser representativo em sua temática, de modo que possa evidenciar o seu objeto de análise. Como salienta também Hartley (2009, p. 110), o conteúdo do *corpus* deve ser relevante para o domínio da pesquisa. Se o *corpus* for criado com teor não representativo para o estudo, os resultados da análise não serão significativos e convincentes para a pesquisa, razão pela qual, nesta investigação, a primeira tarefa foi ter acesso a um grande número de textos bilíngues, devidamente redigidos em português e em chinês.

Como foi mencionado na introdução, a área de tradução do presente estudo está diretamente relacionada com campo jurídico, pelo que optamos

<sup>6</sup> Disponível em: <https://legco.gov.hk/yr02-03/chinese/panels/ajls/papers/aj0623cb2-2566-1c.pdf>. Acesso em: maio 2020.

por iniciar a criação de *corpus* a partir de sentenças judiciais proferidas pelo Tribunal da Última Instância, doravante designado por TUI<sup>7</sup>. Conforme figura 1, esse é o órgão supremo na hierarquia dos tribunais, o qual exerce o poder de julgamento em última instância da Região Administrativa Especial de Macau, conferido pela Lei Básica.

Assim, o *corpus* é constituído pelas sentenças

judiciais publicadas na página eletrônica TUI, sendo todos textos elaborados por juizes relatores experientes com/sem traduções oficiais feitas por tradutores profissionais e, por se tratar de documentos oficiais, possuem confiabilidade no que diz respeito às questões linguísticas, e por isso a página eletrônica TUI é um elemento importante na seleção dos dados.

Figura 2 – Página eletrônica da TUI

**Acórdãos recentes**

- T.U.I.
- T.S.I.
- T.J.B.
- T.A.

Pesquisa de acórdãos

Uniformização de Jurisprudência

Sumários

Análise dos casos típicos

Acórdãos importantes no relatório

Você está aqui: [Home](#) > [Acórdãos](#) > [Acórdãos recentes](#) > T.U.I.

### Tribunal de Última Instância

Data da Decisão	Número	Espécie	Texto integral
29/05/2020	58/2020	Recurso Relativo ao Direito de Reunião e Manifestação	<a href="#">PDF</a> <a href="#">TXT</a>
<b>Resultado</b>			
Acordam em negar provimento ao recurso.			
<b>Votação :</b>	Com declaração de voto		
<b>Relator :</b>	Dra. Song Man Lei		
<b>Juizes adjuntos :</b>	Dr. José Maria Dias Azedo Dr. Sam Hou Fai		

Data da Decisão	Número	Espécie	Texto integral
06/05/2020	31/2020	Recurso de decisão jurisdicional em matéria administrativa	<a href="#">PDF</a> <a href="#">TXT</a>
<b>Votação :</b>	Unanimidade		
<b>Relator :</b>	Dra. Song Man Lei		
<b>Juizes adjuntos :</b>	Dr. Sam Hou Fai Dr. Lai Kin Hong		

Fonte: Website dos Tribunais da RAEM

Para que o *corpus* fosse constituído por um número significativo e suficiente de dados linguísticos, foi necessário selecionar um número elevado de sentenças bilingues. Normalmente, quanto maior o volume de dados arquivados em um *corpus*, maiores as probabilidades de que elementos semelhantes ou idênticos possam ser identificados e/ou selecionados para a reutilização em novas tarefas. Assim, para a

realização deste estudo, selecionamos, inicialmente, as sentenças judiciais emitidas em 2019, devidamente publicadas no TUI.

Os textos fornecidos pela página TUI encontram-se em dois formatos distintos (PDF e TXT) e podem ser acessados e baixados diretamente por meio de formato PDF, bem como copiados na versão TXT. A fim de otimizar a dinâmica na seleção do *corpus*, escolhemos a segunda opção, uma vez

<sup>7</sup> Disponível em: <http://www.court.gov.mo/pt/subpage/tui>. Acesso em: maio 2020.

que os documentos em PDF, para serem utilizados na ferramenta CAT, precisam ser convertidos em formatos editáveis e manipuláveis, como por exemplo DOC, TXT, RTF *etc.*, conforme descrevemos na seção relacionada à criação de base terminológica.

### 3.1.2 Seleção de dados

As razões principais pelas quais foram selecionadas as sentenças judiciais do TUI como base de dados estão relacionadas com a sua precisão de linguagem, especialidade jurídica e acessibilidade pública.

Conforme a figura 3, em 2019, foi disponibilizado pelo TUI um total de 126 sentenças, sendo que o número das sentenças monolíngues é de 36 e o número das sentenças bilíngues é de 90. Como foi mencionado no início da seção de Metodologia, o *corpus* da presente pesquisa é composto por textos bilíngues. Dessa forma, em primeiro lugar, foram excluídas as sentenças monolíngues, ou seja, sentenças proferidas apenas em uma língua: quer na língua chinesa quer na língua portuguesa.

**Figura 3** – 126 sentenças proferidas pelo TUI em 2019

The screenshot shows the website of the Tribunais da RAEM. The main navigation bar includes links for 'Mensagem de Boas-Vindas', 'Situação Geral dos Tribunais', 'Acórdãos', 'Distribuição e Marcação', 'Venda Judicial', 'Estatística', 'Consulta das declarações de rendimentos', and 'Download'. The 'Acórdãos' section is active, showing a search for judgments from 2019. The search results table is as follows:

Data da Decisão	Número	Espécie	Texto integral
11/12/2019	21/2016	Recurso em processo civil	
<b>Resultado</b>			
Acordam em julgar improcedente a presente reclamação.			
<b>Votação :</b>		Unanimidade	
<b>Relator :</b>		Dra. Song Man Lei	

**Fonte:** Website dos Tribunais da RAEM

No que se refere à seleção dos textos e excertos, observaram-se os seguintes pontos:

- Tipo do texto: sentença judicial
- Estrutura do texto: texto estruturado em três partes (relatório, fundamentos e decisão)
- Línguas envolvidas: português e chinês
- Data da Publicação: ano de 2019
- Órgão de Publicação: TUI da RAEM, China

Um recurso utilizado na coleta dos textos foi utilizar uma lista com o nome dos juizes relato-

res pois, a partir destes, foi possível identificar com mais facilidade a sua adequada formação linguística e, em consequência, a língua-fonte das sentenças selecionadas. Em algumas sentenças (e.g. processo n.º 43/2015), a tradução está claramente assinalada em chinês/português, o que comprova o pressuposto anterior de que, segundo o nome do relator, pode-se deduzir qual a língua em que se escreveu o texto original da sentença. Em segundo lugar, no que se refere à seleção das sentenças bilíngues, esta foi realizada de forma aleatória, isto é, 5 sentenças, com número de palavras no texto-fonte entre 3450 e 4053, foram preparadas em duas versões

editáveis, em formato TXT. Essas 5 sentenças foram redigidas pelo mesmo juiz português, em português, traduzidas simultaneamente na língua chinesa.

Assim, o *corpus* constituído possui as seguintes características:

- Todos os textos do *corpus* são sentenças proferidas no TUI da RAEM
- Todos os textos têm o português como língua-fonte (LF) e o chinês como língua-alvo (LA)

➤ Todos os textos são acórdãos emitidos por tribunais coletivos

➤ Todos os textos evidenciam usos linguísticos da atualidade, pois são relativos ao ano de 2019

Como apresenta o Quadro 1, o *corpus* contém um total de 907 elementos bilíngues alinhados, sendo que a parte em português compreende um total de 19027 palavras, enquanto a parte em chinês do *corpus* é composta por 31502 caracteres (sem pontuações).

**Quadro 1** – Demonstrativo Geral do *Corpus*

Informações do <i>Corpus</i>					Português (LF)	Chinês (LA)
Sentença Judicial	Processo N.º	Tipo de TF	Data da decisão	N.º de elementos bilíngues alinhados	N.º de palavras	N.º de caracteres
1ª	118/2019	Recurso de decisão jurisdicional em matéria administrativa	29/11/2019	130	3450	5293
2ª	36/2017	Recurso em processo civil	29/11/2019	209	3753	5933
3ª	44/2017	Recurso de decisão jurisdicional em matéria administrativa	29/11/2019	259	4053	6698
4ª	108/2019	Recurso em processo civil	29/11/2019	136	3726	6539
5ª	111/2019	Recurso em processo civil	29/11/2019	173	4045	7039
Total				907	19027	31502

**Fonte:** elaborado pelos autores (2020)

### 3.1.3 Alinhamento

O alinhamento representa o elemento fundamental nos estudos de *corpus* paralelo; assim, as unidades de alinhamento podem ser classificadas em diferentes níveis: palavra, expressão, frase, parágrafo ou mesmo página. É importante salientar que quanto menor for a unidade de alinhamento, mais precisa será informação linguística e maior

será o valor de utilização (WANG; HAN, 2010). No entanto, considerando que existe uma grande diferença entre o português e o chinês e que quase a metade dos textos utilizados no presente estudo é oriunda da tradução manual, é difícil conseguir um alinhamento perfeito.

No que concerne ao processo de alinhamento, os procedimentos foram realizados em etapas sequenciais como a seguir.

➤ Etapa 1: os 5 documentos bilingues, constituídos por sentenças judiciais apresentadas acima, em formato TXT, foram importados, respectivamente, para o *software* de alinhamento Tmxmall<sup>8</sup> (figura 4), o qual, de acordo

com as regras de segmentação predefinidas pelo sistema, procede automaticamente à separação dos documentos envolvidos, tanto no texto-fonte (TF) quanto no texto-alvo (TA).

**Figura 4** – Exemplo de alinhamento no programa Tmxmall

No.	Detect language: Portuguese	36-2017PT.xlsx	Detect language: Chinese (Hongkong)	36-2017CH.xlsx
24	I – Relatório		一、概述	
25	Companhia de Desenvolvimento B, Limitada, intentou ação declarativa com processo comum ordinário contra A, pedindo a sua condenação no reconhecimento da autora como proprietária da parcela de terreno com a área de 41 m <sup>2</sup> (formada por duas sub-parcelas com as áreas de 37 m <sup>2</sup> e 4 m <sup>2</sup> ), sito na Taipa, parte do prédio descrito sob o n.º XXXXX, na Conservatória do Registo Predial de Macau e a abster-se de usar a mesma.		乙發展有限公司針對甲提起通常訴訟程序宣告之訴，請求判處被告承認原告為一幢位於氹仔面積為41平方米、作為標示於澳門物業登記局第XXXX號之房地產其中一部分的土地(由面積分別為37平方米和4平方米的兩個地段組成的所有權人，並放棄使用該地段。	
26	O réu A deduziu reconvenção, pedindo a declaração de que adquiriu a propriedade da parcela do terreno por usucapião e,		被告甲提出反訴，請求宣告其以時效方式取得該幅土地的所有權；	
27	subsidiariamente, por acesso industrial imobiliária e ainda subsidiariamente, a aquisição por usucapião dos direitos de superfície, de usufruto, uso e habitação.		作為補充，請求宣告其以不動產人工添附的方式取得所有權；若仍不獲法院支持，則請求宣告其以時效方式取得該幅土地的地上權、用益權、使用及居住權。	
28	Por sentença da Ex.ma Presidente do Tribunal Colectivo foi acción julgada improcedente e o réu absolvido do pedido e julgou o réu parte ilegítima para formular o pedido reconvenional, desacompanhado dos outros herdeiros de C e a autora absolvida da instância reconvenional.		合議庭主席透過判決裁定訴訟敗訴，駁回針對被告的訴訟請求，並裁定被告在沒有丙的其他繼承人陪同的情況下沒有提出反訴請求的正當性，駁回了其針對原告的反訴。	
29	Recorreu a autora Companhia de Desenvolvimento B, Limitada para o Tribunal de Segunda Instância (TSI) que concedeu provimento ao recurso, condenou o réu no reconhecimento da autora como proprietária da parcela de terreno e a abster-se de usar a mesma.		原告乙發展有限公司向中級法院提起上訴，該院裁定上訴勝訴，判處被告承認原告為有關地段的所有權人，並放棄使用該地段。	
30	Recorre, agora, o réu A para este Tribunal de Última Instância (TUI), suscitando a seguinte questão:		現被告甲向本終審法院提起上訴，提出以下問題：	
31	- O DL n.º 6/93/M tem por objetivo legislativo apenas a contenção e erradicação das edificações informais.		—第6/93/M號法令的立法原意僅在於遏制及取締僑建物。	
32	Como os terrenos privados podem ser objecto das transacções de direitos, não existe qualquer motivo que obste e impeça os possuidores das barracas de invocar a usucapião quanto aos		由於私有土地可以成為權利交易的標的，因此沒有任何理由妨礙及阻止木屋的占有人就上述木屋所在的私有土地提出時效取得的請求。	

**Fonte:** extrato do programa Tmxmall (2020)

➤ Etapa 2: o segundo momento do alinhamento foi realizado no ambiente de trabalho da ferramenta Tmxmall com as unidades linguísticas segmentadas automaticamente, em que se realizou um ajuste manual, incluindo-se a divisão, a união, a correção etc. O ajuste dos excertos segmentados foi estabelecido em conformidade com a unidade de tradução, que "*normally ranges from the word through the collocation to the clause. It could be described as 'as small as is possible and as large as is necessary' [...] (NEWMARK, 1988, p. 285).*" No caso do presente estudo, as unidades que não constituíam uma frase completa foram consideradas como um segmento independente.

Na investigação em apreço, observamos que a intervenção humana é fundamental, pois permite ao pesquisador ajustar as inconsistências entre os dois textos, isto é, Texto Fonte (TF) e Texto Alvo (TA), o que implica esforço, concentração, e competência linguística, uma vez que o resultado pode influenciar diretamente a qualidade e o valor do *corpus* paralelo pretendido. Concluídas as etapas de segmentação e alinhamento, consoante à necessidade de utilização, os textos alinhados podem ser exportados em formato TMX, XLSX, DOCX, entre outros, conforme apresentado na figura 5.

<sup>8</sup> Disponível em: <https://www.tmxmall.com/aligner>. Acesso em: maio 2020.

Figura 5 – Textos alinhados pelo Tmxmall e exportados em formato xlsx (Excel)

	A	B
	Texto-Fonte (Português)	Texto-Alvo (Chinês)
25	I – Relatório	一、概述
26	Companhia de Desenvolvimento B, Limitada, intentou acção declarativa com processo comum ordinário contra A, pedindo a sua condenação no reconhecimento da autora como proprietária da parcela de terreno com a área de 41 m <sup>2</sup> (formada por duas sub-parcelas com as áreas de 37 m <sup>2</sup> e 4 m <sup>2</sup> ), sito na Taipa, parte do prédio descrito sob o n.º XXXXX, na Conservatória do Registo Predial de Macau e a abster-se de usar a mesma.	乙發展有限公司針對甲提起通常訴訟程序宣告之訴，請求判處被告承認原告為一幢位於氹仔面積為41平方米、作為標示於澳門物業登記局第XXXX號之房地產其中一部分的土地(由面積分別為37平方米和4平方米的兩個地段組成)的所有權人，並放棄使用該地段。
27	O réu A deduziu reconvenção, pedindo a declaração de que adquiriu a propriedade da parcela do terreno por usucapião e, subsidiariamente, por acessão industrial imobiliária e ainda subsidiariamente, a aquisição por usucapião dos direitos de superfície, de usufruto, uso e habitação.	被告甲提出反訴，請求宣告其以時效方式取得該幅土地的所有權； 作為補充，請求宣告其以不動產人工添附的方式取得所有權；若仍不獲法院支持，則請求宣告其以時效方式取得該幅土地的地上權、用益權、使用及居住權。
28	Por sentença da Ex.ma Presidente do Tribunal Colectivo foi acção julgada improcedente e o réu absolvido do pedido e julgou o réu parte ilegítima para formular o pedido reconvençional, desacompanhado dos outros herdeiros de C e a autora absolvida da instância reconvençional.	合議庭主席透過判決裁定訴訟敗訴，駁回針對被告的訴訟請求，並裁定被告在沒有內的其他繼承人陪同的情況下沒有提出反訴請求的正當性，駁回了其針對原告的反訴。
29	Recorreu a autora Companhia de Desenvolvimento B, Limitada para o Tribunal de Segunda Instância (TSI) que concedeu provimento ao recurso, condenou o réu no reconhecimento da autora como proprietária da parcela de terreno e a abster-se de usar a mesma.	原告乙發展有限公司向中級法院提起上訴，該院裁定上訴勝訴，判處被告承認原告為有關地段的所有權人，並放棄使用該地段。
30	Recorre, agora, o réu A para este Tribunal de Última Instância (TUI), suscitando a seguinte questão:	現被告甲向本終審法院提起上訴，提出以下問題：
31	- O DL n.º 6/93/M tem por objetivo legislativo apenas a contenção e erradicação das edificações informais.	—第6/93/M號法令的立法原意僅在於遏制及取締僭建物。
32	Como os terrenos privados podem ser objecto das transacções de direitos, não existe qualquer motivo que obste e impeça os possuidores das barracas de invocar a usucapião quanto aos terrenos privados onde se situam as referidas barracas.	由於私有土地可以成為權利交易的標的，因此沒有任何理由妨礙及阻止木屋的占有人就上述木屋所在的私有土地提出時效取得的請求。
33	O litígio de direito sobre terrenos privados não cai no âmbito de regulação do DL em causa.	私有土地的權利糾紛不在相關法令的規管範圍之內。

Fonte: elaborado pelos autores (2020)

### 3.2 Utilização de corpus paralelo na CAT

#### 3.2.1 Criação de memórias de tradução

Por constituir um sistema de arquivos que reúne traduções anteriores ou materiais preparados previamente em diferentes momentos e situações, a memória de tradução é considerada como um recurso essencial para o tradutor, podendo possibilitar, em muitos casos, a recuperação de trabalhos anteriormente salvos ou importados, quer em nível terminológico quer em nível fraseológico.

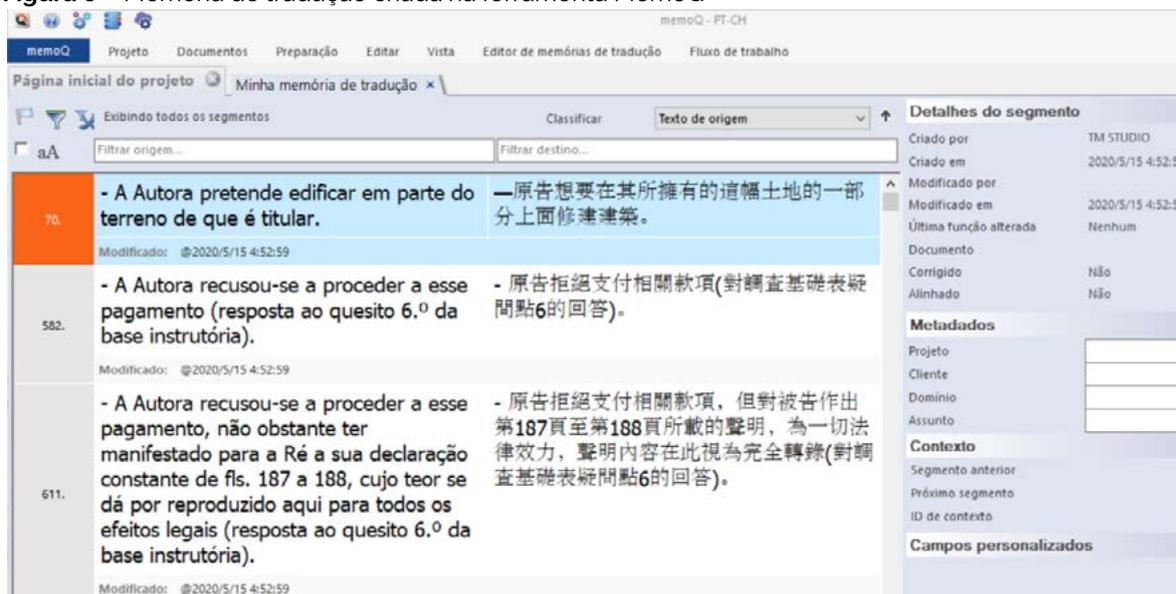
A memória de tradução pode ser formada por diversos meios, conforme Garcia (2007), incluindo o processo de tradução, o alinhamento

de recursos bilingues, a importação de dados existentes, entre outros. Um documento, se tiver duas versões, o texto original e a tradução, depois de ser adequadamente segmentado e alinhado, pode ser importado diretamente para a memória de tradução, tornando-se uma base de dados que permite diretamente a consulta, a referência e o uso.

Neste estudo, optou-se pela forma de construir um banco de dados como memória de tradução, concretizada por meio da importação de recursos bilingues devidamente alinhados (isto é, *corpus* paralelo anteriormente criado mediante Tmxmall) numa ferramenta de tradução assistida por computador – MemoQ<sup>9</sup> –, conforme figura 6.

<sup>9</sup> Disponível em: <https://www.memoq.com/>. Acesso em: maio 2020.

**Figura 6** – Memória de tradução criada na ferramenta MemoQ



Fonte: extrato do programa MemoQ (2020)

### 3.2.2 Criação de base terminológica

Uma terminologia correta e apropriada é uma condição *sine qua non* para a qualidade de qualquer tradução, especialmente no campo jurídico. De acordo com Bastos (2018), uma base de dados de terminologia jurídica contém informações valiosas, que permitem aos tradutores tomar decisões corretas sem a necessidade de uma pesquisa extensiva, podendo acelerar assim o processo de decisão na tradução de documentos jurídicos, melhorando simultaneamente a qualidade do produto final e evitando erros dispendiosos.

Em Macau, a base terminológica jurídica é imprescindível devido não só às enormes diferenças linguísticas entre o português e o chinês, mas também à especialidade da linguagem jurídica. Nesse contexto, a base a qual os tradutores vêm recorrendo é o Glossário Jurídico Chinês-Português e Português-Chinês, elaborado pela Direcção dos Serviços de Assuntos de Justiça<sup>10</sup>, doravante DSAJ, o qual pode ser consultado *online* ou *offline*.

No escopo deste estudo, o foco centra-se,

especificamente, na conversão eletrônica e no tratamento do referido glossário jurídico, com o objetivo de obter uma base de terminologia disponível, a qual pode ser utilizada diretamente pela ferramenta CAT.

O processo de criação de base terminológica divide-se em quatro etapas:

- Etapa 1: *download* – o glossário jurídico em formato PDF é baixado na página eletrônica da DSAJ.
- Etapa 2: formatação – o glossário em formato PDF é convertido para o formato Excel por meio do *software* de Reconhecimento Ótico de Caracteres (OCR) ABBYY FineReader<sup>11</sup>.
- Etapa 3: ajuste manual – trata-se de um processo semelhante ao de fazer o alinhamento de *corpus* paralelo.
- Etapa 4: importação – a terminologia fica bem alinhada no Excel, conforme figura 7, a qual pode ser importada diretamente para ferramenta MemoQ (figura 8).

<sup>10</sup> Disponível em [https://www.dsaj.gov.mo/DicWeb/DictionaryWeb\\_index.html](https://www.dsaj.gov.mo/DicWeb/DictionaryWeb_index.html). Acesso em: maio 2020.

<sup>11</sup> Disponível em <https://www.abbyy.com/en-us/finereader/>. Acesso em: maio 2020.

Figura 7 – Terminologia jurídica criada com base no glossário da DSAJ

	A	B	C	D
1	PI	CH		
3978	crime contra a segurança do Estado	妨害國家安全罪		
3979	crime contra a segurança exterior do Estado	妨害國家外部安全罪		
3980	crime contra a segurança interior do Estado	妨害國家內部安全罪		
3981	crime contra a vida	侵犯生命罪		
3982	crime contra a vida intra-uterina	侵犯子宮內生命罪		
3983	crime contra as pessoas	侵犯人身罪		
3984	crime contra bens patrimoniais	侵犯財產罪		
3985	crime contra direitos patrimoniais	侵犯財產權罪		
3986	crime contra o património	侵犯財產罪		
3987	crime culposo	有罪過的犯罪		
3988	crime de burla	詐騙罪		
3989	crime de carácter político	政治犯罪		
3990	crime de corrupção	貪污罪		
3991	crime de dano	毀損罪		
3992	crime de dano qualificado	加重毀損罪		
3993	crime de desobediência qualificada	加重違令罪		
3994	crime de detenção de armas proibidas	持有禁用武器罪		
3995	crime de emissão de cheque sem cobertura	簽發空頭支票罪		
3996	crime de emissão de cheque sem provisão	簽發空頭支票罪		
3997	crime de emprego	僱用關係方面的犯罪		
3998	crime de emprego	僱用上的犯罪		
3999	crime de especulação	投機罪		
4000	crime de espionagem	間諜罪		
4001	crime de falsas declarações	虛假聲明罪		
4002	crime de falsidade de testemunho	虛假證言罪		
4003	crime de falsificação	偽造罪		
4004	crime de furto	盜竊罪		
4005	crime de furto de uso	使用盜竊罪		
4006	crime de guerra	戰爭犯罪		
4007	crime de guerra	戰爭罪行		
4008	crime de guerra contra civis	侵犯平民的戰爭罪		
4009	crime de imigração clandestina	非法移民罪		
4010	crime de infração de regras de construção e perturbação de serviços	違反建築規則及擾亂事業罪		
4011	crime de mão própria	己手犯		
4012	crime de mão própria	親手犯		
4013	crime de omissão impróprio	不純正不作為犯		

Fonte: elaborado pelos autores (2020)

Figura 8 – Terminologia jurídica importada na ferramenta MemoQ

M	ID	Português (Portugal)	Chinês (Macau S.A.R.)
	203	acção de redução de liberalidades inoficiosas	扣減損害特留份的慷慨行為之訴
	204	acção de reforma	糾正之訴
	205	acção de registo	登記之訴
	206	acção de regresso	求償之訴
	207	acção de regulação de avaria marítima comum	共同海損理算之訴
	208	acção de reivindicação	請求返還之訴
	209	acção de resolução de contrato	解除合同之訴
	210	acção de responsabilidade	責任之訴
	211	acção de responsabilidade	追究責任之訴
	212	acção de restituição	返還之訴
	213	acção de restituição da posse	回復占有之訴
	214	acção de restituição por enriquecimento sem causa	因不當得利而請求返還之訴
	215	acção de simples apreciação	確認之訴
	216	acção de simples apreciação negativa	消極確認之訴

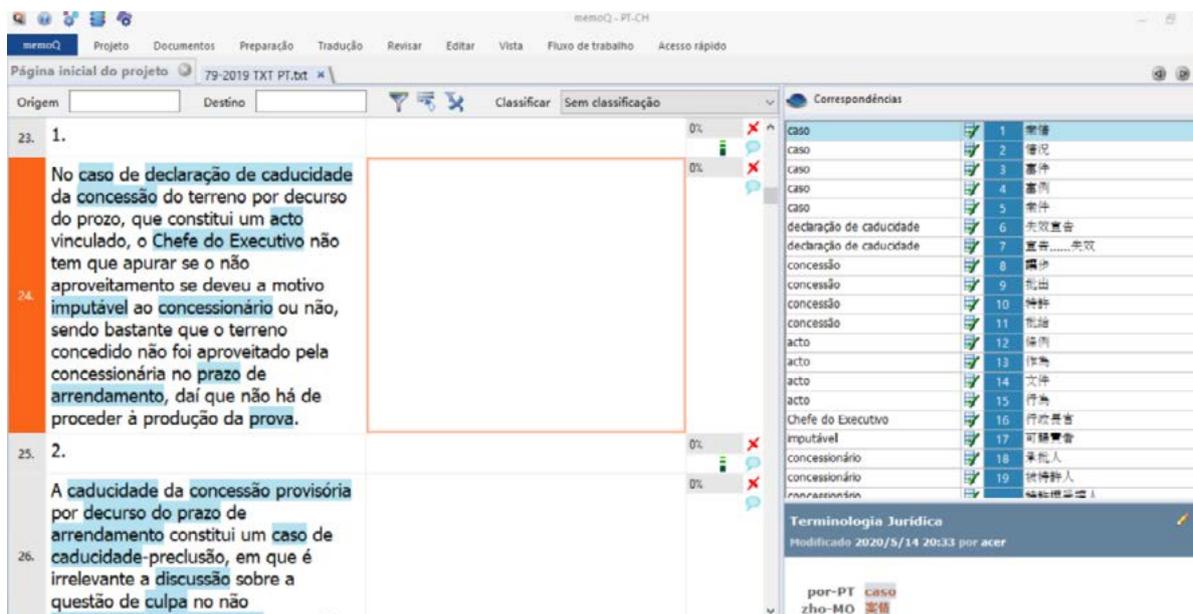
Fonte: extrato do programa MemoQ (2020)

Como se pode ver na Figura 8, foi obtido um total de 15148 termos jurídicos bilíngues, os quais, depois de serem importados para a base de terminologia da ferramenta MemoQ, encontram-se disponíveis para serem automaticamente acessados como fundamentos de referência, para o processo de tradução jurídica português-chinês.

Para um melhor entendimento, quando um

TF, em língua portuguesa, for importado para a ferramenta MemoQ, os respectivos termos jurídicos podem ser identificados e assinalados automaticamente de acordo com a base terminológica. Ao mesmo tempo, as correspondentes traduções encontram-se no canto direito para a referência do tradutor (vide Figura 9).

**Figura 9** – Termos jurídicos identificados e assinalados pela ferramenta MemoQ



**Fonte:** extrato do programa MemoQ (2020)

Na secção seguinte, apresentamos a análise e quantificação da função da ferramenta da CAT durante o processo de tradução de sentenças judiciais entre o par linguístico chinês e português, demonstrando como um tradutor humano pode fazer uso dos dados e recursos linguísticos previamente elaborados.

#### 4 Análise e discussão

Com as sentenças judiciais do TUI e o Glossário Jurídico Chinês-Português e Português-Chinês da DSAJ, conseguimos obter documentos bilíngues disponíveis para serem utilizados na ferramenta de

auxílio de tradução por computador, nomeadamente na memória de tradução e base de terminologia.

##### 4.1 Memória de tradução como recurso

A fim de observar, por meio de uma análise estatística, como um tradutor se pode beneficiar com esses recursos linguísticos, foi realizado um teste experimental. Primeiro, selecionamos de forma estocástica uma sentença monolíngue em 2019, LF em português e sem tradução em chinês. Na presente análise, a sentença selecionada foi o processo n.º 79/2019 (figura 10), a qual é composta por 4829 palavras em português.

**Figura 10** – Dados da primeira folha da sentença n.º 79/2019

Processo n.º 79/2019  
 Recurso jurisdicional em matéria administrativa  
 Recorrente: Sociedade de Investimento Imobiliário Nga Keng Van, S.A.  
 Recorrido: Chefe do executivo da RAEM  
 Data da conferência: 29 de Novembro de 2019  
 Juízes: Song Man Lei (Relatora), Sam Hou Fai e Viriato Manuel Pinheiro de Lima

**Assuntos:** - Caducidade da concessão provisória

- Produção da prova
- Caducidade-preclusão
- Causa impeditiva da caducidade
- Princípios gerais do Direito Administrativo
- Abuso de direito
- Lei Básica da RAEM

**SUMÁRIO**

1. No caso de declaração de caducidade da concessão do terreno por decurso do prazo, que constitui um acto vinculado, o Chefe do Executivo

Processo n.º 79/2019

1

**Fonte:** extrato do website dos Tribunais da RAEM (2020)

Segundo, a referida sentença (em formato TXT) com sua função de "Estatísticas" (Figura 11) foi importada na ferramenta MemoQ e analisada

**Figura 11** – TF analisado pela função de "Estatísticas" da ferramenta MemoQ

**Análise**

Escopo *"79-2019 TXT PT.txt"*  
 Recursos *Todas as TM e corpora*

Tipo	Segmentos	Palavras de origem	Caracteres de origem	Tags de origem	Porcentagem
Todos	187	4829	24631	1	100.00
Com tradução cruzada/contexto duplo	0	0	0	0	0.00
Repetições	12	43	204	0	0.89
101%	0	0	0	0	0.00
100%	10	40	217	0	0.83
95%-99%	6	25	123	0	0.52
85%-94%	1	1	9	0	0.02
75%-84%	11	361	1862	0	7.48
50%-74%	27	591	2924	0	12.24
Sem correspondência	120	3768	19292	1	78.03

**Fonte:** extrato do programa MemoQ (2020)

Para compreendermos como os resultados da análise foram produzidos, é necessário identificar o que significam as taxas de correspondência (*match rates*) demonstradas pela ferramenta

MemoQ. Nesse sentido, de acordo com as definições (Quadro 2) determinadas e reproduzidas na página eletrônica de MemoQ<sup>12</sup>, observa-se que:

<sup>12</sup> Disponível em: <https://docs.memoq.com/current/en/Things/things-match-rates-from-translation-m.html>. Acesso em: maio 2020.

**Quadro 2** – Significado de taxa de correspondência em "Estatísticas" da ferramenta MemoQ

<i>Percent value</i>	<i>Match rates</i>	<i>Meaning</i>
101%	<i>Exact match with context or Context match</i>	<i>The source text in the segment text is exactly the same as the match from the translation memory. In addition, the context of the source text is also the same as the context that was stored in the translation memory.</i>
100%	<i>Exact match</i>	<i>The source text in the segment text is exactly the same as the match from the translation memory. The context of the segment does not match – or we do not know anything about it.</i>
95%-99%	<i>Nearly exact match</i>	<i>The source text of the segment is exactly the same as the match, but there are slight differences: numbers, tags, punctuation marks and spaces might be different.</i>
50%-94%	<i>Fuzzy match</i>	<i>The source text is similar to the source text in the match, but there are already differences in the text, too.</i>

**Fonte:** Elaborado pelos autores (2020) com base nos Memoqdocs

A partir das estatísticas apresentadas na figura 9, podemos observar que a sentença a ser traduzida tem um total de 43 palavras repetidas em 12 unidades de tradução. Em comparação com os dados arquivados na memória de tradução, 10 unidades de tradução (contêm 40 palavras, que representam 0,83% do número total das palavras no TF) apresentam uma taxa de correspondência exata – 100% (*exact match*) –, isso significa que o TF, no segmento, é exatamente o mesmo que o da memória de tradução; 6 unidades de tradução (contêm 26 palavras, que representam 0,52% do número total das palavras no TF) apresentam uma taxa de correspondência quase exata – 95%-99% (*nearly exact match*) –, isso significa que o TF, no segmento, é exatamente o mesmo que o da memória de tradução, mas há pequenas diferenças em termos de números, pontuações *etc.*; 39 unidades de tradução (contêm 951 palavras, que representam 19,74% do número total das palavras no TF) apresentam uma taxa de correspondência entre 50% e 94% (*fuzzy match*), isso significa que o TF, no segmento, é semelhante ao da memória de tradução, tendo já também algumas diferenças no texto.

Considerando o que foi apresentado, 55 unidades de tradução (contêm 1018 palavras, que representam 21,09% do número total das palavras no TF) apresentam uma taxa de correspondência superior a 50%. Assim, ao fazer a tradução da referida sentença, com base na memória de tradução, o tradutor pode ter como referência, pelo menos, um quinto do respectivo conteúdo, podendo, em algumas situações, utilizar diretamente as traduções fornecidas pela memória de tradução em caso de correspondência exata.

#### 4.2 Base de terminologia como recurso de tradução

Na parte inicial da sentença n.º 79/2019, pode-se ver os "**assuntos**" que representam o conteúdo núcleo da sentença, cujos respectivos termos são relevantes no TF e podem ser considerados palavras-chave representativas para a análise. Depois da importação na ferramenta MemoQ, foram identificados pela ferramenta MemoQ, tendo como referência a base de terminologia feita anteriormente, os seguintes termos jurídicos (vide quadro 3):

**Quadro 3** – Termos corretamente identificados pela ferramenta MemoQ

Assuntos	Termos identificados pela ferramenta MemoQ	Traduções automaticamente fornecidas pela ferramenta MemoQ	Número de termos identificados
Caducidade da concessão provisória	caducidade	失效; 除斥期間	2
	concessão provisória	臨時批出	
Produção da prova	prova	考試; 證明; 證據	1
oCaducidade-preclusão	caducidade	失效; 除斥期間	1
Causa impeditiva da caducidade	causa	訴訟; 案件; 原因; 理由; 訴因	2
	caducidade	失效; 除斥期間	
Princípios gerais do Direito Administrativo	princípios gerais	一般原則; 總則	2
	direito administrativo	行政法	
Abuso de direito	abuso	濫用	2
	direito	權; 法	
Lei Básica da RAEM	lei básica	基本法	1
Total			11

**Fonte:** elaborado pelos autores (2020)

De acordo com o referido quadro 3, pode-se ver que foi identificado um total de 11 termos jurídicos, excluindo as repetições, efetivamente, restam 9 termos jurídicos, cujas correspondentes traduções foram automaticamente fornecidas pela ferramenta MemoQ. Porém, devido à falta de informações linguísticas, essas traduções, tal como o dicionário

*online*, podem apenas servir de referência.

Nesta abordagem, analisou-se, estatisticamente, o seu número de ocorrência no TF e na memória de tradução, respectivamente. Considerando-se a função de "**Concordância**" da ferramenta MemoQ, temos a seguinte informação (quadro 4):

**Quadro 4** – Número de ocorrência dos termos no TF e na memória de tradução

Termos	Ocorrência no TF	Ocorrência na memória de tradução
caducidade	45	23
concessão provisória	5	17
prova	15	11
causa	19	29
princípios gerais	7	1
direito administrativo	7	1
abuso	8	0
direito	41	32
lei básica	9	2

**Fonte:** elaborado pelos autores (2020)

Mediante a função de "**Concordância**" da ferramenta MemoQ, é possível saber o número de ocorrência do termo na memória de tradução, assim como as informações contextuais em nível lexical, frasal e textual não só no texto-fonte (TF) como também no texto-alvo (TA). Desse modo, o tradutor pode obter informações detalhadas

sobre o termo, o que lhe permite ter recursos confiáveis durante o processo de tradução; por exemplo, por meio de pesquisa da "caducidade" na "**Concordância**" da ferramenta MemoQ, pode-se ter as informações constantes da infra figura (figura 12):

**Figura 12** – "Caducidade" assinalada pela "**Concordância**" da ferramenta MemoQ

The screenshot shows the MemoQ Concordance tool interface. At the top, there are configuration options for the search, including checkboxes for 'Colocar o texto selecionado entre aspas', 'Adicionar coringas ao texto selecionado', 'Pesquisar no destino', 'Diferenciar maiúsculas e minúsculas', 'Equivalência numérica', and 'Presumir tradução'. The search term 'caducidade' is entered in the 'Texto' field. The results are displayed in a table with three columns: 'Origem', 'Destino', and 'Metainformações'. The table contains four rows of results, each showing a snippet of text from the source and target, along with a 'Minha memória de tradução' entry indicating the modification time (2020/5/15 4:52).

Origem	Destino	Metainformações
Declaração da <b>caducidade</b> do contrato de concessão.	宣告批給合同失效。	Minha memória de tradução Modificado 2020/5/15 4:52
- A <b>caducidade</b> dos autos é sancionatória e não preclusiva;	— 本案例中所涉及的失效為懲罰性失效，而不是過期失效；	Minha memória de tradução Modificado 2020/5/15 4:52
7. Aplicação da Lei de Terras no tempo e <b>caducidade</b> da concessão pelo decurso do prazo da concessão	7. 《土地法》在時間上的適用以及批給因批給期間屆滿而失效	Minha memória de tradução Modificado 2020/5/15 4:52
Declaração da <b>caducidade</b> do contrato de concessão.	宣告批給合同失效。	Minha memória de tradução Modificado 2020/5/15 4:52
VI - Decorrido o prazo de 25 anos da concessão provisória (se outro prazo não estiver fixado no contrato) o Chefe do Executivo deve declarar a <b>caducidade</b> do contrato se considerar que, no mencionado prazo, não foram cumpridas as cláusulas de aproveitamento previamente estabelecidas.	六、25 年的臨時批給期間屆滿後(如合同未另外訂定期間)，行政長官如認為在上述期間內事先訂定的利用條款未獲履行，則應宣告合同失效。	Minha memória de tradução Modificado 2020/5/15 4:52

**Fonte:** extrato do programa MemoQ (2020)

De acordo com as informações demonstradas na referida figura, pode-se saber que, foi encontrado, em 0,21 segundo, um total de 23 ocorrências de "caducidade" na memória de tradução, isto é, no *corpus* paralelo anteriormente preparado com base nas 5 sentenças selecionadas.

É especialmente digno de sublinhar que, por intermédio da localização do termo "caducidade" na memória de tradução, o tradutor pode obter 23 frases adequadamente emparelhadas em português e em chinês. Acredita-se que essas 23 frases possam ajudar o tradutor não só no aprofundamento da interpretação do termo no TF, mas também no estabelecimento da ligação ao contexto da tradução no TA. Conforme podemos identificar no *corpus* Linguee<sup>33</sup>, o tradutor é capaz de ter acesso a informações bilíngues por meio de uma simples busca de uma unidade

linguística, podendo esse ser uma palavra, uma expressão ou até uma frase.

Pelo exposto, acredita-se que a utilização de uma base de dados do campo jurídico possa contribuir para dinamizar e otimizar o processo tradutório, bem como garantir a qualidade das traduções, incluindo a tradução das sentenças judiciais de Macau, além de reforçar o papel da tecnologia CAT como uma ferramenta importante no contexto geral da tradução.

## 5 Considerações finais

O presente estudo analisou a viabilidade da utilização de um *corpus* paralelo português-chinês como um recurso linguístico associado à ferramenta CAT, cujo objetivo destaca-se, especialmente, pela relevância na utilização de *corpus* paralelo

<sup>33</sup> Disponível em: <https://www.linguee.pt/>. Acesso em: maio 2020.

como parte de uma metodologia aplicável na tradução jurídica. Para tal, elaborou-se um *corpus* constituído por 5 sentenças judiciais bilingues emitidas pelo TUI no ano de 2019, o qual foi posteriormente transformado em uma memória de tradução. Com base nesse processo, utilizamos a ferramenta MemoQ para analisar e testar, de forma estatística, uma outra sentença judicial.

Segundo os resultados evidenciados neste estudo, quando a taxa de correspondência de uma unidade de tradução se encontra igual ou superior a 50%, a ferramenta MemoQ pode fornecer traduções precisas e com alto valor de referência, as quais apenas necessitam ser confirmadas ou ajustadas pelo tradutor. Isto representa um ganho para o tradutor, não só por poupar significativamente o tempo no processo tradutório, mas também por aumentar a sua confiança na tomada de decisões e, em consequência, assegurar a qualidade das traduções efetuadas.

Além disso, em contrapartida, as memórias de tradução ainda conseguem ser um *corpus* bilingue comparável, que disponibiliza ao tradutor ricos recursos de referência, pois lhe permite entender melhor o significado do termo na LF e, conseqüentemente, estabelecer uma cadeia linguística adequada na LA. É de se presumir que, à medida que se vai aumentando a quantidade do *corpus*, as informações com valor de referência serão também em maior número, representando um valor simétrico e correspondente.

Apesar de as análises terem sido realizadas em um estudo piloto, as respectivas conclusões oferecem dados concretos de que tal estudo possa ser ampliado e aprofundado, considerando-se a elaboração de um *corpus* constituído por um número mais significativo de sentenças judiciais, ou seja, mais robusto em termos de vocabulário e formas estruturais. Em simultâneo, vale a pena mencionar que uma importante contribuição deste estudo está ainda relacionada com a possibilidade do desenvolvimento de uma metodologia para auxiliar no processo de solução da questão de tradução técnica e especializada, não apenas entre o par linguístico português-chinês, como também para outras línguas.

## Referências

- BAKER, Mona. Corpora in translation studies: An overview and some suggestions for future research. *Target. International Journal of Translation Studies*, v. 7, n. 2, p. 223-243, 1995. <https://doi.org/10.1075/target.7.2.03bak>
- BASTOS, Maria do Céu Henriques de. *Terminology for translators: a termbase in the domain of company law*. PhD diss. Vigo: Universidade de Vigo, 2018.
- BOWKER, Lynne. *Computer-aided translation technology: A practical introduction*. University of Ottawa Press, 2002.
- CAPEZ, Fernando. *Curso de processo penal*. 19. ed. São Paulo: Saraiva, 2012.
- CHEN, Pearl. *Using comparable specialized corpora with machine translation for extracting N-gram translation equivalents: a case study of Chinese and English contracts*. PhD diss. Taipei: National Taiwan Normal University, 2012.
- GARCIA, Ignacio. Power shifts in web-based translation memory. *Machine Translation*, v. 21, n. 1, p. 55-68, 2007. <https://doi.org/10.1007/s10590-008-9033-6>
- HARTLEY, Tony. Technology and translation. In: MUNDAY, J. *The Routledge companion to translation studies*. London and New York: Routledge, 2009. p. 106-127.
- HU, Xianyao. *A Corpus-based Study on the Translational Norms of Contemporary Chinese Translated Fiction*. PhD diss. Shanghai: East China Normal University, 2006.
- HU, Xingwen; WU, Amiao; SHU, Xuejin. Research on School Classification of Western Translation Theory. *Journal of Hefei University of Technology (Social Science)*, v. 25, n. 6, p. 142-149, 2011.
- HU, Zhihua. *Sobre a Análise e a Tradução das Partículas Auxiliares do Chinês para o Português com base num corpus paralelo chinês-português*. PhD diss. Aveiro: Universidade de Aveiro, 2019.
- LOPES, Alba Valéria Sabóia Teixeira. *O plano de texto no gênero sentença judicial*. Textos e discursos no Direito e na Política: análises e perspectivas interdisciplinares. Coimbra: Grácio Editor. 2016. p. 27-37.
- MAI, R. *Ensino de chinês a falantes de português: o caso da Universidade de Aveiro*. 2012. 482 p. Tese (Doutoramento em Linguística) – Departamento de Línguas e Culturas, Universidade de Aveiro, 2012.
- NEWMARK, Peter. *A textbook of translation*. New York: Prentice hall, 1988. Vol. 66.
- SARDINHA, Tony Berber. *Linguística de corpus*. Editora Manole Ltda., 2004.
- STUPIELLO, Érika Nogueira de Andrade. *Ética profissional na tradução assistida por sistemas de memórias*. 2014.
- WANG, Li; HAN, Xiwu. Application of Bilingual Dictionary in Statistical Machine Translation. *Computer Engineering and Applications*, n. 16, p. 135-139, 2010.

---

### Wang Tianlong

Doutorando em Linguística Aplicada, Departamento de Português, FAH, Universidade de Macau. Tradutor da C&C Sociedade Gestora de Escritórios de Advogados Limitada.

---

### Ana Luisa V. Leal

Doutora em Informática – Universidade de Évora, Portugal, Mestre em Linguística PUCRS, e Especialista em LP, FAPA. Professora Assistente e Pesquisadora na Universidade de Macau.

---

### Endereço para correspondência

Wang Tianlong

Avenida da Praia Grande, 759, 5º andar

Macau, China

Ana Luisa V. Leal

Avenida da Universidade

E21, 3º andar, gabinete 3078

Macau, China