

---

HU, Kaibao. **Introducing Corpus-based Translation Studies.** *New Frontiers in Translation Studies*, 2016, 258 p.



Letícia Freitas de ASSIS\*  
Universidade Federal do Ceará, Brasil.

**RECEBIDO EM:** 20 de agosto de 2019

**APROVADO EM:** 10 de dezembro de 2019

**ACEITO EM:** janeiro de 2020

O livro *Corpus-based Translation Studies*, de Kaibao Hu, foi escrito para responder questionamentos gerais sobre os Estudos da Tradução Baseados em *Corpus*. Os objetivos de sua publicação foram: proporcionar considerações sobre as principais áreas dos Estudos da Tradução Baseados em *Corpus* (doravante, ETBC) e tentar fornecer fundamentação teórica tencionando estabelecer o campo de pesquisa como um novo paradigma nos Estudos da Tradução (EdT).

297

Os ETBC podem ser considerados um marco nos EdT, visto que trouxeram a possibilidade de mudança da abordagem tradicional prescritiva dos EdT para métodos descritivos típicos das pesquisas empiristas. Os ETBC focam em descrever características e normas da tradução, na tentativa de descobrir a natureza tradutória e a relação entre tradução e cultura. Para tanto, a tradução é analisada como produto e como processo por meio investigação estatística de *corpora*. O termo foi introduzido pela primeira vez em 1996 por Mona Baker em seu artigo *Corpus-Based Translation Studies: The Challenges that Lie Ahead*. O campo surgiu da confluência entre os Estudos Descritivos da Tradução e a Linguística de Corpus. Primeiramente, a pesquisa em ETBC é baseada na análise de uma coleção considerável de textos produzidos naturalmente. O foco dessa observação é identificar padrões nos textos traduzidos. Outra característica importante é a combinação de análise quantitativa e qualitativa. O uso de *corpus* permite que o pesquisador obtenha dados estatísticos de textos traduzidos, sendo esses relativos tanto ao nível lexical, quanto ao sintático e ao textual. A interpretação dos dados acontece tencionando comprovar características gerais e normas de tradução. Os ETBC, método que vem sendo aplicado desde os anos 90, são considerados uma abordagem que

---

objetiva analisar o conjunto de normas abstratas que regem uma língua natural. Para tal, julgamentos subjetivos e conceitos pré-estabelecidos pelos tradicionais EdT são evitados nessa perspectiva. As pesquisas na área podem ser descritas como *bottom-up*, tendo ênfase na verificação de hipóteses e sendo baseadas na análise de textos autênticos. Além disso, a Linguística de Corpus fornece metodologia para os ETBC, enquanto os Estudos Descritivos da Tradução proporcionam a fundamentação teórica.

Entre os anos de 1993 e 1998, a importância do uso de *corpus* foi questionada e discutida mais detalhadamente. Baker (1993, p. 243-245 *apud* HU, 2016, p. 6) cunhou o termo “universais da tradução” como características linguísticas que acontecem nos textos traduzidos e que estão livres da influência dos pares linguísticos específicos envolvidos no processo tradutório. Baker desenvolveu uma metodologia baseada em corpus para investigar os universais da tradução. Essa pesquisa ocorreu por meio da análise da coletânea de diversos tipos de *corpora*, tanto em língua de partida (LP) como em língua de chegada (LC), visto que a pesquisadora indicava que “textos traduzidos registram eventos comunicativos genuínos que não são nem inferiores tão pouco superiores em relação a outros eventos comunicativos em qualquer língua”<sup>1</sup> (BAKER, 1993, p. 248 *apud* HU, 2016, p. 6, tradução minha). Os universais da tradução podem ser considerados produtos da tradução que são inerentes ao processo tradutório, fator que comprova a validade da nomeação de “universal”. Em 1998, com o lançamento da série de artigos *The Corpus-Based Approach: A New Paradigm in Translation Studies*, foram discutidas várias particularidades dos ETBC. O volume foi dividido em duas partes, na primeira foram discutidos escopo, objeto de estudo e metodologia. A segunda parte compreendeu o empirismo aplicado aos EdT, bem como a formação de tradutores. Baker e outros pesquisadores construíram o primeiro *corpus* de traduções do mundo, nomeado *Translational English Corpus* (TEC). Composto por mais de 6,6 milhões de palavras, o TEC é um *corpus* monolíngue de traduções de línguas europeias e não-europeias para Língua Inglesa. Existem outros *corpora* famosos, como o *The Multilingual Parallel Corpus of the Bible*, *The English-Norwegian Parallel Corpus*, entre outros. A expansão da coletânea de *corpora* é fator que apoia a disseminação dos ETBC, bem como o crescimento no número de artigos, monografias e publicações em geral também auxiliam no desenvolvimento das pesquisas na área.

Para os ETBC, *corpora* podem ser divididos em *corpus* paralelo, *corpus* comparável, *corpus* de traduções e *corpus* de interpretação. A compilação desses envolve seleção de texto, de anotação e de alinhamento. O primeiro tópico trata dos critérios para seleção, visto que para

---

Zanettin (2000, p. 105 *apud* HU, 2016, p. 17) a comparabilidade de textos é elemento questionável, independente da diligência na escolha. Ainda, o pesquisador estabelece como parâmetros para triagem o meio em que o texto foi publicado, a disponibilidade tanto do texto de partida como do texto de chegada, a data de publicação e a aceitação/popularidade do texto. Para Kenny (2001, p. 108 *apud* HU, 2016, p.17), a escolha é baseada no domínio ao qual os textos pertencem, na situação em que eles foram produzidos e recebidos, no período em que eles foram elaborados e no seu meio de divulgação. Anotações são necessárias para obter informações do autor e da fonte de um texto, bem como de partes do discurso e de propriedades sintáticas. Para Monti *et al* (2005 *apud* HU, 2016, p. 17) temos as anotações linguísticas, paralinguísticas e extralinguísticas. O alinhamento entre textos é essencial para o agrupamento de *corpora* paralelos. Para Kenny (2001, p. 120 *apud* HU, 2016, p. 17) o método básico de alinhamento é feito ao estabelecer equivalência entre a primeira e a segunda frase de cada texto.

O uso de *corpora* nos EdT é fundamental para a pesquisa em tradução em virtude da necessidade de comprovação metodológica-científica para solucionar hipóteses. Por esse motivo, a constituição de *corpora* é considerada tarefa primordial em ETBC. Ainda, esse paradigma tem a propriedade de ser um campo de estudo independente dos EdT e fornece evidências para sustentar comprovações. As áreas de pesquisa desenvolvidas dentro dos parâmetros dos ETBC são: compilação de *corpora* para EdT, características de tradução, estilo do tradutor, normas de tradução, prática de tradução, formação de tradutores e estudos de interpretação. A compilação de *corpora* para EdT compreende a elaboração de *corpus*, seleção de texto, captação de dados, anotação e alinhamento de textos. Os ETBC sobre características de tradução abrangem os universais da tradução e as características de traduções entre pares linguísticos. O primeiro tem importância teórica e discorre sobre padrões encontrados nos textos traduzidos. O último relata as particularidades da tradução nos níveis léxico, sintático e discursivo, além de investigar o uso de um certo léxico ou estrutura sintática. Esse campo de pesquisa descreve as diferenças entre as línguas de chegada e de partida. Ademais, também mostra a escolha do tradutor e o consenso entre as línguas. Sobre estilo do tradutor, a pesquisa na área investiga as idiossincrasias do tradutor quanto a léxico, estruturas sintáticas, pontuação e estruturas do discurso. Também são examinadas as motivações para essas manifestações, sejam elas de cunho linguístico, cognitivo, social ou cultural. Para normas de tradução, são exploradas as escolhas tradutórias dentro de um contexto sociocultural. A linha de pesquisa de prática de tradução explora tanto o uso de metodologia de *corpus* para resolução de problemas encontrados durante o ato tradutório, bem como otimização da qualidade da tradução. Podem

---

ser considerados subcampos de estudo a aplicação de *corpora* para: identificar equivalência de LP em LC, averiguar as regularidades na transferência linguística e observar estratégias e métodos de tradução. Os ETBC voltados para formação de tradutores são baseados na compreensão de regularidades a serem seguidas durante a transferência linguística. Os aprendizes entendem padrões a partir da observação de variadas amostras de tradução. Por fim, os estudos de interpretação são fundamentados nas regras e nos aspectos principais da interpretação. Os pesquisadores da área fazem uso de grande quantidade de dados analisados estatisticamente ou quantitativamente, com objetivo de assegurar confiabilidade para suas constatações.

Hu (2016, p. 35) define *corpus* como um banco de dados composto por materiais linguísticos de ocorrência natural e coletado de acordo com processo seletivo específico. As amostras de *corpus* devem ser digitais, bem como representativas para a pesquisa. Podemos dividir *corpora* para ETBC segundo diversos critérios. Geralmente os tipos são: *corpus* paralelo, *corpus* comparável, *corpus* de tradução e *corpus* de interpretação. Hu (2016, p.35) indica a definição de *corpus* paralelo como textos A e B, sendo A o texto fonte e B a sua tradução. Também é possível dividir o *corpus* paralelo, de acordo com o número de línguas envolvidas, entre duas subcategorias: bilíngue e multilíngue. O primeiro inclui textos em duas línguas que mantêm uma relação tradutória entre si. O segundo é composto por um texto em uma língua e vários outros textos em línguas diferentes. Além disso, podemos dividir *corpus* paralelo conforme a direção da tradução: unidirecional (língua X para língua Y), bidirecional (língua X para língua Y, língua Y para língua X) e multidirecional (língua X para língua Y, Z, W e outras). Para selecionar *corpus* paralelo é necessário fazer o alinhamento dos textos, que significa identificar a relação tradutória entre o texto de partida e o texto de chegada nos níveis do léxico, da sentença ou do parágrafo. Essa análise é baseada na investigação do léxico ou do tamanho das sentenças. O uso desse tipo de *corpora* é interessante no sentido da possibilidade de visualização simultânea das amostras, assim como de análise e de exploração de padrões tradutórios. Hu (2016, p. 41) explica que *corpus* comparável consiste em textos que têm formato e estrutura comparáveis entre si. Johansson (1998, p. 136 *apud* HU, 2016, p. 41) afirma a existência de dois tipos de *corpora* comparável. O primeiro é composto por textos originais, sem relação de tradução entre si, mas ainda assim comparáveis e feitos em mais de uma língua. O segundo tipo é formado por textos originais comparados às suas traduções em uma ou mais línguas. O *corpus* comparável pode ser monolíngue, bilíngue ou multilíngue. Estabelecer a relação de comparabilidade dos textos é essencial para esse modelo. Para isso, é necessária a

---

avaliação por meio de critérios extra e intralinguísticos. Os critérios extralinguísticos são determinados em conformidade com o período de publicação e o impacto social dos textos, tal como status social e gênero do autor. Referente aos parâmetros intralinguísticos, temos observações sobre tamanho, assunto e gênero do texto. O *corpus* de tradução consiste exclusivamente de textos traduzidos. Apesar disso, deve ser acompanhado por *corpus* contendo o texto original. Por último, o *corpus* de interpretação é composto por textos transcritos de arquivos de vídeo ou de áudio. O propósito dessa prática é a investigação de estratégias de interpretação, assim como de suas características linguísticas, de duas normas e dos processos cognitivos realizados durante o desempenho da atividade.

A pesquisa em ETBC é iniciada com a coleta de *corpus*. Para tal, é necessário que o *corpus* seja planejado, e que também exista propósito e significância em seu uso. Os objetivos na coleta de *corpus* devem ser: o estudo das regularidades na transferência linguística, as estratégias para efetuar essa tarefa, o estudo das características e normas tradutórias, o estudo da interpretação, a formação de tradutores e o estudo da influência da tradução em uma língua específica. Outros itens a serem levados em consideração durante a escolha de *corpora* são representatividade e tamanho. O pesquisador deve estar consciente da dimensão das suas amostras desde o momento inicial da coleta de *corpora*. De modo geral, conseguiremos identificar melhor a manifestação do uso da língua em questão caso utilizemos *corpora* maiores. Entretanto, o tamanho dos exemplares não representa o uso da língua em sua totalidade, outrossim uma pequena parcela do idioma em questão. Portanto, deveremos definir o tamanho do *corpus* de acordo com sua idiossincrasia e seu propósito, em vez de apenas coletar o maior número de dados possíveis. Representatividade do *corpus* é um tópico relacionado à capacidade dos *corpora* de caracterizar um gênero ou uma variação linguística. O pesquisador em ETBC deve estar atento ao aspecto supracitado, visto que a falta de representatividade nos *corpora* afeta a legitimidade e a confiabilidade de sua pesquisa.

Ferramentas para compilação de *corpora* são indispensáveis nos ETBC. Processos como formatação, anotação de parágrafos, segmentação por palavras, alinhamento, pesquisa e contabilização de frequência podem ser realizados por meio do uso de recursos como o *EmEditor*, *ICTCLAS*, *ParaConc* e *Wordsmith*. O *EmEditor* é usado primariamente para formatar e marcar partes de textos para processamento. O *ICTCLAS* é um sistema de análise de *corpus* em nível lexical para o Mandarim. O *ParaConc* é desenvolvido para linguistas, tradutores e pesquisadores em EdT. É um *software* multifuncional de análise linguística que possui ferramentas para alinhamento de texto, contabilidade de frequência de palavras, dentre

---

outros. O *WordSmith* é uma das ferramentas mais utilizadas nos estudos baseados em corpus, e é geralmente utilizado para produção de concordâncias e de listas de palavras-chave. O recurso “*Concord*” é usado para consultar a frequência de uso de uma palavra ou de uma frase, já o “*WordList*” é utilizado para gerar uma lista de frequência de uso de palavras. Por último, o recurso “*KeyWords*” fornece uma lista de palavras-chave que pode ser comparada e analisada para obtenção de informações sobre o estilo do tradutor. Por conta da necessidade de pesquisa em ETBC, existem diversos *corpora* já compilados em diferentes línguas. Podemos citar alguns como o *Translational English Corpus* (TEC) e o *The Babel Chinese-English Parallel Corpus*.

Textos traduzidos manifestam diversos aspectos linguísticos. Dentre eles podemos considerar os universais da tradução como fatores resultantes da realização do processo tradutório. Recuperando a definição anteriormente explicitada, os universais da tradução são expressões linguísticas que são resultado do processo tradutório, sendo esses independentes das diferenças entre as línguas de partida e de chegada. Os universais da tradução são “típicas expressões translinguísticas genéricas presentes nos textos de chegada, sendo esses diferentes dos textos de partida.”<sup>2</sup> (KE, 2005, p. 303-307 *apud* HU, 2016, p. 85, tradução minha). Antes dos ETBC, alguns pesquisadores também exploraram as características de tradução. Suas pesquisas eram baseadas em poucos textos fonte e, por isso, tinham perspectivas limitadas. Por exemplo, durante esse período, traduções de máquina feitas em larga escala não faziam parte dos EdT. Os procedimentos de análise de textos traduzidos eram: coleta, comparação e análise manual. Em resumo, os estudos dessa época eram de natureza qualitativa e dissertavam sobre léxico, sintaxe, nível estilístico. A maioria dos estudos da fase anterior aos ETBC eram baseados em transformações interlinguísticas, e centravam esforços em desvendar aspectos contrastivos dentro do contexto LP *versus* LC. Essa metodologia tradicional sofria com problemáticas de tamanho e de qualidade das amostras coletadas.

Baker (1993 *apud* HU, p. 89) apresentou hipóteses sobre as características universais de textos traduzidos, além de dissertar sobre a viabilidade e as implicações do uso de *corpora* no estudo dos universais da tradução. A pesquisadora defendeu o uso de *corpus* comparável para observar diferenças entre os originais e suas traduções, e para entender traços deixados pelo processo tradutório ou pela interferência das duas línguas em questão. Em relação a pesquisas mais recentes, os ETBC têm como foco estudos de universais da tradução como explicitação, implicação, simplificação e normalização. A explicitação, de acordo com o que foi apontado inicialmente por Blum-Kulka (1986 *apud* HU, p. 89), é o ato tradutório de inserir interpretação sobre informações do texto de partida no texto de chegada, tencionando dar a esse maior nível

---

de coesão e clareza. Essa ferramenta pode ser considerada uma estratégia universal na mediação entre línguas, cujo uso carrega intenção de esclarecimento sobre quaisquer sentidos implícitos em itens do texto fonte.

Os tipos de explicitação podem ser organizados em diferentes categorias: explicitação formal *versus* explicitação de sentido, explicitação sintática *versus* explicitação semântica, explicitação obrigatória *versus* explicitação opcional, dentre outras classificações que serão regidas de acordo com a perspectiva adotada. As causas para esse aspecto tradutório são variadas, visto que existem diferentes fatores que influenciam o uso desse universal da tradução. Dentre os elementos motivadores estão as diferenças sistêmicas e culturais das línguas em questão, além do próprio tradutor. O autor usa como exemplo línguas cujos níveis de formalização são diferentes. Para esse caso, teremos uma grande tendência de uso da explicitação no ato tradutório. Outro exemplo é de traduções envolvendo línguas cujos enquadramentos culturais são muito diferentes entre si. Quando isso ocorre a ponto do leitor não conseguir compreender o texto, torna-se necessária a intervenção mediadora do tradutor em relação às diferenças culturais realizando o processo de explicitação do tópico. Podemos definir a implicação como uma transformação do conteúdo explícito em conteúdo implícito e compreensível por meio de análise contextual. Essa omissão não exerce influência negativa no entendimento do receptor, além de ajudar a tornar o texto alvo suscinto e conciso. A implicação pode ser categorizada de duas formas: sintática e semântica. A primeira ocorre quando uma estrutura sintática explícita é transformada em material implícito. Para a segunda, temos a conversão de significado exposto de forma clara para forma implícita. Segundo Zipf (1949, *apud* HU, 2016, p.100), o Princípio do Menor Esforço indica que tarefas podem ser performada com maior benefício e pouco esforço. Essa teoria fundamenta a motivação para implicação, tendo em vista que, desde que a comunicação seja assegurada, o tradutor tem a função de facilitar o efeito comunicativo de um texto. A simplificação é considerada por Baker (1996, p. 176 *apud* HU, p. 101) uma redução da estrutura informacional original baseada na reestruturação complexidades linguísticas e na utilização de palavras funcionais na tradução. Antes do uso de *corpora* nos EdT, a maioria dos estudos relacionados a simplificação eram produzidos analisando nível lexical, sintático ou estilístico. Essas pesquisas indicaram que a simplificação ocorre em quatro aspectos: a) substituição de hipônimos por hiperônimos, b) simplificação de sentenças complexas e redundantes, c) redução de repetição, omissão de informação redundante, exclusão de modificadores e frases, e d) adequação e reescrita por conta de diferenças culturais. A normalização é definida como a tendência de textos traduzidos a levar

---

em consideração as normas da língua alvo em termos de pontuação, escolha lexical, estrutura sintática e textual. Essa categoria de universal da tradução leva em consideração o ato tradutório de corrigir uso impróprio de pontuação, de adequar termos de acordo com questões culturais, de usar estruturas sintáticas mais simples para traduzir unidades peculiares, de transformar linguagem informal em formal e de organizar sentenças, parágrafos e capítulos, com objetivo de conferir maior nível de coesão ao texto.

É importante notar que o campo de estudo de aspectos tradutórios, como os universais da tradução, tem progredido substancialmente principalmente após o surgimento dos ETBC. O desenvolvimento rápido da tecnologia computacional agilizou a compilação de *corpora* e, por consequência, gerou crescimento do seu uso para diferentes propósitos. Dentre esses, também podemos citar o estudo do estilo do tradutor baseado em *corpus* como uma das maneiras da utilização de análise de *corpus* para obtenção de dados. O livro apresenta um panorama sobre as pesquisas, tanto antigas como atuais, na área de estudos sobre estilo do tradutor e normas de tradução baseados em *corpus*. O estilo do tradutor é, de acordo com Hu (2016, p. 123), referente às escolhas linguísticas e também às opções não-linguísticas que o profissional faz durante a tradução. Também estão inclusos os métodos e estratégias de tradução, o uso de notas de tradução e notas de rodapé. Em linhas gerais, o estilo do tradutor está conectado diretamente às preferências de uso da língua e os padrões utilizados durante seu trabalho. Dentre as acepções teóricas antigas dos EdT, podemos esclarecer o conceito de invisibilidade do tradutor em relação ao estilo. O ponto foi adotado com objetivo de fornecer suporte à ideia de que a tradução é um derivado do texto original, visto que a tradução era considerada apenas uma troca linguística, uma imitação ou uma reprodução do original. De acordo com a antiquada interpretação de que a principal intenção da atividade tradutória era a transmissão de conteúdo, o maior nível de fidelidade possível ao original era o resultado esperado da finalização do processo tradutório.

Durante os anos 60, os EdT sofreram uma transformação e passaram a compreender o valor cultural do texto fonte, levando em consideração que o trabalho tradutório tem influência de outros aspectos senão o linguístico. A partir disso, as pesquisas começaram a investigar elementos dos textos fonte e dos textos alvo, tencionando analisar comparativamente e chegar a conclusões quanto a qualidade da tradução, tomando como base o registro e o gênero apresentados nos textos em questão. A voz do tradutor também começou a ser estudada por conta do argumento de Hermans (1996, p. 27 *apud* HU, 2016, p. 126) de que na tradução estavam presentes mais de uma voz discursiva e que a do tradutor ficava por vezes escondida

---

atrás da voz do narrador. Além disso, o pesquisador em questão também afirma que o tradutor deve usar sua voz de forma auxiliar ao texto, de uma perspectiva de fora, mas ainda assim contextualizada por meio de uso de notas de rodapé ou notas do tradutor ao final do texto traduzido. O estilo do tradutor também é explicado como um tipo de impressão digital expressada na forma de itens linguísticos ou não linguísticos. Os estudos que tratam dos aspectos do estilo do tradutor trazem descrições, análises, comparações e interpretação de padrões linguísticos sutis encontrados com a investigação de *corpora* na intenção de encontrar provas convincentes de que existem marcas estilísticas deixadas pelo tradutor. As pesquisas sobre a questão abordam, majoritariamente, duas perspectivas. A primeira está relacionada a análise linguística de aspectos tradutórios, a segunda trata do uso de estratégias e métodos de tradução. A abordagem analítica de elementos é conduzida no nível do léxico, da sentença, da colocação e da estrutura discursiva. A outra compara as estratégias e métodos de tradução adotados por diferentes tradutores ao realizar seu trabalho com léxico e estruturas sintáticas específicas. Os diferentes estilos de tradução podem ser comparados por meio da observação de itens em textos traduzidos e não traduzidos. Essas análises deverão ser feitas com foco em identificar ampliações, omissões ou ajustes de estruturas sintáticas, bem como em identificar as criações do tradutor.

305

A área de estudo referente à pesquisa sobre normas de tradução tem um histórico de mais de 2000 anos. O autor levanta um questionamento importante: o que é a norma tradutória? De acordo com Hermans (1999, p.80 *apud* HU, 2016, p. 137), as normas são regularidades de comportamento. Esse conceito pode ser estendido de regularidades para regras e convenções. A atividade tradutória é executada de acordo com o processamento de decisões, sendo essas motivadas por atos pessoais ou por restrições padronizadas. As normas também são vistas como um fenômeno sociocultural com suas idiossincrasias, além de carregarem restrições e imposições sob atos sociais e comportamentais. Durante a tradução, o tradutor entra num processo de tomada de decisão desde o momento da seleção de textos a serem traduzidos até a escolha de estratégias e métodos que reflitam conformidade em relação aos valores e princípios evidenciados naquele contexto tradutório. As normas de tradução retratam principalmente a conexão do tradutor com o mundo e com a sociedade, visto que ficam estabelecidas ligações abstratas entre a criatividade e a regra. O tradutor tem a liberdade de escolha de concordar ou não com as normas tradutórias desde que entenda as consequências do seu ato. O ideal é que o tradutor adeque seu trabalho ao contexto dos leitores da LC enquanto usa sua subjetividade e criatividade para execução da tarefa.

---

A pesquisa nessa área tem histórico de longa data, por esse motivo é dividida em três períodos de pesquisa. No primeiro é identificada uma abordagem filológica, que inclui pesquisas investigando séries de normas sobre a constituição de uma boa tradução e sobre como essa atividade deve ser feita. Essas regras têm cunho apenas prescritivo e são geradas com base em comparações entre os textos de partida e de chegada. Entretanto, as normas prescritivas elaboradas pelas pesquisas baseadas em análises filológicas não levam em consideração a interação entre tradução, determinações socioculturais e ideológicas. No segundo período surgem as pesquisas com abordagem linguística-funcional, cujos princípios dissertavam em torno das seguintes especificações: a) o propósito do texto alvo determina a seleção de estratégias e métodos usados da tradução, visto que a funcionalidade do texto de chegada é valorizada como a motivação para sua tradução, b) o texto alvo é a oferta de informação na língua alvo, c) a função do texto alvo não necessariamente é a mesma do texto fonte, portanto não é possível fazer uma simples reversão de um para o outro, d) o texto de chegada deve manter a coerência interna e também externa, em relação ao texto fonte. O último período antes dos ETBC tratava das normas tradutórias a partir de um ponto de vista descritivo. Sendo assim, as pesquisas nesse período indicavam que a tradução deveria ser determinada de acordo com a história, a sociedade e a ideologia. O argumento em comum para essa etapa era de que a tradução deve ser regida de acordo com características socioculturais específicas.

O uso de *corpus* nas pesquisas de normas de tradução facilitou a mudança do cunho prescritivo para aspectos cada vez mais descritivos de análise. Baseado em *corpora*, o pesquisador pode chegar a conclusões gerais sobre normas, tendo como suporte a descrição das características do texto traduzido. A importância em entender as diferentes normas tradutórias vem da natureza de ferramenta de pesquisa que elas apresentam, visto que cada língua apresenta um tipo de procedimento. “O comportamento tradutório é, na verdade, relativo a cada língua traduzida em particular, portanto as características da língua traduzida podem ser percebidas como fatores distintivos para a tradução em relação à textos não traduzidos na mesma língua.”<sup>3</sup> (HU XIANYAO, 2008 *apud* HU, 2016, p. 147, tradução minha). É possível reconstruir as normas de tradução a partir da investigação baseada em análise de *corpora*, além de analisar suas características com base no contexto sociocultural e histórico. A reconstrução deve ser feita de acordo com as condições de que os dados linguísticos sejam autênticos, abundantes e representativos. De fato, a ligação entre norma, hipótese e *corpus* é bidirecional, ou seja, a pesquisa baseada em *corpus* serve para comprovar hipóteses relacionadas ao uso de normas. Além disso, o pesquisador pode fazer uso da ferramenta de duas principais formas. A primeira

---

é a possibilidade de listagem de normas restritivas por meio de análise comparativa entre texto fonte e texto alvo. A segunda é a descoberta de normas relativas ao par linguístico em questão mediante análise de características da tradução. Portanto, é interessante notar que o tradutor está sujeito às restrições de variadas normas, sendo essas regidas por diversos fatores como sociedade, cultura, língua, convenções literárias, aceitação por parte dos leitores e decisões tomadas pelo tradutor por conta de particularidades culturais.

A investigação da norma deve ser feita com fundamento no estudo de ações e produtos tradutórios específicos. Também é importante lembrar que a existência de regularidades no objeto de pesquisa é pré-requisito para que a pesquisa descritiva seja executada. Segundo Toury (1999, *apud* HU, 2016, p. 150) as normas tradutórias podem ser investigadas de acordo com contexto intratextual e extratextual. Para o primeiro, constam as verificações de partes do texto como prefácio, apêndice, bem como de elementos linguísticos da tradução em pauta. No segundo estão incluídas as observações sobre comentários e teorias acerca da tradução. Os procedimentos indicados para o estudo das normas podem ser divididos em três etapas. Durante a primeira, é necessário descrever e analisar as particularidades da tradução a serem investigadas. Isso deve ser feito a nível lexical, sintático ou textual. Em segundo lugar, as normas devem ser organizadas a modo que comprovem ou modifiquem as hipóteses previamente estabelecidas. Na última etapa, é feita uma descrição com perspectiva sociocultural das motivações para aquelas normas. Dessa maneira é possível traçar um percurso comportamental da tradução, bem como explicar retraduições ou adaptações necessárias no texto de chegada.

A seguir, é explicado o uso de *corpora* nas pesquisas sobre correspondência entre LP e LC, sobre o uso de estratégias de tradução, sobre avaliação de qualidade da tradução e sobre tradução auxiliada por computador. O estudo da prática tradutória faz referência direta à análise de correspondência entre LP e LC à nível lexical e sintático, bem como à investigação de equivalentes terminológicos para itens da LP. Os resultados desses estudos têm influência direta na solução de unidades problemáticas, além de ajudar diretamente os tradutores a aplicar diferentes métodos de tradução. Para isso, é indicado o uso de *corpora* paralelo, visto que esse material torna possível a avaliação comparativa em contexto bilíngue. Além disso, *corpora* paralelos podem ser considerados fonte de recurso informativo para o tradutor, dado que funcionam como um sistema de memória de tradução. O estudo de identificação lexical baseado nas análises desse tipo de *corpora* consiste no reconhecimento de equivalentes linguísticos da LP na LC. As pesquisas sobre estrutura sintática usam *corpus* paralelo alinhado para

---

comparação sintática de LP e LC. Ainda, o uso de *corpora* paralelo alinhado é considerado útil para avaliar o nível de qualidade da tradução. Isso ocorre em três etapas. Durante a primeira é realizada a coleta de textos naturais da área pré-selecionada. Na segunda etapa, a investigação de estruturas decorrentes de contextos espontâneos ocorre por meio do uso de tecnologias, tencionando apontar padrões decorrentes de situações genuínas. Por último, as características quantitativas e qualitativas são combinadas e avaliadas. Hu (2016, p. 161-162) afirma que os dicionários bilíngues são uma ferramenta cada vez menos prática na identificação de equivalentes terminológicos quando comparados ao uso de *corpus*. Isso pode ser justificado pela automação causada pela análise de *corpora*. Apesar de todas as vantagens apresentadas sobre o uso de *corpora*, o autor também aponta que a confiabilidade excessiva em seu uso pode afetar negativamente a competência e a criatividade do profissional durante a tradução.

O autor também disserta sobre o uso de *corpora* na formação de tradutores e na elaboração de livros de tradução. O uso de *corpora* na tradução tem sido útil no ensino-aprendizagem de alunos de tradução nos sentidos de conscientização sobre o ato tradutório e de desenvolvimento de competências. Por um longo tempo, o ensino de tradução era apenas baseado na experiência e nas intuições do professor. Porém, o uso de *corpora* começou a ser discutido nos anos 90, bem como surgiram publicações de estudos mais aprofundados sobre o uso de *corpora* até mesmo em avaliações. De fato, a ligação entre a treinamento de tradutores e o uso de *corpora* é uma inovação para a didática da tradução, visto que esse formato pode superar limitações presentes no ensino tradicional de tradução. O uso de *corpora* comparável, por exemplo, é útil para que os estudantes entendam normas e características de tradução. Isso auxilia na compreensão do que é traduzir. Um dos usos didáticos apontados pelo autor é a atividade de análise de comportamento tradutório, que pode ser realizada pedagogicamente como uma investigação de estruturas sintáticas ou de características semânticas de uma palavra encontrada em diferentes contextos. A partir disso, os alunos poderiam assimilar estratégias e métodos de tradução empregados por profissionais e, portanto, adquirir habilidades tradutórias. De acordo com Zanettin (1998, p. 618-621 *apud* HU, 2016, p.181), a função do uso de *corpus* comparável na formação de tradutores segue três princípios: oferecer estudantes assistência em tradução, suprir esses alunos com informação útil de terminologia e conteúdo, e explorar similaridades entre diferentes línguas.

Os Estudos de Interpretação Baseados em *Corpus* (doravante, EICB) começaram nos anos 50. A pesquisa em EIBC segue a psicologia experimental, a teoria interpretativa da tradução como proposta pela Escola de Paris, a neurolinguística, a neuropsicologia, entre outras

---

áreas. Esse campo explora o papel do intérprete e a atividade de interpretação, além de serem caracterizados como estudos qualitativos baseados em análises de pequenas quantidades de dados. Os EIBC ainda estão em crescimento e precisam estabelecer maior objetividade e cientificidade. Apesar disso, esse objeto tem importância teórica e já foram criados *corpora* de interpretação como o K6 (MEYER, 2008), o CIAIR *corpus* (TOHYAMA, 2006), o EPIC (BENDAZZOLI; SANDRELLI, 2005), ou o CECIC (HU; TAO, 2010). Os EIBC podem ser organizados em quatro períodos nomeados pré-pesquisa, psicologia experimental, pesquisa em teoria interpretativa da tradução e pesquisa interdisciplinar. Entretanto, umas das dificuldades encontradas na área é a característica da interpretação ser, geralmente, mantida secreta, visto que lida com questões de privacidade ou confidencialidade. O *corpus* de interpretação é encontrado na forma de vídeo ou áudio, o que pode ser considerado outro obstáculo, posto que a transcrição é necessária e será feita manualmente pelo pesquisador ou por sua equipe. Esses fatores tornam a compilação de *corpora* de interpretação mais desafiadora do que a coleta de *corpora* de traduções em texto escrito. Ademais, para conduzir uma pesquisa na área é necessário conhecimento sobre os diferentes tipos de serviços e modos de interpretação. O primeiro cobre as possibilidades de papéis do intérprete, bem como de participantes da interpretação. O segundo faz referência à interpretação simultânea, consecutiva ou *liaison*. Ainda que tenha sido feito progresso nos EIBC, ainda são necessários maiores esforços para compilação de *corpora* tencionando desenvolver e expandir o escopo da área.

309

Na conclusão do livro, Hu apresenta inicialmente as implicações dos ETBC e são feitas análises dos problemas que podem ser encontrados na área. A primeira adversidade a ser tratada é a limitação tecnológica. Mesmo com o grande número de *corpora* já compilado, a coleta ainda está longe de ser suficiente para os ETBC. Sendo assim, as pesquisas podem ter sua validade afetada por esse motivo. Além disso, o *corpus* pode ter sua serventia questionada por suprir apenas contextualização linguística, e não social. Outra restrição é encontrada em relação ao intervalo de tempo dos *corpora* e ao número de pares linguísticos. O pesquisador pode não conseguir coletar dados referentes às épocas mais remotas da tradução ou à correspondência linguística. Não obstante, o uso de *corpus* não é útil para análise do discurso, visto que a concordância analisada via investigação de *corpora* não leva em consideração os níveis de coesão e coerência textual. Dentre as limitações metodológicas, o autor cita o impedimento dos estudos baseados em *corpus* em explicar a natureza da tradução, visto que uma grande quantidade de dados é fornecida, mas não é suficiente para explicar motivações para normas ou mesmo para pesquisas. Em seguida, o autor disserta sobre as perspectivas dos ETBC mesmo

---

com os entraves apresentados. Uma das expectativas é a de crescimento no tamanho dos *corpora* disponíveis em conjunto com a divulgação desse material. Outra tenciona a possibilidade de condução de pesquisas interdisciplinares, bem como de incorporação de análises qualitativas e não só quantitativas. Também está presente o anseio de maior envolvimento em trabalhos na área de EIBC, de tradução de gêneros pragmáticos e de formação de tradutores.

*Corpus-based Translation Studies* discorre sobre temas importantes para a evolução dos Estudos da Tradução Baseados em *Corpus*. A área é apresentada como um campo promissor, apesar de ainda estar em seu estágio inicial de desenvolvimento. Suas referências metodológicas e suas práticas são validadas por meio da pesquisa empirista que os ETBC propõem, além de facilitarem a difusão de investigações com abordagens tanto qualitativas como quantitativas para o campo em questão. Através da leitura do livro é possível entender em detalhes as possibilidades de aplicação dos estudos de *corpus* nos EdT. Ademais, a edição confirma os ETBC como um ramo plenamente estabelecido e com grande possibilidade de crescimento.

310

## REFERÊNCIAS

HU, Kaibao. **Introducing Corpus-based Translation Studies**. *New Frontiers in Translation Studies*, 2016, 258 p.

---

\* Letícia Freitas de ASSIS – Mestranda em Estudos da Tradução na Universidade Federal do Ceará (UFC) e licenciada em Letras Português e Inglês (2018) pela mesma instituição. Fortaleza, Ceará, Brasil.  
Currículo acadêmico: <http://lattes.cnpq.br/3401358815887054>  
ORCID: <https://orcid.org/0000-0002-6958-3844>  
E-mail: [leticiaassis23@gmail.com](mailto:leticiaassis23@gmail.com)

<sup>1</sup> “Translated text record genuine communicative events in any language” (BAKER, 1993, p. 248)

<sup>2</sup> “Typical cross-lingual features with certain universality in target texts which are different from those of source texts” (KE, 2005, p. 303-307)

<sup>3</sup> “Translational behavior is actually relative to a particular translated language, so features of translated language can be perceived as distinctive features of translation in relation to non-translated texts in the same language” (HU XIANYAO, 2008 *apud* HU, 2018, p. 147)