O XBENCH COMO UM RECURSO PARA GARANTIR A CONVENCIONALIDADE NA TRADUÇÃO DE COLOCAÇÕES

Xbench as a Resource to Ensure Conventionality in the Translation of Collocations

DOI: 10.14393/LL63-v41S-2026-04

Patrícia Helena Freitag*

RESUMO: Este trabalho se insere nos Estudos da Tradução Baseados em Corpus. Atualmente, tradutores de áreas técnicas utilizam programas de controle de qualidade, como o Xbench (ApSIC, 2020), para identificar, de forma automatizada, problemas como inconsistências. Este artigo descreve a elaboração de uma checklist do Xbench para ajudar a garantir a convencionalidade na tradução de colocações em artigos de blogs de coworking. Partiu-se de um quadro que registra quebras de convencionalidade na tradução de colocações (Freitag, 2025). Demonstrou-se como elaborar regras utilizando expressões regulares. Um exemplo é o caso de *office space*, que aparece como *espaço de escritório* frequentemente no corpus traduzido, ao passo que o corpus autêntico revela uso simplesmente de *escritório*. Explicou-se que, no Xbench, "espaços? de escritório" detecta as traduções *espaço de escritório*, *espaços de escritório*, *espaço de escritórios* e *espaços de escritórios*, automatizando a identificação dessas quebras de convencionalidade. A metodologia apresentada aqui pode ser usada para criar checklists de outros gêneros textuais.

PALAVRAS-CHAVE: Tradução. Convencionalidade. Colocações. Xbench. Controle de qualidade.

ABSTRACT: This study is situated within Corpus-Based Translation Studies. Nowadays, translators working in technical fields use quality assurance tools, such as Xbench (ApSIC, 2020), to automatically identify issues like inconsistencies. This paper describes the creation of an Xbench checklist designed to ensure the use of conventional collocations in the translation of articles published in coworking blogs. The checklist was based on a table with deviations from conventions in the translation of collocations (Freitag, 2025). We show how to create rules using regular expressions. One example is *office space*, which frequently appears as *espaço de escritório* in the translated corpus, even though the corpus of authentic texts reveals the use simply of *escritório*. We explain that, in Xbench, "espaços? de escritório" can detect *espaço de escritório*, *espaços de escritório*, *espaço de escritórios* e *espaços de escritórios* in translations, automatizing the identification of these deviations from conventions. The methodology presented here can be used to create checklists for other text genres.

KEYWORDS: Translation. Conventionality. Collocations. Xbench. Quality Assurance.

* Doutora em Letras. Professora Assistente do curso de Licenciatura em Letras da Universidade Estadual do Ceará. ORCID: 0000-0002-2446-3718. E-mail: patricia.freitag(AT)uece.br.

1 Introdução

Atualmente busca-se, na maioria das profissões, a automatização de processos para aumentar a produtividade e qualidade dos serviços ou produtos entregues. No caso da tradução, isso não poderia ser diferente. Tal mentalidade já estava presente na década de 1990, quando os programas de auxílio à tradução começaram a fazer parte da rotina de tradutores (Stupiello; Esqueda, 2017), em substituição a traduções realizadas em documentos de texto simples. Novo momento decisivo foi o da popularização da tradução automática, quando os tradutores passaram, aos poucos, a se ocupar mais de pós-edição do que de tradução propriamente dita. Hoje, com a combinação desses elementos, somados à inteligência artificial e aos prazos cada vez mais curtos, é essencial que se saiba utilizar bem o tempo.

Na literatura relevante, há trabalhos que abordam o uso do Xbench (ApSIC, 2020). Trata-se de uma ferramenta específica de controle de qualidade, popular no setor, em que é possível carregar um arquivo bilíngue após a tradução (geralmente no formato XLIFF) e executar verificações de qualidade. Os trabalhos encontrados que mencionam essa ferramenta em projetos de tradução não se aprofundam no potencial das checklists (listas de verificação), m dos recursos do programa. É o caso de Streckwall (2023), que apresenta as funções padrão do Xbench e menciona a existência das checklists, mas sem explorar suas possibilidades. O mesmo ocorre em Preto (2021), que traz um relatório de estágio em tradução em uma empresa na área. Em tal relatório, Preto (2021) explica que o uso do Xbench é obrigatório em muitos projetos, mas não menciona o uso de checklists. Esses exemplos reforçam a ideia de que o Xbench é uma ferramenta de controle de qualidade importante e prevalente no setor de tradução, mas que acaba sendo usada principalmente de forma padronizada, sem que tradutores e empresas aproveitem seu potencial de auxiliar no controle de elementos personalizados.

Nesse contexto, defendo que é de suma importância o desenvolvimento de pesquisas acadêmicas que investiguem formas de contribuir para a automatização do trabalho do tradutor, sem reduzir seu protagonismo de especialista. Assim, no presente trabalho, descrevo o desenvolvimento de um recurso que se propõe a automatizar, em alguma medida, a identificação de quebras de convencionalidade na tradução de colocações no gênero textual

artigos de blogs de coworking. Trata-se de uma checklist do Xbench[1] (ApSIC, 2020). Ao mesmo tempo que o recurso visa à automatização do controle de qualidade, sua criação exige conhecimento especializado por parte do tradutor, tanto da área da Linguística de Corpus quanto do uso de expressões regulares, explicadas na próxima seção.

2 Pressupostos teóricos

Nos próximos parágrafos, serão cobertos os aspectos de fundamentação teórica mais relevantes para a pesquisa, quais sejam: Estudos da Tradução Baseados em Corpora, colocações e qualidade na tradução. A origem dos Estudos da Tradução Baseados em Corpora é marcada pelos trabalhos de Baker no início da década de 1990. Em seu artigo de 1993, a estudiosa afirma que os corpora permitiriam observar características de textos traduzidos e os princípios que os governam. Considerando que a linguística de corpus permite a observação de padrões em textos traduzidos (Baker, 1993), a presente pesquisa utiliza corpora para identificar possíveis influências na tradução de colocações do inglês para o português. Para isso, o foco recai sobre os padrões colocacionais que desviam dos usos típicos em textos autênticos, ou seja, textos não traduzidos. Os desvios podem ocorrer devido a, por exemplo, combinações não consagradas pelo uso, como *engano crasso* em vez de *erro crasso*, ou frequência mais alta ou mais baixa do que o esperado pela comunidade ou gênero textual.

Neste trabalho, o ponto de partida é um quadro resultante de uma análise prévia de colocações extraídas de artigos de blogs de coworking traduzidos em comparação com artigos do mesmo gênero escritos originalmente em português (Freitag, 2025). Tal quadro contém recomendações para a tradução e colocações. Entendo colocação como "combinação lexical consagrada de duas ou mais palavras de conteúdo" (Tagnin, 2013). Nessa perspectiva, convém retomar o princípio idiomático de Sinclair (1991), segundo o qual temos, na linguagem, expressões já estruturadas que se apresentam aos falantes como escolhas individuais. Tomemos como exemplo o substantivo *custos* e a intenção de dizer que o cliente deverá assumir o pagamento de determinados valores. Nesse exemplo, há alguns verbos que poderiam ser usados, como *assumir os custos* e *responsabilizar-se pelos custos*, mas *arcar com*

[1] Versão 3.0.0, Build 1520.

os custos é a forma mais comum, consagrada pelo uso. Assim, segundo o princípio idiomático, o falante tem à disposição a combinação *arcar com os custos* já estruturada para uso, não precisando escolher palavra por palavra.

Foram abordados quatro tipos de colocações com base no sistema proposto por Tagnin (2013): nominais, adjetivas, verbais e adverbiais. Além disso, também consideraram-se colocações que contêm uma letra isolada (ex.: *geração Z*) ou duas palavras estrangeiras (ex.: *home office*). Freitag (2025) e Freitag e Rebechi (2025) investigam possíveis motivos para as quebras de convencionalidade nas traduções e fazem recomendações com base no corpus de textos autênticos. Convém ressaltar que, em grande parte, as recomendações envolvem o uso de colocações no texto-alvo, mas nem sempre. Isso porque, em alguns casos, os dados do corpus de textos autênticos sugerem o uso de uma só palavra. É o caso de *espaço de escritório* (combinação usada para traduzir *office space*) que, apesar de ser formada por duas palavras lexicais, não foi a forma encontrada para expressar esse conceito nos textos autênticos, sendo o uso simplesmente de *escritório* mais típico nesses textos.

A convencionalidade é, portanto, um norteador no presente trabalho. Sua relevância é reconhecida não só em pesquisas acadêmicas sobre padrões colocacionais, mas também no setor da tradução técnica, onde é um dos elementos observados ao se avaliar a qualidade de textos traduzidos, como demonstra Oliveira (2019). A preocupação com a qualidade se intensificou no setor da tradução técnica nos últimos anos devido a, entre outros fatores, o aumento no volume de textos a serem traduzidos e a redução nos prazos, o que exige uma produtividade maior de tradutores e leva à divisão de trabalhos longos em uma equipe de tradutores (Drugan, 2013).

Convém, então, esclarecer o que se entende por qualidade. É preciso considerar as demandas do setor: para a maioria dos clientes, uma tradução de qualidade é aquela que não contém erros. Nesse setor, as traduções costumam ser verificadas por avaliadores humanos, que buscam erros previstos na tipologia de erros adotada pela empresa. Um exemplo é o MQM Core[2], que traz categorias de erros como Accuracy (que engloba, entre outras questões, omissão ou inclusão de informações em relação ao texto-fonte), Linguistic conventions (que abrange pontuação, gramática e outros erros) e Terminology (que cobre não conformidade

[2] https://themqm.org/

com o glossário fornecido e outros erros de terminologia). Além dessas categorias mais objetivas, o modelo também conta com a categoria Style, que abrange, entre outros, construções que não soam naturais. Nota-se que o conceito de erro extrapola o que um leigo poderia esperar, incluindo questões que podem ser consideradas um tanto subjetivas (como combinações atípicas de palavras), mas que podem ser observadas utilizando-se ferramentas de linguística de corpus que permitem identificar padrões de língua geral ou de determinado gênero, comunidade ou registro.

Esse contexto de expectativa de alta produtividade e qualidade revela a importância de se ter ferramentas de controle de qualidade eficientes, que deem conta não só de problemas já bastante conhecidos do setor (como inconsistências entre segmentos, segmentos não traduzidos e não conformidade com os glossários) como também de questões específicas identificadas por recursos personalizáveis, como as checklists do Xbench.

Apesar de os recursos-padrão do Xbench serem bastante utilizados por tradutores (e, inclusive, exigidos por muitas empresas de tradução), o recurso de checklist, que permite personalização, é pouco utilizado[3]. É provável que isso ocorra porque essas checklists precisam ser construídas e exigem o domínio de expressões regulares. As expressões regulares, ou RegEx, são padrões formados por caracteres literais (ou seja, que devem ser entendidos como tal pelo programa) e caracteres especiais (ou seja, que representam uma função) (Moreira Filho, 2021). A sintaxe usada no Xbench está disponível no guia do usuário online da ferramenta[4]. Nas checklists do Xbench, por exemplo, poder-se-ia elaborar a expressão regular *salas?,* sendo que *s*, *a*, *l*, *a* e *s* são caracteres literais e *?* é um caractere especial cuja função é buscar ocorrências em que o caractere imediatamente anterior aparece zero ou uma vez. Assim, essa expressão de exemplo identificaria *sala* e *salas*, mas não *salass*.

Uma busca no Google Scholar por *Xbench checklist* retorna apenas 34 resultados. Em alguns dos trabalhos, observa-se que, quando utilizadas, as checklists do Xbench costumam abarcar preferências do cliente, como aspas retas ou curvas (Pinto, 2020). Em outros, relata-se o uso de checklists fornecidas pelo cliente, sem especificar o que elas contêm (Pereira, 2025). Há, ainda, trabalhos que indicam que se pode criar checklists pessoais com possíveis erros,

[3] Com base em minha observação como tradutora no período de 2012 a 2025.
[4] https://docs.xbench.net/user-guide/regular-expressions/#microsoft-word-wildcards-syntax

como quando se tem o advérbio de negação *not* no texto-fonte em inglês, mas não se tem um advérbio de negação no texto-alvo (Streckwall, 2023). Neste trabalho, além de dar mais destaque para as checklists, proponho um uso inovador: em vez de partir de erros cometidos anteriormente ou de feedback de clientes, proponho a construção de uma checklist com foco em convencionalidade. Não encontrei, na literatura disponível, outro trabalho que utilize a linguística de corpus para extrair e analisar colocações de corpora paralelo e, em seguida, utilizar os achados para elaborar uma checklist do Xbench com foco em convencionalidade. Tal recurso pretende auxiliar em traduções que respeitem os padrões colocacionais de textos autênticos do gênero e, ao mesmo tempo, evitar alguns erros de estilo que poderiam ser apontados por avaliadores. A seguir, apresento a metodologia utilizada.

3 Metodologia

Esta seção descreve brevemente os procedimentos de coleta dos corpora, de extração de n-gramas para a identificação de colocações e da análise desses itens (Freitag, 2025). Logo em seguida, os procedimentos de elaboração da checklist do Xbench, que são o cerne deste trabalho, são apresentados em detalhes.

3.1 Procedimentos de coleta

Foram compilados dois corpora: um corpus monolíngue de artigos de blogs de coworking escritos originalmente em português (PTBR-AUT, com 396.296 tokens), e um corpus bilíngue paralelo composto por textos do mesmo gênero escritos originalmente em inglês e suas respectivas traduções (ENG-AUT e PTBR-TRAD, com 370.734 e 390.788 tokens, respectivamente). A coleta se deu de forma manual, acessando os blogs e copiando os textos para documentos de texto. Para o corpus paralelo, foi feito o alinhamento das traduções usando o LF Aligner (Farkas, 2019). Os corpora foram armazenados no Sketch Engine (Kilgarriff *et al*., 2014).

3.2 Procedimentos de extração de dados e análise

Após a compilação dos corpora, foram extraídos n-gramas-chave lematizados com extensão entre 2 e 4 palavras (para abarcar colocações que porventura aparecessem com

alguma palavra entre a base e o colocado), com no mínimo 5 ocorrências, para o PTBR-AUT e o PTBR-TRAD. Foram 18.975 n-gramas-chave no PTBR-AUT e 16.110 no PTBR-TRAD. As listas de n-gramas-chave foram exportadas para MS Excel (incluindo dados como frequência) e reorganizadas em três planilhas, quais sejam: n-gramas-chave presentes nos dois corpora, n-gramas-chave presentes só no PTBR-TRAD e n-gramas-chave presentes só no PTBR-AUT. Realizou-se a leitura linha a linha, excluindo os n-gramas que não se enquadravam nas estruturas previstas de colocações, por exemplo, *de coworking* e *um espaço de trabalho*, que se iniciam por preposição e artigo indefinido, respectivamente. Como resultado, restaram 1.465 n-gramas-chave no PTBR-AUT e 1.175 no PTBR-TRAD. Nesse momento, também foi feita a classificação dos itens mantidos, podendo ser colocação nominal, adjetiva, verbal, adverbial ou outro (sendo que essa última classe abarca colocações que contêm uma letra isolada ou duas palavras estrangeiras).

Em seguida, as planilhas foram reorganizadas para permitir a análise de cada categoria. O Quadro 1 apresenta a quantidade de n-gramas-chave em cada caso.

Quadro 1 – Quantidade final de n-gramas-chave.

	Adjetivas	Nominais	Adverbiais	Verbais	Outras
Ambos os corpora	161	73	55	96	4
PTBR-AUT	434	318	85	226	13
PTBR-TRAD	330	238	65	144	9

Fonte: elaborado pela autora.

Procedeu-se à análise qualitativa dos dez itens de maior chavicidade em cada caso, por exemplo, colocações nominais presentes nos dois corpora, só no PTBR-AUT e só no PTBR-TRAD. No total, foram analisadas 143 colocações (Freitag, 2025). Nessa análise, contrastaram-se as combinações de palavras encontradas em português autêntico e traduzido. No caso de combinações iguais nos corpora, analisou-se o entorno para saber se os contextos de uso eram também semelhantes. No caso de combinações diferentes, buscou-se entender a motivação para as combinações usadas no PTBR-TRAD, inclusive possível influência da língua-fonte ou do texto-fonte.

Após a análise, os itens passíveis de comporem regras do Xbench e que foram identificados como possíveis pontos de dificuldade para tradutores foram registrados em um quadro com quatro colunas: PTBR-AUT, PTBR-TRAD, ENG-AUT e Comentário. Em PTBR-AUT, foram registradas as colocações encontradas no corpus de textos escritos originalmente em português; em PTBR-TRAD, as colocações correspondentes encontradas no corpus de textos traduzidos; em ENG-AUT, o texto em inglês que deu origem à tradução; e, por fim, em Comentário, anotações relevantes para a construção da regra, como contextos de uso e recomendações de traduções mais adequadas do que a *prima facie*. O quadro é composto por 61 itens (Freitag, 2025).

Em alguns casos, nem todas essas informações são relevantes, então algumas células do quadro foram deixadas em branco. É o caso de *sala de reunião*, colocação encontrada com frequência no PTBR-AUT, e *sala de reunião privativa*, encontrada com frequência apenas no PTBR-TRAD. Percebe-se que os textos desse gênero em português autêntico não costumam utilizar o adjetivo *privativa* nesse contexto, possivelmente porque essa característica já é inerente às salas de reunião do ponto de vista dessa comunidade linguística. Nesse caso, o texto na porção em inglês do corpus bilíngue não é relevante para a criação da regra. Assim, apenas as colunas PTBR-AUT, PTBR-TRAD e Comentário foram preenchidas. O Quadro 2 apresenta alguns dos itens.

Quadro 2 – Itens para a elaboração das regras do Xbench.

PTBR-AUT	PTBR-TRAD	ENG-AUT	Comentário
escritório, espaço de escritório	espaço de escritório	office space	Considerar o uso de *escritório* simplesmente.
sala de reunião	sala de reunião privativa	—	Não é necessário utilizar *privativa* (subentendido).
fazer home office, trabalhar em home office	trabalho em casa, trabalhar de/em casa	work from home	Em português autêntico, *home office* pode se referir ao espaço físico, mas é usado principalmente no sentido de atividade.
exercer impacto, gerar impacto, causar impacto	causar impacto	make an impact, have an impact, cause an impact e impact (verbo).	Prosódia semântica neutra: exercer impacto. Positiva: gerar impacto. Negativa: causar impacto.
Geração Z (menos frequente), geração Z (mais frequente)	Geração Z (mais frequente), geração Z (menos frequente)	Gen Z, Generation Z	Priorizar escrever geração com g minúsculo.

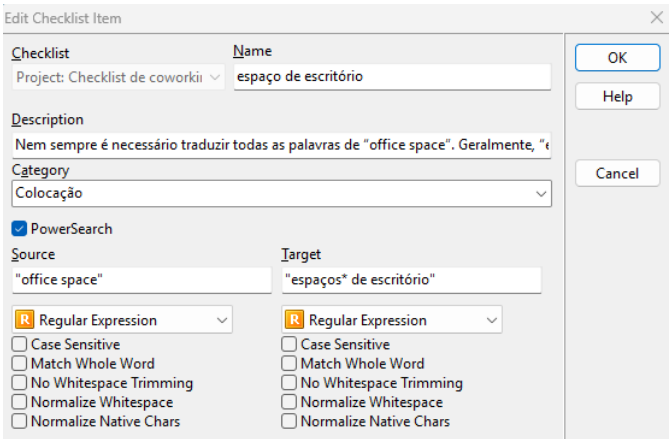
Fonte: modificado de Freitag (2025).

3.3 Procedimentos de construção da checklist

Com base no quadro, procedeu-se à criação da lista do Xbench diretamente no programa. Para isso, foi necessário abrir o Xbench, clicar em *View* na faixa de opções superior e, então, em *Checklist manager* no menu suspenso que se abre. Em seguida, clicar no ícone *New Checklist* na faixa de opções, nomear a checklist, clicar em OK e selecionar o local onde o arquivo deve ser salvo. O nome escolhido foi *Checklist de coworking*.

Com o arquivo aberto, foram criadas as regras individualmente. Para cada regra, foram preenchidos (quando relevante) os seguintes campos: *Name*, *Description*, *Category*, *PowerSearch*, *Source*, *Config source*, *Target*, *Config target*. A Figura 1 exemplifica o preenchimento desses campos, que são explicados logo em seguida.

Figura 1 – Exemplo de preenchimento dos campos.



Fonte: captura de tela do programa ApSIC Xbench.

O campo *Name* foi preenchido com um identificador curto, que pode ser a combinação a ser evitada, o texto em inglês com o qual ter atenção ou a opção recomendada com base no PTBR-AUT, a depender de cada caso. *Description* recebeu a explicação mais longa e detalhada, que permite ao tradutor entender qual é a recomendação e as razões para isso. *Category* foi preenchido com uma classificação da regra, conforme a necessidade. A maioria foi classificada como *Colocação*, mas alguns casos ficaram como *Cultura*, *Estrangeirismo* ou *Palavra isolada*. *PowerSearch* é um campo que pode ser marcado para fazer buscas que: contenham mais de uma palavra, em qualquer ordem; contenham mais de uma palavra, na ordem dada (sendo necessário usar aspas); não contenham determinado texto (sendo necessário adicionar um hífen antes do texto); ou contenham uma *ou* outra palavra (sendo necessário adicionar *or* entre as palavras). *Source* foi preenchido com o texto a ser buscado no segmento em inglês do arquivo traduzido que passará pelo controle de qualidade, e *Target* com o texto a ser buscado na respectiva tradução em português. *Config source* e *Config target* são as nomenclaturas que usamos para o conjunto de configurações do texto-fonte e do texto-alvo.

A criação das regras partiu dos itens no quadro previamente explicado nesta seção e apoiou-se no guia do usuário online do Xbench[5], especificamente no tópico sobre uso de expressões regulares. Conforme explicado anteriormente, as expressões regulares, ou RegEx, permitem usar caracteres especiais para representar funções. Por exemplo, o asterisco é usado

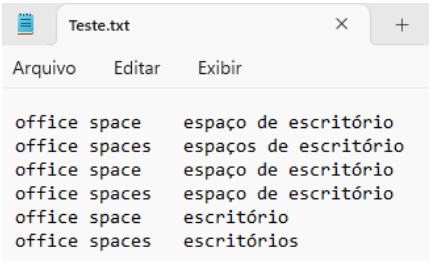
[5] https://docs.xbench.net/

para identificar palavras com zero ou mais ocorrências do caractere imediatamente anterior. Assim, a expressão *travel*ing* encontraria tanto ocorrências de *traveling* quanto de *travelling*.

Após a escrita de cada regra, foi feita sua testagem. Para isso, criou-se um documento em formato .txt nomeado como *Teste*, que foi carregado no Xbench como *Ongoing Translation*, ou seja, documento traduzido a ser verificado. O conteúdo desse arquivo foi alterado para a testagem de cada regra individualmente. Com ele, simularam-se possíveis textos-fonte e textos-alvo em traduções, inserindo, a cada linha, um possível texto-fonte, seguido de um Tab e, então, do respectivo texto-alvo. É preciso incluir todas as variações consideradas pontos de atenção, para confirmar que a regra esteja identificando todos os casos necessários.

A regra na Figura 1, por exemplo, deve encontrar todos os casos que contenham *office space* ou *office spaces* no texto-fonte, além de todos os casos em que esse trecho foi traduzido como *espaço de escritório*, *espaços de escritório*, *espaços de escritórios* ou *espaços de escritórios*. Além disso, o arquivo .txt de teste também precisa conter opções tradutórias adequadas (no caso, as palavras *escritório* e *escritórios* isoladamente). Isso porque, no momento do teste, busca-se verificar se a regra está identificando apenas possíveis traduções problemáticas, e não falsos-positivos, ou seja, traduções adequadas. Assim, o arquivo .txt deste exemplo foi preenchido da forma indicada na Figura 2.

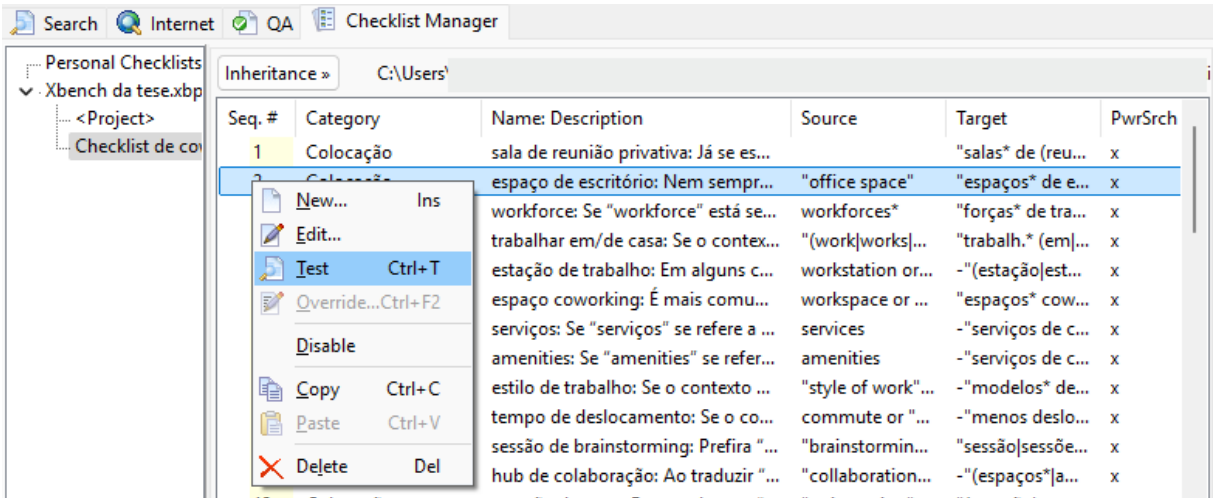
Figura 2 – Exemplo de preenchimento do .txt de teste de regras.



Fonte: captura de tela do programa Bloco de notas.

Organizado o .txt que simula traduções específicas de interesse para a regra, foi rodada a testagem da regra em si. Para isso, na tela *Checklist Manager* do Xbench, é necessário clicar com o botão direito sobre a regra e, então, na opção *Test* no menu suspenso. A Figura 3 mostra essa opção.

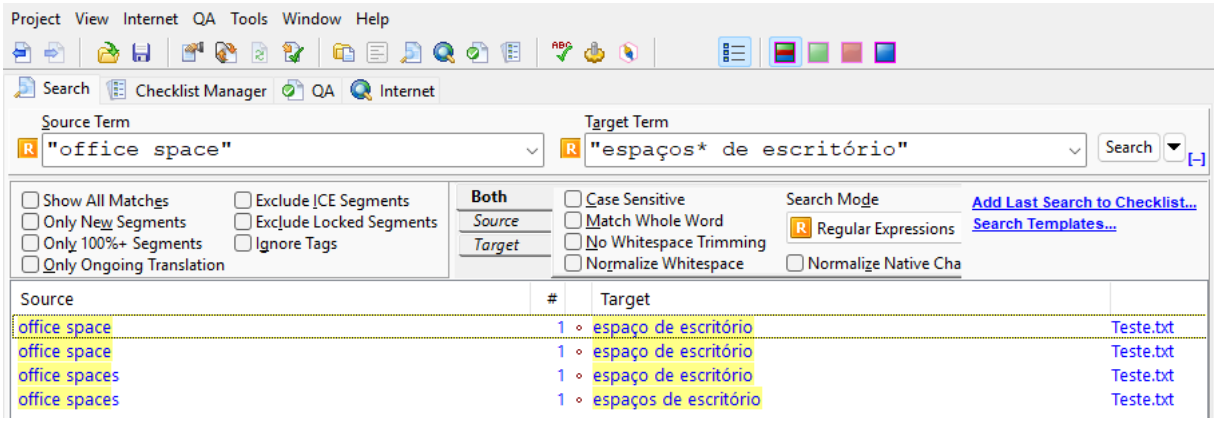
Figura 3 – Passo a passo para testar uma regra.



Fonte: captura de tela do programa ApSIC Xbench.

A Figura 4 mostra o resultado do teste da regra do exemplo. Apenas os segmentos com traduções possivelmente inadequadas (ou seja, contendo *espaço de escritório* ou *espaços de escritório* no texto-alvo) foram identificadas. Os segmentos com traduções julgadas adequadas (as palavras *escritório* ou *escritórios* isoladamente) não foram identificados. Isso demonstra que a regra funciona como o esperado.

Figura 4 – Exemplo de resultado de teste de regra individual.



Fonte: captura de tela do programa ApSIC Xbench.

Nos casos em que uma dada regra não funcionou como esperado, foram feitas alterações nas configurações e expressões de busca no texto-fonte e texto-alvo da regra e repetido o teste, até que se chegasse ao resultado desejado. Esse processo é recomendável

como prova real mesmo para pesquisadores que dominam o uso de expressões regulares, mas principalmente para pesquisadores que ainda estão adquirindo esse conhecimento, como forma de garantir o correto funcionamento e evitar erros.

Após a elaboração de todas as regras, uma última etapa foi o teste de aplicação da checklist inteira usando os arquivos alinhados do corpus. Para isso, foram carregados no Xbench os arquivos do corpus paralelo alinhados em .tmx (um dos formatos gerados pelo LF Aligner). Apesar de ser um formato de memória de tradução, pode ser carregado como *Ongoing translation* para passar pelo controle de qualidade do programa. Em seguida, foi rodado o controle de qualidade do Xbench aplicando apenas a checklist, e não os recursos-padrão da ferramenta.

4 Resultados

Os próximos parágrafos apresentam a maneira como foi desenvolvida a regra referente a cada item do Quadro 2. A combinação de palavras *espaço de escritório* ocorre significativamente mais no corpus de textos traduzidos (freq. 175) do que no corpus de textos autênticos (freq. 17) (LL +153,52). Isso se dá, possivelmente, devido à influência da colocação do texto-fonte, *office space*. Ainda que, nos textos em inglês, o convencional seja o uso de uma colocação, observou-se que, no PTBR-AUT, costuma-se usar simplesmente *escritório*. Esse caso demonstra que é preciso ter conhecimento de como as colocações funcionam, mas mantendo em mente que, por se fundamentar no conceito de convencionalidade, os padrões típicos em cada idioma podem ser distintos: o que se expressa como colocação em inglês pode ser expresso como uma palavra isolada em português, e vice-versa. Assim, essa regra devia identificar segmentos que contivessem, no texto-fonte, *office space* ou *office spaces*, e, ao mesmo tempo, no texto-alvo, *espaço de escritório*, *espaços de escritório*, *espaço de escritórios* ou *espaços de escritórios*.

Para isso, em *Source*, usou-se "office space". As aspas garantem a busca da expressão exatamente na ordem e com os elementos dados. Ou seja, não identifica ocorrências apenas de *office* ou de *space*, nem a ordem *space office*, nem sequências com outras palavras no meio, como *office or desk space*. No entanto, as aspas não impedem a identificação de outras letras antes ou depois do texto, ou seja, também capta *office spaces*, com *s* ao final. A configuração

de busca de *Source* foi marcada como *Simple*, pois não usa nenhum caractere especial que configure uma expressão regular.

Em *Target*, usou-se "espaços? de escritório". Novamente, as aspas limitam o texto e a ordem de busca. O ponto de interrogação em *espaços?* é usado para buscar ocorrências dessa palavra que contenham zero ou uma ocorrência da letra *s*. Em *escritório* isso não é necessário, pois se trata do fim da expressão de busca, portanto o programa identifica todas as ocorrências da expressão, sejam elas seguidas por outros caracteres ou não. No entanto, se desejado, é possível utilizar a mesma estratégia do ponto de interrogação sem prejuízo. A configuração de busca de *Target* foi marcada como *Regular Expression*, para que o programa entenda o asterisco como uma função de busca, e não como texto literal.

Nesta regra, a caixa *PowerSearch* foi marcada para indicar que todo o texto entre aspas deve ser buscado na sequência dada, incluindo os espaços em branco. Uma observação importante é que as aspas nos campos *Source* ou *Target* precisam ser retas para serem compreendidas como caractere delimitador na busca com PowerSearch. O Quadro 3 mostra as informações inseridas para esta regra na checklist do Xbench.

Quadro 3 – Regra *espaço de escritório*.

Name	espaço de escritório
Description	Nem sempre é necessário traduzir todas as palavras de “office space”. Geralmente, “escritório” basta.
Category	Colocação
PowerSearch	Sim
Source	"office space"
Config source	Simple
Target	"espaços? de escritório"
Config target	Regular Expression

Fonte: elaborado pela autora.

O próximo item diz respeito ao uso do adjetivo *privativa* logo após *sala de reunião*. Essa adjetivação ocorre apenas no corpus PTBR-TRAD (freq. 6), possivelmente devido à presença do

cognato *private* no texto-fonte. Supõe-se que essa ausência no corpus PTBR-AUT ocorra porque os brasileiros já esperam que as salas de reunião sejam espaços mais reservados.

Além da combinação *sala de reunião privativa* no singular, o corpus de textos traduzidos também contém ocorrências com a marcação de plural em diferentes pontos — *salas de reunião privativas* e *sala de reuniões privativa* —, por isso, a regra do Xbench referente a essa questão precisa ser capaz de identificar essas variações no texto-alvo. Chegou-se à expressão regular "salas? de (reunião|reuniões) privativas?". Como nos casos anteriores, as aspas são necessárias para localizar essa sequência de palavras, nesta ordem, com espaço entre os elementos. Também como no caso anterior, o ponto de interrogação antes do *s* é usado para buscar ocorrências da palavra que contenham zero ou uma ocorrência da letra imediatamente anterior, ou seja, *sala* e *salas* e *privativa* e *privativas* no caso desta regra.

Na língua portuguesa, nem sempre o plural se dá pela simples adição da letra *s* no fim da palavra. É o caso de *reunião*, cujo final *ão* é substituído por *ões*. Assim, o uso do ponto de interrogação não provocaria o efeito desejado. Para dar conta dessa questão, optou-se por indicar as duas formas da palavra — singular e plural — separadas por uma barra vertical, sem espaços. Esse caractere especial funciona como o operador OR em outros sistemas, ou seja, significa que a regra deve necessariamente encontrar *reunião* ou *reuniões*. Os parênteses são necessários para indicar prioridade, assim como em operações matemáticas. Caso eles não fossem usados, o resultado esperado não seria atingido, pois a regra buscaria segmentos que contivessem "*salas? de reunião"* ou "*reuniões privativas?*", ou seja, toda a porção antes da barra vertical e toda a porção depois da barra vertical. A configuração de busca de *Target* foi marcada como *Regular Expression*, para que o programa entenda o ponto de interrogação, a barra vertical e os parênteses como funções de busca, e não como texto.

Nesta regra, optou-se por não especificar o texto-fonte, pois não era relevante. Além disso, assim como na regra anterior, a caixa *PowerSearch* foi marcada para indicar que todo o texto entre aspas deveria ser buscado. O Quadro 4 mostra as informações inseridas para esta regra na checklist do Xbench.

Quadro 4 – Regra *sala de reunião privativa*.

Name	sala de reunião privativa
Description	Já se espera que as salas de reunião sejam privativas (pois não se deseja realizar uma reunião em ambientes aos quais outras pessoas tenham acesso no mesmo momento). Dessa forma, não é necessário utilizar esse adjetivo, pois é uma propriedade subentendida de salas de reunião.
Category	Colocação
PowerSearch	Sim
Source	—
Config source	—
Target	"salas? de (reunião|reuniões) privativas?"
Config target	Regular Expression

Fonte: elaborado pela autora.

O próximo item, *trabalhar em/de casa*, foi criado de modo a alertar para a tradução palavra por palavra de *work from home*. A colocação *trabalhar em casa* é mais frequente no corpus PTBR-TRAD (freq. 143) do que no PTBR-AUT (freq. 92), sendo essa diferença estatisticamente significativa (LL +11,88). Aparentemente, esse sobreuso é resultado de influência do texto-fonte. Como alternativa, para que os textos traduzidos mantenham um padrão colocacional parecido com o do PTBR-AUT, pode-se usar opções que envolvam o estrangeirismo *home office*, mesmo que ele não esteja presente no texto-fonte. Essa recomendação é respaldada pela maior frequência de construções como *fazer home office* nos textos escritos originalmente em português em comparação com os textos traduzidos.

A regra da checklist, portanto, deve ser capaz de identificar segmentos que contenham, na tradução, *trabalho em casa*, *trabalhar em casa* (ou conjugações de *trabalhar*), ou *trabalhar de casa* (ou conjugações de *trabalhar*) e, ao mesmo tempo, no texto-fonte, *work from home*, *working from home* ou *work-from-home*, que são todas formas que ocorrem no ENG-AUT. Considerando essas necessidades, elaborou-se, para o campo *Source*, a expressão regular "(work|works|worked|working) from home" or "work-from-home". Na regra anterior, havia apenas uma palavra que poderia variar no texto-alvo — reunião/reuniões —, por isso, usou-se

a barra vertical para separar as opções e os parênteses para delimitar essas opções. Porém, na presente regra, as opções envolvem espaços ou hifens, não sendo possível utilizar apenas a estratégia anterior, que não dá conta desses elementos. Assim, foi necessário, também, separar as opções entre aspas (para considerar os espaços ou hifens) e o operador *or*. Enquanto "(work|works|worked|working) from home" identifica *work from home*, *works from home*, *worked from home* ou *working from home*, "work-from-home" identifica apenas *work-from-home*. Vale destacar que, muitas vezes, existe mais de uma expressão para chegar ao mesmo resultado. Aqui, poder-se-ia optar por "work from home" or "works from home" or "worked from home" or "working from home" or "work-from-home". No entanto, essa expressão ficaria bastante longa, não sendo essa uma boa prática, justamente porque o uso de expressões regulares visa à eficiência na busca.

Para o campo *Target* desta regra, elaborou-se a expressão "trabalh.* (em|de) casa". Em expressões regulares, o ponto-final representa qualquer caractere. Como o asterisco representa zero ou mais ocorrências do caractere imediatamente anterior a ele, temos, nessa expressão, a busca por quaisquer caracteres entre *trabalh* e *em/de casa,* como *trabalh**ando** em casa*, *trabalh**ava** em casa* e *trabalh**a** de casa*. Assim como nas regras anteriores, a caixa *PowerSearch* foi marcada para indicar que todo o texto entre aspas deveria ser buscado. O Quadro 5 mostra as informações inseridas para esta regra na checklist do Xbench.

Quadro 5 – Regra *trabalhar em/de casa*.

Name	trabalhar em/de casa
Description	Ao traduzir “work from home” e similares, em vez de utilizar “trabalhar em casa”, se o contexto permitir, prefira “fazer home office” ou “trabalhar em home office”. Aposte no uso do estrangeirismo.
Category	Colocação
PowerSearch	Sim
Source	"(work|works|worked|working) from home" or "work-from-home"
Config source	Regular Expression
Target	"trabalh.* (em|de) casa"
Config target	Regular Expression

Fonte: elaborado pela autora.

A próxima regra envolve várias possibilidades de verbo no texto-fonte, e a recomendação do verbo no texto-alvo também depende de alguns fatores. Trata-se da regra *exercer, gerar ou causar impacto?*. A colocação *causar impacto* ocorre com mais frequência no PTBR-TRAD (freq. 17) do que no PTBR-AUT (freq. 2). Realizou-se uma investigação de outras formas de expressar isso nos dois corpora do português e concluiu-se que, no PTBR-AUT, o verbo usado em português depende da prosódia semântica: *gerar impacto* quando positiva, *causar impacto* quando negativa e *exercer impacto* quando neutra. Além disso, constatou-se que, no texto-fonte, ocorre tanto *make an impact* quanto *have an impact*, ou simplesmente o verbo *to impact*.

Dessa forma, esta regra precisa abarcar todas essas opções do texto-fonte. Para o campo *Source*, foi elaborada a expressão ((makes*|made|making|have|has|had) impact) or ((impacts*|impacted|impacting) -"the impact" -"impacts* of"). Alguns caracteres usados em expressões regulares já foram explicados nas regras anteriores. Cabe ressaltar que se pode usar um par de parênteses no interior de outro, sendo que a expressão no interior é considerada como um bloco. Isso quer dizer que temos dois grandes blocos de opções nesta expressão: ((makes*|made|making|have|has|had|causes?|caused?) impact), que precede o operador *or*, e ((impacts*|impacted|impacting) -"the impact" -"impacts* of"), que sucede o operador *or*.

A primeira parte, ((makes*|made|making|have|has|had|causes?|caused?) impact), orienta a regra a identificar conjugações dos verbos *make*, *have* ou *cause*, seguido do substantivo *impact*. Note que esse primeiro bloco não é limitado por aspas, pois isso informaria à regra para desconsiderar palavras entre o verbo e o substantivo de busca, não sendo identificados casos como *makes* ***an immediate*** *impact* ou *has* ***a huge*** *impact*, bastante frequentes no corpus. A segunda parte inicia por (impacts*|impacted|impacting), que identifica diferentes conjugações do verbo *impact*. Porém, convém lembrar que o Xbench não trabalha com lematização, portanto não sabe imediatamente que esta regra busca apenas o verbo, e não o substantivo *impact*. Assim, se a expressão terminasse nesse ponto, a regra buscaria todas as ocorrências de *impact* e *impacts* também como substantivo, gerando muitos falsos-positivos, como em *The* ***impact*** *is being felt in the office, too* e *There are also other, more subtle* ***impacts***. Para contornar parte desse problema, completou-se a regra com -"the impact" -"impacts* of". Em expressões regulares, o hífen indica texto que não deve estar presente no segmento. Assim, esta regra desconsidera segmentos com as estruturas *the impact, impact of* e *impacts of*.

Quanto ao campo *Target* da regra, optou-se por deixá-lo em branco. Apesar de haver pelo menos três possibilidades convencionais de tradução (com os verbos gerar, causar ou exercer), é o contexto que indicará qual dessas opções é a mais adequada. Assim, independentemente da tradução usada, é importante que o controle de qualidade aponte esses itens para que o tradutor verifique se a prosódia semântica foi levada em conta.

Assim como nas regras anteriores, a caixa *PowerSearch* foi marcada para indicar que todo o texto entre aspas deve ser buscado. O Quadro 6 mostra as informações inseridas para esta regra na checklist do Xbench.

Quadro 6 – Regra *exercer, gerar ou causar impacto?*

Name	exercer, gerar ou causar impacto?
Description	Prefira “exercer impacto” quando a prosódia for neutra, “gerar impacto” quando for positiva e “causar impacto” quando for negativa. Exemplos: “O ambiente é um fator que exerce um grande impacto sobre o desempenho de qualquer profissional.” “É sobre ter uma função em que a ou o funcionário acredite que gera um impacto positivo para a comunidade e para si mesmo.” “A pandemia causou um impacto profundo no mercado de trabalho, afetando trabalhadores com menor proteção social e baixa escolaridade.”
Category	Colocação
PowerSearch	Sim
Source	((makes?|made|making|have|has|had|causes?|caused?) impact) or ((impacts?|impacted|impacting) -"the impact" -"impacts? of")
Config source	Regular Expression
Target	—
Config target	—

Fonte: elaborado pela autora.

A última regra trazida como exemplo refere-se ao uso de *geração Z*. Tal colocação tem frequência semelhante no PTBR-AUT (freq. 47) e PTBR-TRAD (freq. 44). No entanto, no PTBR-TRAD, o uso de caixa alta no substantivo é mais recorrente (*Geração Z*), espelhando o texto-fonte em inglês (*Gen Z*). Assim, a recomendação é utilizar letra minúscula (*geração Z*), como no PTBR-AUT.

Dessa forma, a regra precisa identificar segmentos que contenham *Gen Z* ou *Generation Z* no texto-fonte (independentemente de uso de caixa alta ou baixa) e, ao mesmo tempo, que tenham *Geração Z* no texto-alvo (em caixa alta). Para isso, elaborou-se a expressão "gen z" or "generation z" para o campo *Source* e "Geração Z" para o campo *Target*. Nessa regra, a opção *Regular Expression* só foi marcada na configuração de *Source*, pois faz uso do operador *or*. Além

disso, a opção *Case Sensitive* foi marcada na configuração de *Target*, para evitar que a regra mostrasse resultados com *geração Z*, que já utiliza caixa baixa como recomendado.

Assim como em regras anteriores, a caixa *PowerSearch* foi marcada para indicar que todo o texto entre aspas deveria ser buscado. O Quadro 7 mostra as informações inseridas para esta regra na checklist do Xbench.

Quadro 7 – Regra *geração Z*.

Name	geração Z
Description	Prefira escrever “geração” com “g” minúsculo.
Category	Colocação
PowerSearch	Sim
Source	"gen z" or "generation z"
Config source	Regular Expression
Target	"Geração Z"
Config target	Simple, Case Sensitive

Fonte: elaborado pela autora.

Estes exemplos trazem alguns dos caracteres especiais mais utilizados nas regras elaboradas e explicam seu funcionamento.

5 Considerações finais

Este trabalho parte de uma análise de colocações em um corpus de artigos de blogs de coworking escritos originalmente em português em comparação com as combinações encontradas em um corpus de textos do mesmo gênero traduzidos do inglês para o português. Tal análise é descrita em detalhes em Freitag (2025), sendo o foco do presente trabalho o compartilhamento de como se pode construir uma checklist do Xbench (ApSIC, 2020) para automatizar o processo de controle de qualidade especificamente para questões de convencionalidade.

Atualmente, os tradutores devem entregar um volume cada vez maior de textos traduzidos, em cada vez menos tempo (Drugan, 2013), sendo importante elaborar recursos

que auxiliem no controle de qualidade. As checklists do Xbench têm sido pouco utilizadas até o momento, sendo que um dos motivos pode ser a exigência de conhecimento sobre expressões regulares. Assim, este trabalho mostra como transformar uma análise com corpora em uma checklist do Xbench por meio de expressões regulares. Ressalta-se que esse procedimento pode ser aplicado a textos de qualquer gênero.

Além de trazer exemplos da elaboração de regras, destacando como determinados caracteres especiais funcionam, este trabalho se destaca por propor o uso desse recurso não para alertar sobre erros cometidos anteriormente ou preferências do cliente, mas sim para o uso adequado de colocações, considerando a convencionalidade dos textos. O uso de checklists nesse sentido pode conferir destaque ao papel do tradutor como especialista que não só domina a tradução na teoria e na prática, como também se apropria de ferramentas complexas, destacando-se num mundo em que os avanços tecnológicos não podem ser contidos, e comprovando que, apesar de excelentes, as tecnologias precisam ser manipuladas e controladas por especialistas para serem aproveitadas ao máximo.

Agradecimentos

Agradeço minha orientadora de doutorado, a profª drª Rozane Rebechi, por me conduzir nos quatro anos de pesquisa que possibilitaram publicações como esta. Agradeço também todos na empresa TextTrans, especialmente Claudio Costa, diretor executivo, por me colocar em posições de liderança e elaboração de materiais, como listas de Xbench voltadas para fins específicos, o que me permitiu abordar esse assunto com maestria em meus estudos acadêmicos.

Referências

APSIC. Xbench. Versão 3.0, build 1520. Barcelona, 2020. Programa de computador. Disponível em: https://www.xbench.net/index.php/download. Acesso em: 1 jan. 2020.

BAKER, Mona. Corpus Linguistics and Translation Studies: implications and applications. Text And Technology: In honour of John Sinclair, [S.l.], p. 233-250, 1993. John Benjamins Publishing Company. https://doi.org/10.1075/z.64.15bak.

DRUGAN, Joanna. Quality In Professional Translation: assessment and improvement. London: Bloomsbury, 2013.

FARKAS, András. **LF Aligner**. Versão 4.21. 2019. Disponível em: https://sourceforge.net/projects/aligner/. Acesso em: 20 jun. 2022.

FREITAG, P. H. **Colocações em artigos de blogs de coworking**: um olhar descritivo sobre textos autênticos e traduzidos. 2025. 284 f. Tese (Doutorado em Letras) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2025. Disponível em: https://lume.ufrgs.br/bitstream/handle/10183/295363/001290098.pdf?sequence=1&isAllowed=y. Acesso em: 09 jan. 2026.

FREITAG, P. H.; REBECHI, R. R. Analysing collocations in articles published in coworking blogs in Portuguese: a comparative study on authentic and translated texts. **Cadernos de Tradução**, Florianópolis, v. 45, p. 1-25, 2025. https://doi.org/10.5007/2175-7968.2025.e102039

KILGARRIFF, A. *et al*. The Sketch Engine: ten years on. **Lexicography**, [S.l.], v. 1, n. 1, p. 7-36, jul. 2014. Equinox Publishing. https://doi.org/10.1007/s40607-014-0009-9.

MOREIRA FILHO, José Lopes. **Python para Linguística de Corpus**: guia prático. São Paulo: Ed. do Autor, 2021.

OLIVEIRA, Tatiane Rocha. **Análise da tradução de combinatórias lexicais em tarefas de pós-edição de tradução automática**. 2019. 168 f. Dissertação (Mestrado) - Curso de Tradução, Universidade de Lisboa, Lisboa, 2019. Disponível em: https://www.inesc-id.pt/member/a35661e1-ac57-4468-b565-56b6b90e58f7/publications/. Acesso em: 23 out. 2025.

PEREIRA, Vera Lúcia Peralta. **As competências e o Workflow do tradutor em contexto empresarial**: experiência profissional na SMARTIDIOM. 2025. 50 f. Dissertação (Mestrado) - Curso de Tradução e Comunicação Multilingue, Escola de Letras, Artes e Ciências Humanas, Universidade do Minho, Braga, 2025. Disponível em: https://repositorium.uminho.pt/handle/1822/95388. Acesso em: 15 jul. 2025.

PINTO, Joana Inês Filipe. **Relatório de estágio**: Kvalitext - serviços de tradução, Lda. 2020. 90 f. Dissertação (Mestrado) - Curso de Tradução e Serviços Linguísticos, Faculdade de Letras, Universidade do Porto, Porto, 2020. Disponível em: https://sigarra.up.pt/flup/pt/PUB_GERAL.PUB_VIEW?pi_pub_base_id=410651. Acesso em: 15 jul. 2025.

PRETO, Ana Margarida Lemos Tavares. **A importância das tecnologias de apoio à tradução na formação e no trabalho do tradutor**: relatório de estágio curricular na SMARTIDIOM. 2021. 190 f. Dissertação (Mestrado) - Curso de Tradução e Interpretação Especializadas, Instituto Superior de Contabilidade e Administração do Porto, Porto, 2021. Disponível em: https://recipp.ipp.pt/bitstream/10400.22/19538/1/Ana_Preto_MTIE_2021.pdf. Acesso em: 10 jul. 2025.

SINCLAIR, John. **Corpus, Concordance, Collocation**. Oxford: Oxford University Press, 1991.

STRECKWALL, Fernando. Aseguramiento de la calidad y herramientas informáticas en traducción. In: CAGNOLATI, Beatriz Emilce; GENTILE, Ana María; SPOTURNO, María Laura (org.). **Problemas de traducción**: enfoques plurales para su identificación y tratamiento. La Plata: Edulp, 2023. p. 59-76. Disponível em: https://libros.unlp.edu.ar/index.php/unlp/catalog/book/2328. Acesso em: 10 jul. 2025.

STUPIELLO, Erika Nogueira de Andrade; ESQUEDA, Marileide Dias. Tecnologias e formação de tradutores. **Domínios de Lingu@Gem**, [S.L.], v. 11, n. 5, p. 1764-1781, 21 dez. 2017. https://doi.org/10.14393/DL32-v11n5a2017-20.

TAGNIN, S. E. O. **O jeito que a gente diz:** combinações consagradas em inglês e português. São Paulo: Disal, 2013.

Recebido em: 22.07.2025 Aprovado em: 26.01.2026