



## Analysing collocations in articles published in coworking blogs in Portuguese: A comparative study on authentic and translated texts

**Patrícia Helena Freitag**

Universidade Estadual do Ceará

Aracati, Ceará, Brasil

patricia.freitag@uece.br

<https://orcid.org/0000-0002-2446-3718> 

**Rozane Rodrigues Rebechi**

Universidade Federal do Rio Grande do Sul

Porto Alegre, Rio Grande do Sul, Brasil

rozane.rebechi@ufrgs.br

<https://orcid.org/0000-0003-0401-5691> 

**Abstract:** This paper reports on a descriptive and comparative study that falls within corpus based translation studies (Baker, 1993, 1995, 1996, 2004). The goal was to identify and describe similarities and differences in the collocations used in articles published in coworking blogs written originally in Portuguese and in texts of the same genre written in English and translated into Portuguese. Tagnin's (2013) taxonomy was used to inform our view of collocation. We assume that translated texts may use word combinations that do not sound natural in the target language because they are atypical in texts written originally in this language, following Mauranen's (2007) findings. Also, this paper's first author has worked with specialized translation for over a decade and observed that, in this context, there is a generalized expectation for texts to sound natural. We therefore argue that translators should be aware and have a good knowledge of collocations. The articles were collected from websites maintained by coworking companies, processed and analysed following corpus linguistics principles, a methodology that allows the identification of linguistic patterns. Two corpora were used: 1) a parallel corpus, with texts written originally in English and their corresponding translations into Portuguese and 2) a monolingual corpus with texts originally written in Portuguese. The corpora were stored in Sketch Engine (Kilgarriff et al., 2014), and the collocations were identified from n-gram lists from both corpora in Portuguese. The analyses found collocations that were frequently used in both corpora, such as *trabalho flexível* and *espaço de trabalho*, which describe professional environments and ways of working. As expected, differences were also found. One example is *escritório privativo*, which appears only in the translated corpus. In the authentic corpus, this concept is expressed as *sala privativa*. Thus, this article may be used to raise awareness, among translation students, professional translators, and professors, of the importance of looking for conventional word combinations in the target language while translating texts.

**Keywords:** collocation; translation; conventionality; blog; coworking.



## I. Introduction

This research study is situated within the Corpus-based Translation Studies. The leading scholar in this area is Mona Baker (1993, 1995, 1996, 2004), who proposed the use of corpora to explore patterns in translated texts. This research study compares translated texts and authentic texts. In particular, it focuses on the collocations used in articles published in coworking blogs. In short, collocations are words that go together “naturally; and, in general, there is no explanation for this fact”<sup>1</sup> (Tagnin, 2013, p. 63). That is, word A occurs with word B, and not with word C (which has the same semantic value as B), simply because this pattern has become conventionalized. A comprehensive definition is provided in Section 3, “Collocations: word combinations conventionalized by combinability”; for now, it is worth noting that one of the aspects that reveals whether a combination is typical is its frequency of use. We will also discuss in detail the types of collocations analyzed, namely: adjectival (formed by an adjective and a noun), nominal (two nouns), verbal (a verb and a noun), and adverbial (an adverb that modifies a verb or a noun).

An illustrative example is the nominal collocation *redução de custo*, which occurs 2,844 times in Brazilian texts in Corpus do Português: NOW (Davies, 2012-2019, hereinafter NOW), a corpus of news articles published online from 2012 to 2019 in Portuguese-speaking countries. The use of *diminuição de custo* complies with grammatical and semantic rules, but this word combination occurs only 131 times in NOW.

The same can be observed in the adjectival collocation *mudança radical* (1,073) compared to *mudança dramática* (61); in the verbal collocation *obter resultados* (1,324) compared to *entregar resultados* (252); and finally, in the adverbial collocation *criticar duramente* (79) compared to *criticar severamente* (3). Atypical combinations may sound awkward to the reader. For technical texts, we also believe they may lead the reader to question the author’s (or even the translator’s!) level of expertise.

Therefore, even when the use of atypical combinations in translation does not result in grammatical or meaning errors, it affects the quality of the translated text. In the technical and specialized translation industry, it is common practice for translation agencies to evaluate the quality of the texts and provide feedback to the translators involved in the work. To do this, evaluators usually use a form, in which they record the inadequacies they found and classify them according to the error typology adopted by the agency. TAUS, an organization dedicated to translation efficiency, provides its template free of charge<sup>2</sup>. The form includes the following categories: accuracy, fluency, terminology, style, design, locale convention, verity, and other, indicating that evaluators take multiple aspects into account. Drawing on professional experience, we observe that atypical combinations are most often recorded under Style. Some recurring comments from evaluators when recording style issues include: *the translation is too literal*, *the text lacks naturalness* and *this is not how people say it*. Considering these comments, we argue that translations not only need to be correct, but also sound natural. This highlights the need for research into the naturalness of translations, aiming to raise translators’ awareness of this issue.

<sup>1</sup> In the original: “de forma natural, não havendo, via de regra, explicação para o fato”. All translations from Portuguese for the quotes are our own.

<sup>2</sup> <https://info.taus.net/dqf-mqf-error-typology-template-download>

Tagnin and Teixeira (2004, p. 315) explain that naturalness is “how things are, in fact, said in a given area of a given language or linguistic variant”.<sup>3</sup> Drawing on corpora comprised of culinary recipes, the authors identify differences in the ways terms and phrases are used in Brazilian Portuguese and European Portuguese within the genre, and advocate for the publication of country-specific translations. A few years later, Lamparelli (2007) examined elements that promote naturalness in translations of texts from *National Geographic* magazine. The author argues that collocations are one of these key elements.

These studies reinforce that collocations (the focus of our study) contribute to textual naturalness, thus justifying our investigation. As Berber Sardinha explains in his book, which contributes to consolidate corpus linguistics (CL) in Brazil, this methodology allows for the identification of language patterns, with collocations being one example (Berber Sardinha, 2004). Therefore, we chose to employ CL in this research.

When we investigate patterns, we explore language conventions. Regarding linguistic forms, Nord (2018, p. 43) states that “a form that is conventional in one culture may be unconventional in another”. From this, we conclude that certain combinations may be conventional in the source language, but their literal translation might be unconventional in the target language, once again reinforcing the need for translators to pay attention to collocations.

We chose articles published in coworking blogs as the genre for this study due to the importance of coworking spaces in a world where remote work is becoming more prevalent. In 2022, Woba, a network of coworking spaces, conducted an analysis of the Brazilian market. A survey involving decision makers in coworking offices—such as CEOs, directors, shareholders, and managers—revealed a 63% increase in the number of coworking spaces in the country (Woba, 2023). The growth in the number of these spaces suggests that international coworking companies operating in Brazil would not want to miss the opportunity to expand their market share. To do so, they will likely need to have more of their content translated.

By shedding light on the translation of collocations, this research aims to support translation agencies in evaluating the naturalness of translations, recruiting translators who perform well in this regard, and training their current pool of translators. From an academic perspective, we expect the results to reinforce the status of Translation Studies as a scientific discipline and to inform and enrich translator training.

The paper is organized as follows: Sections 2, 3 and 4 present the theoretical framework of the study and provide definitions of key concepts, including naturalness, conventionality, and collocations. Section 5 presents the study reported in this paper, detailing the compilation and organization of the corpora as well as the procedures of analysis. The results and interpretation are presented in Section 6. The paper concludes with final remarks.

## 2. Naturalness and conventionality

Lopez-Rodriguez (2016) points out that creativity is fundamental to cognitive processes, including translation, and is also present in specialized translation. The author provides the following examples of linguistic creativity: neologisms, innovative uses of existing words, intentional

---

<sup>3</sup> In the original: “como as coisas, de fato, são ditas numa dada área de uma dada língua ou variante linguística”.

manipulation of nonstandard syntax, and other forms of linguistic invention. Hence, we infer that the creative process—whether in the production of authentic texts or in translation—results from conscious choices made when attested linguistic possibilities prove insufficient.

Atypical word combinations can be used intentionally and creatively, for example, to depict a character in a narrative as a non-native speaker of a particular language. For instance, one might illustrate this by associating the expression *eu precisar um livro* with an Anglophone character who, unfamiliar with the correct first-person singular in the simple present tense—*preciso*—, and unaware of the required preposition *de*, reproduces the structure of their native language, resulting in a literal and grammatically incorrect rendering in the target language.

In the existing literature of Translation Studies, the terms *naturalness* and *conventionality* are sometimes used interchangeably, as both refer to the use of words and structures that are established in the target language. Jiménez-Crespo (2013, p. 24) uses the term *naturalness* to refer to “translations that recipients perceive as natural productions in the target language”, while Putranti (2018, p. 99) argues that the term “can be seen in the use of appropriate TL expressions as well as TL structure”<sup>4</sup>. Regarding *conventionality*, Lopez-Rodriguez (2016, p. 92), while advocating for the use of corpora to study both conventional and creative translation solutions, states that it “refers to the idea [...] that much of the language use is routine”.

In this paper, *naturalness* refers to the extent to which texts sound fluent and not awkward, while *conventionality* refers to the use of recurrent patterns that are standard within the given language and genre. Obviously, we believe that the use of conventions contributes to the naturalness of texts.

In her book published in 2013, Brazilian researcher Stella Tagnin points out that there are several levels of conventionality in language and highlights three of them: syntactic, semantic, and pragmatic. At the semantic level, conventionality manifests itself in the “unmotivated relationship between an expression and its meaning”<sup>5</sup> (Tagnin, 2013, p. 26). An example is the expression *piece of cake*, used to refer to something easy to accomplish, despite there being no direct or motivated connection between the idea of ease and the image of a sliced baked treat.

At the pragmatic level, conventionality is revealed in the process of verbal expression that is appropriate to the communicative situation. For example, when a person is greeted with *How do you do?*, it is expected that they reply by repeating the same question, regardless of its literal meaning.

It is at the syntactic level that the author situates the aspect of combinability, a characteristic of certain words that are strongly associated due to their established use. The word order in a clause is also considered at this level. In the context of coworking blogs, for example, although expressions such as *distant work* and *remote job* comply with grammar rules, only *remote work* was conventionalized through usage.

### 3. Collocations: word combinations conventionalized by combinability

Sequences of words that occur systematically in language are given different names, depending on their characteristics and the combinability patterns they exhibit. N-grams, for example,

<sup>4</sup> For a more detailed discussion on naturalness, see Oliveira (2017), which provides definitions from various authors, in both monolingual and translation contexts.

<sup>5</sup> In the original: “relação não motivada entre uma expressão e seu significado”.

are continuous sequences of words or characters extracted from a text or corpus, regardless of whether they constitute a complete unit of meaning. Examples include the bigram *from home* and the trigram *work from home*.

Other examples of linguistic patterns include proverbs and idioms, which have specific characteristics. Structures such as *A problem shared is a problem halved* and *to bite one's tongue* exemplify, respectively, these types of prototypical word combinations, whose idiomatic meanings can hardly be inferred by the mere sum of the meanings of the individual words that compose them.

Hoey (2005) uses the term *collocation* to refer to the high probability of certain words appearing in each other's company, a phenomenon motivated by mental and contextual associations. *Remote work*, mentioned at the end of the previous section, is an example. One of the pioneers in this field, Firth (1957, p. 12) points out that, when discussing collocations, we are dealing with "mutual expectancy". This means that when we hear or read a word in a particular context of use, we expect it to occur alongside a specific word. For example, in the context of medical news, when we hear *excruciating* we already expect the next word to be *pain*. The formation of other combinations, such as *excruciating ache* or *massive pain*, would sound awkward. This could be used intentionally in humor and newspaper headlines, for example.

Provided that grammar norms are observed, there are no rules that determine which words should co-occur rather than others. In the previous example, we have an adjective of intensity followed by a noun. As long as this syntax is kept, there is nothing in the language itself that determines the use of the adjective *excruciating* instead of *massive*. It is all about usage.

Several authors have proposed classifications of collocations, differing slightly in scope, perspective, or definition. For example, in Pastor's (1996) taxonomy, which builds on Hausmann's (1989), noun-verb collocations are categorized separately from verb-noun collocations, rather than being grouped together as a single category of verbal collocations. Although many authors have discussed collocations, this paper adopts Tagnin's (2013) classification due to the emphasis she places on conventionality. The researcher is one of the most prominent scholars in this field in Brazil, with publications on the subject dating back to the 1980s. The author has an extensive body of work focused on the identification of conventional patterns. In addition, she also has research that addresses deviations from conventions as an intentional strategy to create humor (Tagnin, 2005).

According to the researcher, adjectival collocations are combinations composed of an adjective and a noun, regardless of the order of the elements (e.g., *agile business*, *heir apparent*). Note that this is especially true in languages such as Portuguese, in which adjectives may precede or follow the noun. In this type of collocation, both the adjective and the noun can be conventionalized. For example, we could say *collective area*, but the norm is to use *common area* to refer to a space in a coworking office where people can relax and interact with one another. In this example, the adjective is conventionalized. In the case of *social distancing*, the noun is conventionalized, in the sense that other nouns, such as *spacing* or *separation*, would be grammatically correct and could convey the same meaning, but the word combination is not widely accepted.

Nominal collocations are word combinations consisting of two nouns, sometimes linked by a preposition or hyphen. Some examples are *coworking space*, *board of directors*, and *cost-efficiency*. Verbal collocations consist of a verb and a noun, sometimes linked by an article or preposition. Some examples are *boost productivity*, *prompt the question*, and *work from home*. Verb-adjective combinations are also classified as verbal collocations, as in *go public*. Adverbial collocations consist

of word combinations in which an adverb modifies an adjective (*fully remote*) or a verb (*work remotely*). Tagnin (2013) also presents partitive expressions as a type of collocation, which she refers to as unit-specifying expressions. These are useful in languages such as English to indicate quantity when using uncountable nouns. This specification is less frequently needed in Portuguese, so we will not address it in detail. However, for the sake of clarification, we offer an example from Tagnin (2013, p. 14): *a piece of advice*, which in Portuguese is simply expressed as *um conselho*. Finally, Tagnin (2013) also classifies collective nouns formed by two nouns as collocations. These refer to a group of elements belonging to the same category, such as a *panel of experts*. We regard them simply as nominal collocations, given their noun-noun structure.

#### 4. Translating collocations

We have seen that collocations are words that frequently occur together. Readers of a given genre and community are thus familiar with such combinations, even if they are not consciously aware of them. On this aspect, Nord (2018, p. 43) states that “The more conventional the linguistic form, the less notice we take of it”. Therefore, to avoid that translated texts sound awkward and to ensure fluency, translators are advised to draw on collocations found in authentic texts representative of the genre they are working with.

However, for cognitive reasons, it can be challenging for translators to recall collocations in the target language. From studies in bilingualism and psycholinguistics, it is known that bilinguals’ languages are not stored in separate compartments in the mind; instead, they remain in constant interaction (Grosjean, 1989). It is also known that there are memory links between words when their form, word class or meaning are similar in the two languages. This occurs, for example, with cognates, which share certain semantic and orthographic features in the source and target languages. In such cases, Hansen-Schirra et al. (2017) claim that cognates are more frequently chosen when words are presented without contextual information. The authors conducted a study with trainee translators whose first language was German and additional language was English. The study involved two English to German translation tasks: a list consisting of 20 words and a text in which the same 20 words appeared. Each word could be translated either as a cognate or as a non-cognate. The results showed that the words were translated using cognates 57.39% of the time under the list condition, and 37.27% under the text condition. This indicates that the similarity between words can lead to translation choices that are not necessarily the most conventional in the target language, revealing source-text influence. However, the degree of this influence varies depending on some factors such as the presence of context and the level of expertise.

In light of these findings, we assume that translators can more readily recall specific translations for some words due to the links between languages. However, they do not always take collocational patterns into account, which can result in unnatural word combinations in the target language. For example, in our corpus of English-to-Portuguese translated texts, *altamente flexível* appears as a translation for *highly flexible*. The words *altamente* and *highly* share similar meanings in that both can refer to height. This memory link in the translator’s mind may explain the use of



*altamente* in the translation<sup>6</sup>. However, in our corpus of authentic Portuguese texts, the adverbs that typically occur with *flexível* are  *muito*, *bastante*, and *totalmente*. The word combination *altamente flexível* does not appear in this corpus. These three adverb options in the corpus—*muito*, *bastante*, and *totalmente*—seem to share less semantic similarity with *highly* when considered in isolation, without context. Therefore, they may not be as readily available as *altamente* in the minds of translators.

This example suggests that the individual words that make up the collocations in the source text may influence the translation if the translators approach the text word by word, sometimes resulting in atypical combinations in the target text. Whether due to lack of knowledge and research or to cognitive influence of the source language on the target language, atypical combinations within the given textual genre create a sense of awkwardness and negatively impact the clients' perception of translation quality.

In the literature, there are studies that use authentic bilingual corpora to aid in the translation of collocations. This is the case of Goldschmidt et al. (2017), who compiled a comparable corpus of authentic journalistic texts in Spanish and Portuguese. They extracted verbal and adjectival collocations from the Spanish subcorpus and used the Portuguese subcorpus to validate the translations. The authors were able to find equivalent collocations in Portuguese for most of the items extracted from the Spanish subcorpus, demonstrating the usefulness of corpora in identifying translation equivalents for collocations. Studies such as this consider that translations should not only conform to grammatical rules and ensure meaning accuracy, but also use word combinations that are typically used by language speakers within a given genre. The present study adopts this view and contributes to the academic literature on collocations in the English-Portuguese language pair, specifically in articles published on coworking blogs.

Nesselhauf (2003) investigates the production of word combinations in English by learners whose first language is German. The author classifies these combinations according to their degree of combinatorial restriction, ranging from least to most restricted: free combinations (e.g., *want a car/the truth/a drink*), combinations with inconclusive restriction (e.g.: *perform an experiment/task*), collocations with low restriction (e.g. *reach a decision/conclusion/verdict*), collocations with high restriction (e.g. *dial a number*), and idioms (e.g. *sweeten the pill*)<sup>7</sup>. Although we do not adopt this author's classification, one of the results of her research may shed light on some deviations from conventions we identified in our data. According to the researcher, learners made more mistakes in collocations with little restriction and fewer errors in collocations with high restriction. The researcher interprets this result as evidence that learners are more aware of restriction when it is high. Although the study is on English teaching and not translation, these observations may apply to translation. When a word has a high level of combinatorial restriction (that is, it occurs with only one or a few specific words), perhaps translators notice more easily when a given combination sounds awkward. On the other hand, when a word has lower combinatorial restriction (that is, it occurs with several words from the same semantic field), perhaps translators are not so aware of the restrictions or do not notice atypical combinations so easily. The correlation between the level

---

<sup>6</sup> We will not explore in detail cognitive concepts such as the priming effect, lexical access, or the notion of prototypes, as this goes beyond the scope of our research. To learn more about these phenomena, we refer readers to studies in the fields of psycholinguistics, cognitive psychology, and cognitive linguistics.

<sup>7</sup> Examples obtained from Nesselhauf (2003).

of combinatorial restriction and deviations from conventions while translating collocations is not in our scope of investigation. However, we argue that it is a relationship worth investigating in the future.

According to Bernardini (2016), the use of corpora to analyse translations encourages the development of critical thinking and detection of linguistic patterns, such as collocations, in real contexts of use. This methodological approach supports the empirical analysis proposed here, which seeks to identify conventional patterns and specific collocational variations in articles published in coworking blogs, both in authentic texts and in translations.

## 5. The study

The goal of the study was to identify and describe similarities and differences in the collocations used in articles published in coworking blogs written originally in Portuguese and in texts of the same genre written in English and translated into Portuguese. We assume that the links between the two working languages—English and Portuguese—may influence translation choices, sometimes resulting in unusual and unconventional word combinations. Therefore, we expected to find differences between the collocational patterns employed in the corpus of authentic Portuguese and the corpus of translated Portuguese. We also expected to explain some of these differences based on the identification of source-text influence.

Corpus linguistics was adopted as the methodological approach, as the use of corpora enables the identification and analysis of recurring patterns in a set of texts (Baker, 1993). As previously discussed, collocations are word combinations that became conventionalized through use, and translators should be familiar with them to produce texts that sound natural in the target language.

### 5.1 Corpora collection and organization

To compose the parallel corpus, only three blogs from international coworking companies that publish articles in English and their respective translations were found: Spaces, Regus, and WeWork. Since no other blogs meeting this criterion were found, it can be argued that the corpus comprises the entirety of available texts. Considering that “Representativeness refers to the extent to which a sample includes the full range of variability in a population” (Biber, 1993, p. 243) and that we had enough time to collect the totality of available articles, we decided to do so. The sum of articles found in the three blogs to compose the parallel bilingual corpus is 329 texts. The Portuguese subcorpus contains 12,795 types and 390,788 tokens<sup>8</sup>. The fact that there are few blogs to compose the bilingual corpus can be seen as a limitation, because it does not provide many words for analysis. However, this can also be considered a positive aspect, because the low number of texts may be due to the novelty of this genre. Thus, although modest, this bilingual corpus enables exploratory analysis of a genre not yet explored.

For the monolingual corpus, composed of articles originally written in Portuguese, there were dozens of national blogs available. The number of blogs and articles was excessive for our capacity. So, criteria were established for the selection of texts. The first criterion was that the blog

---

<sup>8</sup> *Type* refers to the unique words in a corpus, while *token* is the total number of word occurrences, including repetitions.



should have at least 40 articles published. This criterion was set because the international blog in the bilingual corpus with the least number of articles was Spaces, with 43 articles. In our view, this cut-off point excludes blogs that are not yet well established and includes only national blogs which are comparable to the international ones in terms of number of publications. The second criterion was to disregard blogs in which articles were too short compared to their counterparts. We observed that articles in most national blogs had at least 500 words, in general, so we disregarded blogs in which most of the articles were around 200 to 300-words long. These criteria guaranteed certain similarity between national and international articles and homogeneity among national blogs.

Six national blogs met the criteria. At least 50 articles were collected from each. It was not necessary to collect all the articles from the six blogs to reach a balanced number of tokens in relation to the Portuguese portion of the bilingual corpus. Because it contains articles from all blogs that met the criteria, we believe that the monolingual corpus is representative of the genre, allowing inferences and generalizations in this specific context (cf. Leech, 2007). In total, 535 articles were collected to compose the monolingual corpus (8,437 types and 396,296 tokens). A higher number of articles originally written in Portuguese was necessary to match the number of tokens of the monolingual and the bilingual corpora.

It should be noted that all the blogs were publicly accessible, not requiring registration, subscription, or payment. It was not possible to set a date range for the publication of the articles because not all of them contained this information. However, when this information was available, preference was given to articles published between 2017 and 2023, to ensure the inclusion of the most up-to-date texts representative of the genre. It is worth noting that, in most cases, there was no information about authorship either. When this information was given, the author was either a person or a department, but there was usually no specific information about the authors' background. Thus, it was not possible to control this variable.

The text collection was carried out manually. Each article was saved in a .txt file named the following way: LANGUAGE\_TYPE\_COMPANY\_NUMBER.txt, where LANGUAGE could be either ENG for English or PTBR for Portuguese, TYPE could be AUT of authentic or TRAD for translation, COMPANY was the company name and NUMBER was the number regarding the order of text selection for each company. An example is ENG\_AUT\_REGUS\_01.txt to indicate the first article in English, authentic, collected from the Regus blog. Metadata was recorded in a Microsoft Excel spreadsheet containing, in each row, the name of the .txt file, language, type, company, date of access, date of publication (when available), author (when available), and URL. The articles were saved in three folders: PTBR-AUT (authentic Portuguese), PTBR-TRAD (translated Portuguese), and ENG-AUT (authentic English). They were then uploaded into Sketch Engine (Kilgariff et al., 2014).

Also, for organization and preparation, the parallel corpus was aligned using LF Aligner (Farkas, 2019). The aligned .txt files were uploaded to AntPConc (Anthony, 2017), a concordance tool for parallel corpora.

## 5.2 Processing and analysis

Lists of key n-grams were generated for both Portuguese corpora, with 2 to 4 words. Even though collocations are mostly composed of two content words, n-grams with up to 4 words were extracted to include cases in which there are particles—such as articles and prepositions—between



the content words. *Espaço de escritório* was one of the extracted key n-grams. It is a nominal collocation consisting of two nouns, despite the presence of a preposition between them. This collocation would not be identified if the n-gram length was limited to two words<sup>9</sup>. Also regarding the key n-grams settings, the option A = a was selected to ensure that both uppercase and lowercase instances of a word were counted together in the overall frequency. The *lemma* option was used to consider all word inflections<sup>10</sup>. The minimum frequency was set to 5, and Portuguese Web 2020 (ptTenTen20) was chosen as the reference corpus. This allowed the identification of atypical and infrequent combinations, which was important for the purpose of this study. The lists were then exported to MS Excel and manually reorganized into three spreadsheets: one with n-grams that appeared in the lists of both corpora, another with n-grams that appeared only in the PTBR-AUT list, and finally one with n-grams that appeared only in the PTBR-TRAD list.

The cleaning of the lists was carried out simultaneously to the classification of the items<sup>11</sup>. For the three key n-gram lists, each row was checked. Items that did not meet the structure of collocations were deleted (such as *de coworking*, *sua equipe*, and *alugar um*), and the remaining items were classified as adjectival, nominal, verbal, or adverbial collocations. For such classification, Tagnin's (2013) structures were adopted. For example, verb-adverb combinations (such *trabalhar remotamente*) and adverb-adjective combinations (*totalmente remoto*) were classified as adverbial collocations.

Analyses to identify and investigate similarities and differences started with these lists. Several tools were used in this process, as needed and suitable for each case: 1) concordance lines in Sketch Engine were used to check the co-text in Portuguese (for both authentic and translated texts); 2) concordance lines in AntPConc were used to check the source text in English for items found in PTBR-TRAD; and 3) Word Sketch in Sketch Engine was used to identify other collocates of a search word. An example of each situation will be presented along the following paragraphs.

Concordance lines in Sketch Engine allow the search for one or more words to check their co-text and investigate their usage. By default, the tool applies lemmatization; however, this setting can be adjusted using the advanced search option. Take as an example the adverbial collocation *trabalhar colaborativamente*, from the PTBR-TRAD list. When searching for this word combination in PTBR-TRAD using Sketch Engine, the tool automatically lemmatizes the search words and presents results with other inflections of the verb *trabalhar*, such as *trabalharmos* and *trabalhando*. Figure 1 shows the search results from PTBR-TRAD.

---

<sup>9</sup> We recognize that, due to the collocation window, some collocations might not have been identified. This limitation was overcome during the analysis of the collocations in context with the use of the Concordance and Word Sketch tools.

<sup>10</sup> A lemma is the dictionary form of a word. For example, the lemma for *finding*, *finds*, and *found* is *find*.

<sup>11</sup> As lemmatized n-grams were used, the lists contained elements such as *sala privativo* and *escritório compartilhar*, because lemmatized forms are expressed as singular, masculine, and infinitive. For better readability, manual adjustments were made. For the given examples, the adjustments were *sala privativa* and *escritório compartilhado*.

Figure 1: Concordance lines for the query *trabalhar colaborativamente* in PTBR-TRAD using Sketch Engine

<input type="checkbox"/> Details	Left context	KWIC	Right context
1 <input type="checkbox"/> ⓘ doc#190	valioso para corretores, que podem	<b>trabalhar colaborativamente</b>	com um parceiro de vendas da We
2 <input type="checkbox"/> ⓘ doc#194	aremos mudar nossos hábitos para	<b>trabalharmos colaborativamente</b>	de maneira eficaz e segura. </s><s>
3 <input type="checkbox"/> ⓘ doc#197	Nova York estavam ansiosos para	<b>trabalhar colaborativamente</b>	e em pessoa. </s><s>Então, eles cr
4 <input type="checkbox"/> ⓘ doc#197	nários se animaram com a ideia de	<b>trabalhar colaborativamente</b>	e em pessoa. </s><s>Um funcionári
5 <input type="checkbox"/> ⓘ doc#210	s funcionários pudessem se reunir,	<b>trabalhar colaborativamente</b>	e manter sua cultura de união Com
6 <input type="checkbox"/> ⓘ doc#213	diretamente envolvidos e continuar	<b>trabalhando colaborativamente</b>	, com o auxílio dos recursos e do es
7 <input type="checkbox"/> ⓘ doc#214	o vídeo é sua melhor opção para	<b>trabalhar colaborativamente</b>	com a equipe. </s><s>Porém, com c

Source: search performed with Concordance in Sketch Engine.

[Description] Four-column header: Details, Left context, KWIC, and Right context. Seven sample lines, with document identifier, document excerpt and highlighted search words in red, that is, *trabalhar colaborativamente*, and the inflections *trabalharmos colaborativamente* and *trabalhando colaborativamente*. Concordance line 1: doc#190 valioso para corretores, que podem trabalhar colaborativamente com um parceiro de vendas da, Concordance line 2: doc#194 mudar nossos hábitos para trabalharmos colaborativamente de maneira eficaz e segura, Concordance line 3: doc#197 Nova York estavam ansiosos para trabalhar colaborativamente e em pessoa. Então, eles, Concordance line 4: doc#197 se animaram com a ideia de trabalhar colaborativamente e em pessoa. Um funcionário, Concordance line 5: doc#210 funcionários pudessem se reunir, trabalhar colaborativamente e manter sua cultura de união, Concordance line 6: doc#213 diretamente envolvidos e continuar trabalhando colaborativamente, com o auxílio dos recursos e do, Concordance line 7: doc#214 o vídeo é sua melhor opção para trabalhar colaborativamente com a equipe. Porém, com [End of description].

AntPConc is also a concordance tool, specifically designed for parallel aligned corpora. It was used to check the source texts corresponding to specific translated excerpts of interest. Unlike Sketch Engine, AntPConc does not lemmatize results. In other words, a search for *trabalhar colaborativamente* with AntPConc would retrieve only occurrences of the verb in the infinitive form. To overcome this limitation, it is possible to use an asterisk as a wildcard, allowing the tool to retrieve all occurrences of words containing the specified root. For example, *trabalh\** returns all inflections of *trabalhar*, as well as related words, such as *trabalhador*, *trabalhista* etc. However, this approach is not effective for verbs whose stems differ across inflected forms, such as *vou* and *ir*. Figure 2 shows the result of the search *trabalh\* colaborativamente*.

Figure 2: Concordance lines for the query *trabalhar colaborativamente* in AntPConc

Line		KWIC
1	riatividade e a segurança em espaços onde os membros da equipe	<b>trabalham colaborativamente</b> .
2	is remotos podem permanecer diretamente envolvidos e continuar	<b>trabalhando colaborativamente</b> , com o auxílio dos recursos e do es
3	A WeWork é um recurso valioso para corretores, que podem	<b>trabalhar colaborativamente</b> com um parceiro de vendas da WeWo
4	Quando você está em casa, o vídeo é sua melhor opção para	<b>trabalhar colaborativamente</b> com a equipe.
5	O escritório se tornou um lugar para	<b>trabalhar colaborativamente</b> com propósito, em vez de apenas o lo
Line		Reference
1	And they're doing all of this while finding new solutions to support both creativity and safety in spaces where their team members collaborate.	
2	By arranging a shared office or desk space in the city where they live, remote workers can stay closely involved and continue collaborating, supported by the resources and space they need to be productive.	
3	WeWork can be a valuable resource for brokers, who'll collaborate with one WeWork sales partner to find the ideal spot for their client.	
4	When you're home, video is your best bet for collaborating with your team.	
5	The office has become a place to collaborate with purpose, rather than just the default place to work.	

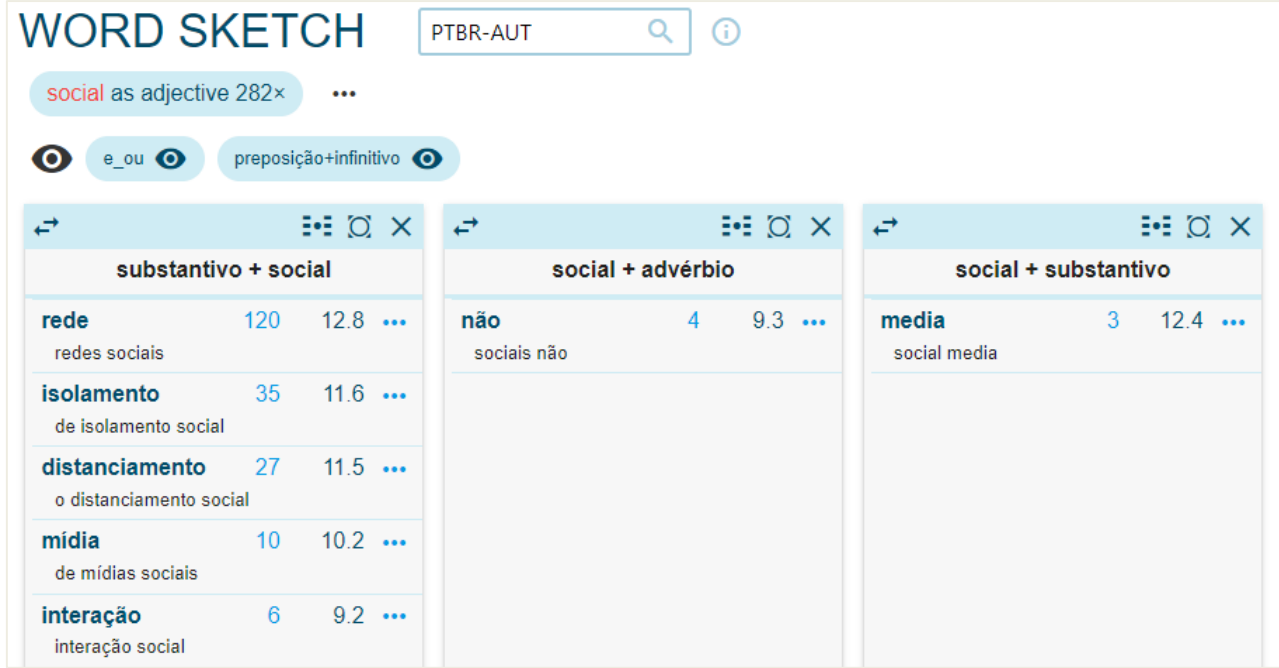
Source: Search performed in AntPConc.

[Description] Screen capture of AntPConc, separated in three parts. At the top, the tool ribbon with the search text *trabalh\* colaborativamente* for PTBR-TRAD. In the middle, the results in Portuguese, which are excerpts from the corpus,

with the verb *trabalhar* with different inflections in red, the adverb *colaborativamente* in blue and the following word in green. Concorcance line 1: criatividade e a segurança em espaços onde os membros da equipe trabalham colaborativamente. Concorcance line 2: remotos podem permanecer diretamente envolvidos e continuar trabalhando colaborativamente, com o auxílio dos recursos e do. Concorcance line 3: A WeWork é um recurso valioso para corretores, que podem trabalhar colaborativamente com um parceiro de vendas da WeWork. Concorcance line 4: Quando você está em casa, o vídeo é sua melhor opção para trabalhar colaborativamente com a equipe. Concorcance line 5: O escritório se tornou um lugar para trabalhar colaborativamente com propósito, em vez de apenas o. At the bottom, the English excerpts that contain the source text of the translations. Concorcance line 1: And they're doing all of this while finding new solutions to support both creativity and safety in spaces where their team member collaborate. Concorcance line 2: By arranging a shared office or desk space in the city where they live, remote workers can stay closely involved and continue collaborating, supported by the resources and space they need to be productive. Concorcance line 3: WeWork can be a valuable resource for brokers, who'll collaborate with one WeWork sales partner to find the ideal spot for their client. Concorcance line 4: When you're home, video is your best bet for collaborating with your team. Concorcance line 5: The office has become a place to collaborate with purpose, rather than just the default place to work [End of description].

The third type of query used in the analysis was the Word Sketch feature in Sketch Engine. The user searches for a word, and the tool shows the combinations identified, per grammatical category. To exemplify, we searched for the adjective *social* in both corpora to know whether they used *distanciamento social* and *isolamento social* in articles published during the covid-19 pandemic or whether only one of these word combinations was frequent.

Figure 3: Word Sketch results for the query *social* in PTBR-AUT



Source: Search using Word Sketch in Sketch Engine.

[Description] Screen capture of Word Sketch with results for the search of *social* as an adjective in PTBR-AUT. The results are shown in three columns, with the header of each column identifying the grammatical structure of the collocation. The collocates appear in the columns along with their respective number of occurrences. Column 1: substantivo + social: rede 120, isolamento 35, distanciamento 27, mídia 10, interação 6. Column 2: social + advérbio: não 4. Column 3: social + substantivo: media 3. [End of description].

As shown in Figure 3, Word Sketch automatically found collocates of *social* in the corpus of study, more specifically combinations formed by a noun followed by the adjective *social*, the adjective *social* followed by an adverb, and the adjective *social* followed by an adverb. Only the noun + social

structure is of interest to the example, showing that both *isolamento social* and *distanciamento social* are frequent in PTBR-AUT, with 35 and 27 occurrences, respectively. In PTBR-TRAD, the same query shows that only *distanciamento social* is frequent, with 144 occurrences. Word Sketch did not identify the noun *isolamento* as a collocate for *social* in this corpus. These examples demonstrate our analysis process.

## 6. Results and discussion

This section begins with a summary of the results represented in tables to exemplify the findings. Tables 1, 2, and 3 show examples of each collocation type. These are the items with the highest keyness scores in each case<sup>12</sup>. Due to space limitations, only three examples are presented for each collocation type. Table 1 shows collocations identified in both corpora—authentic and translated texts—with the numbers in bold highlighting the keyness scores that are higher in one of them.

Table 1: Collocations identified in both corpora

Collocation	Category	PTBR-AUT		PTBR-TRAD	
		Occurrences	Keyness	Occurrences	Keyness
trabalho flexível	Adjective	38	85.05	323	<b>790.86</b>
escritório virtual	Adjective	291	<b>776.83</b>	14	35.83
escritório compartilhado	Adjective	175	<b>600.61</b>	22	49.33
espaço de trabalho	Noun	170	341.42	678	<b>1505.49</b>
espaço de coworking	Noun	377	<b>950.97</b>	121	300.31
sala de reunião	Noun	304	<b>637.83</b>	238	490.45
trabalhar em casa	Verbal	92	189.04	143	<b>310.67</b>
fazer networking	Verbal	55	<b>145.61</b>	11	26.27
receber clientes	Verbal	79	<b>110.61</b>	6	13.62
trabalhar remotamente	Adverbial	22	54.91	96	<b>239.66</b>
mais produtivo	Adverbial	66	<b>131.94</b>	64	129.76
mais acessível	Adverbial	39	<b>76.13</b>	14	28.20

Source: Authors (2025).

[Description] A table with six columns, whose titles are: Collocation, Category, PTBR-AUT occurrences, PTBR-AUT keyness, PTBR-TRAD occurrences, and PTBR-TRAD keyness. The data are as follows, left to right, top to bottom: trabalho flexível, Adjective, 38, 85.05, 323, 790.86, escritório virtual, Adjective, 291, 776.83, 14, 35.83, escritório compartilhado, Adjective, 175, 600.61, 22, 49.33, espaço de trabalho, Noun, 170, 341.42, 678, 1505.49, espaço de coworking, Noun, 377, 950.97, 121, 300.31, sala de reunião, Noun, 304, 637.83, 238, 490.45, trabalhar em casa, Verbal, 92, 189.04, 143, 310.67, fazer networking, Verbal, 55, 145.61, 11, 26.27, receber clientes, Verbal, 79, 110.61, 6, 13.62, trabalhar remotamente, Adverbial, 22, 54.91, 96, 239.66, mais produtivo, Adverbial, 66, 131.94, 64, 129.76, mais acessível, Adverbial, 39, 76.13, 14, 28.20. [End of description].

<sup>12</sup> In corpus linguistics, keyness score is a measure used to compare the occurrences of a word or n-gram in a study corpus with the occurrences in a reference corpus. The items with higher keyness scores are those that occur significantly more frequently in the study corpus. The keyness score used in Sketch Engine is *simple maths*, and the reference corpus used was ptTenTen20.

Table 2 shows collocations identified only or mostly in PTBR-AUT.

Table 2: Collocations identified only or mostly in PTBR-AUT

Collocation in PTBR-AUT	Category	Occurrences	Keyness
sala privativa	Adjective	359	910.38
endereço comercial	Adjective	163	460.66
endereço fiscal	Adjective	141	372.81
espaço coworking	Noun	73	184.34
redução de custo	Noun	81	156.71
rotina de trabalho	Noun	70	142.81
alugar sala	Verbal	61	109.09
continuar a leitura	Verbal	43	98.67
contratar serviço	Verbal	53	73.96
bem localizado	Adverbial	47	96.53
entender melhor	Adverbial	28	87.05
saber mais	Adverbial	102	75.77

Source: Authors (2025).

[Description] A table with four columns, whose titles are: Collocation in PTBR-AUT, Category, Occurrences, and Keyness. The data are as follows, left to right, top to bottom: sala privativa, Adjective, 359, 910.38, endereço comercial, Adjective, 163, 460.66, endereço fiscal, Adjective, 141, 372.81, espaço coworking, Noun, 73, 184.34, redução de custo, Noun, 81, 156.71, rotina de trabalho, Noun, 70, 142.81, alugar sala, Verbal, 61, 109.09, continuar a leitura, Verbal, 43, 98.67, contratar serviço, Verbal, 53, 73.96, bem localizado, Adverbial, 47, 96.53, entender melhor, Adverbial, 28, 87.05, saber mais, Adverbial, 102, 75.77. [End of description].

Table 3 shows collocations identified only or mostly in PTBR-TRAD.

Table 3: Collocations identified only or mostly in PTBR-TRAD

Collocation in PTBR-TRAD	Category	Occurrences	Keyness
trabalhador remoto	Adjective	47	120.61
corretor imobiliário	Adjective	37	103.03
novo mercado	Adjective	38	79.21
estilo de trabalho	Noun	36	94.24
plano de negócio	Noun	41	89.97
tempo de deslocamento	Noun	37	86.14
retornar ao escritório	Verbal	37	93.01
reduzir a densidade	Verbal	22	49.65
adotar abordagem	Verbal	17	37.47
mais propenso	Adverbial	21	45.99
trabalhar colaborativamente	Adverbial	14	36.64
mais velho	Adverbial	30	27.99

Source: Authors (2025).

[Description] A table with four columns, whose titles are: Collocation in PTBR-TRAD, Category, Occurrences, and Keyness. The data are as follows, left to right, top to bottom: trabalhador remoto, Adjective, 47, 120.61, corretor imobiliário, Adjective, 37, 103.03, novo mercado, Adjective, 38, 79.21, estilo de trabalho, Noun, 36, 94.24, plano de negócio, Noun, 41, 89.97, tempo de deslocamento, Noun, 37, 86.14, retornar ao escritório, Verbal, 37, 93.01, reduzir a densidade, Verbal, 22, 49.65, adotar abordagem, Verbal, 17, 37.47, mais propenso, Adverbial, 21, 45.99, trabalhar colaborativamente, Adverbial, 14, 36.64, mais velho, Adverbial, 30, 27.99. [End of description].





The next subsections present the results and discussions for each collocation type. In this paper, due to space limitations, only the item with the highest keyness score of each type is analyzed in detail.

## 6.1 Adjectival collocations

This subsection analyses three adjectival collocations that illustrate different distribution patterns across the corpora—frequent in both, exclusive to authentic texts, and predominant in translated texts.

### 6.1.1 Adjectival collocation with the highest keyness score present in both corpora

The adjectival collocation with the highest keyness score present in both corpora is *trabalho flexível*, with 38 occurrences (keyness 85.05) in PTBR-AUT and 323 occurrences (keyness 790.86) in PTBR-TRAD. The numbers indicate that this collocation occurs significantly more in the translated texts. This may suggest that national coworking companies do not focus on flexibility. To verify this, we checked the frequency of *flexível* and *flexibilidade* in both corpora. The adjective *flexível* appears 194 times in PTBR-AUT and 628 in PTBR-TRAD, while the noun *flexibilidade*, 225 times in PTBR-AUT and 281 in PTBR-TRAD. Although the difference in the frequency of *flexível* is considerable, the difference for *flexibilidade* is not. Therefore, we believe that flexibility is a valued feature for both national and international coworking companies.

An examination of the concordance lines revealed that, in both corpora, the collocation occasionally appears nested, such as in *ambientes de trabalho flexíveis* and *condições de trabalho flexível*. In these cases, we found that, among the nouns that precede *trabalho flexível*, in PTBR-AUT, the most frequent one is *espaço*, but there are also instances of *cultura* and *rotina*. In PTBR-TRAD, *espaço* is also the most common noun, but there are instances of many others, such as *acordo*, *condições*, *cronograma*, *horário*, *local*, *modelo*, *política*, and *prática*.

We then investigated whether, in authentic texts, these nouns also appear immediately followed by *flexível*, without the word *trabalho* in between. Using CQL, a feature that allows advanced searches for grammatical patterns – we applied the query `[tag="N.*"] [lemma="Flexible"]` –, we indeed found occurrences such as *contrato flexível*, *horário flexível*, *modelo flexível*, among others. Based on these findings, a suggestion for translators is to alternate the use of nested collocations by sometimes including and other times omitting the word *trabalho*. For example, varying between *horário flexível* and *horário de trabalho flexível* would help reduce the repetition of the collocation *trabalho flexível* in translations, thereby aligning more closely with the patterns found in authentic texts.

### 6.1.2 Adjectival collocation with the highest keyness score present only in PTBR-AUT

The adjectival collocation with the highest keyness score present only in PTBR-AUT is *sala privativa*, with 359 occurrences (keyness 910.38). It is a quieter, less exposed area in a coworking space, intended for use only by the person or group that booked it. Although most coworking spaces

are usually booked per hour, this is not the case with private offices, which are usually booked for longer periods.

Although this collocation does not appear in PTBR-TRAD, we believe that the concept might be expressed differently, given that this type of space is valuable for businesses seeking a consolidated presence. Using Word Sketch, we searched for the node word *sala* hoping to identify another frequent adjective with a similar meaning to *privativo*. However, the combinations found included *sala comercial*, *sala espaçosa*, *sala pequena*, *sala grande*, and *sala interna*.

Another attempt was to search for the collocate, that is, the adjective *privativo*, using Word Sketch in PTBR-TRAD. There were 50 occurrences of *escritório privativo*. The concordance lines for *escritório privativo* in PTBR-TRAD reveal that the meaning for this combination seems to be the same as the one for *sala privativa* in PTBR-AUT.

To assess potential source text influence, a search for *escritório privativo* was performed using AntPConc. The corresponding English collocation identified was *private office*, suggesting both positive and negative influence from the source text. The influence was positive in the translation of the adjective: the cognates *private* and *privativo* are the adjectives typically used in authentic English and Portuguese texts, respectively, to form the collocation of interest. However, it was negative in the translation of the noun *office*, whose *prima facie* equivalent *escritório* does not correspond to the noun conventionally used in authentic Portuguese texts for this collocation. Thus, it is recommended that translators prioritize the use of *sala privativa* to convey the meaning of *private office* more appropriately.

### 6.1.3 Adjectival collocation with the highest keyness score present mainly in PTBR-TRAD

The adjectival collocation with the highest keyness score, appearing mainly in PTBR-TRAD, is *trabalhador remoto*, with 47 occurrences (keyness 120.61). In contrast, PTBR-AUT contains only three occurrences. Using AntPConc, we found that the English source text corresponding to *trabalhador remoto* was most often *remote worker*. This collocation refers to a person who does not work on-site at a company's headquarters or branch office, but rather works from home, coworking spaces, coffee shops, or other locations. Given that this profile aligns closely with the target audience of coworking companies, it is likely that a more established way of referring to such professional exists in PTBR-AUT.

The search for *remoto* using Word Sketch for both corpora revealed that the adjective occurs 249 times in PTBR-AUT and 453 in PTBR-TRAD. In addition, in PTBR-AUT, the only nouns that co-occur significantly with *remoto* are *trabalho* (freq. 150<sup>13</sup>) and *forma* (freq. 19), whereas in PTBR-TRAD there are more nouns: *trabalho* (freq. 190), *trabalhador* (freq. 47), *funcionário* (freq. 27), *profissional* (freq. 21), and *colaborador* (freq. 11). Based on these findings, we interpret that the authentic texts in Portuguese tend not to use the adjective *remoto* to describe people, whereas the use of *remote* appears to be common in English texts. This contrast often leads to deviations from conventions in translated texts. Therefore, when translators encounter remote modifying nouns that refer to

<sup>13</sup> From here, the abbreviation *freq.* is used to represent the word *frequency*.

people—such as *worker* and *employee*—, we suggest opting for alternative solutions, such as *que trabalha de forma remota*.

It is also worth mentioning that this appears to be a case of exclusively negative source-text influence. Translating *remote* using the cognate *remoto* results in a deviation from normative grammatical structures in Portuguese, where the typical formulation involves the verb *trabalhar* followed by adverbial phrase *de forma remota*, rather than the use of an attributive adjective. Likewise, translating *worker* with the *prima facie* equivalent *trabalhador* also demonstrates negative source-text influence. In authentic texts in Portuguese, we found expressions such as *profissionais que trabalham de forma remota*, but not *trabalhadores que trabalham de forma remota*, possibly to avoid repeating the root *trabalh*\*.

In addition, it should be noted that in Brazil the word *trabalhador* is sometimes used in political contexts to advocate for the rights of people with low-status jobs, while *profissional* is used to refer to people occupying prestigious jobs. Thus, it is important for translators to be aware of these nuances and make an informed decision, knowing that if they use *trabalhador* they may be adding a political trait absent in the source text.

## 6.2 Nominal collocations

Here we examine nominal collocations with the highest keyness scores, focusing on their presence in both corpora, exclusively in the authentic corpus, or predominantly in the translated corpus.

### 6.2.1 Nominal collocation with the highest keyness score present in both corpora

The nominal collocation with the highest keyness score present in both corpora is *espaço de trabalho*, with 170 occurrences (keyness 341.42) in PTBR-AUT and 678 occurrences (keyness 1,505.49) in PTBR-TRAD. Although this collocation is more frequent in translated texts, it is also commonly used in authentic texts. Despite the disparity in frequency, the concordance lines reveal similar usage patterns across both corpora.

In PTBR-AUT, we observed that the collocation *espaço de trabalho* is sometimes accompanied by an adjective that modifies the node word (e.g., *espaços de trabalho compartilhados*), and less frequently by an adjective modifying the collocate (e.g.: *espaços de trabalho colaborativo*), with the former being significantly more common. In some instances, the adjective is adjacent to the node, avoiding ambiguity (e.g.: *espaço compartilhado de trabalho*).

In PTBR-TRAD, the same pattern was found: many cases where the adjective modifies *espaço* (e.g. *espaços de trabalho flexíveis*) and few in which it modifies *trabalho* (e.g. *espaços de trabalho colaborativo*). In contrast to PTBR-AUT, PTBR-TRAD contains no instances in which the adjective appears between the node and the collocate. The absence of this position may be due to source text influence since the word in English is *workspace*, which is the agglutination of *work* and *space*. In any case, as the use of the collocation *espaço de trabalho* is similar in both corpora, it does not seem to cause any translation difficulty.

### 6.2.2 Nominal collocation with the highest keyness score present only in PTBR-AUT

The nominal collocation with the highest keyness score present only in PTBR-AUT is *espaço coworking*, with 73 occurrences (keyness 184.34). Although this construction, without the preposition *de*, occurs 73 times in this corpus, the construction with preposition, *espaço de coworking*, occurs 377 times. In any case, the 73 occurrences of *espaço coworking* should not be neglected.

Comparing the two forms, we believe that in *espaço de coworking*, the word *coworking* refers to an activity, that is, the space is where the activity of working with colleagues and peers is performed. Meanwhile, in *espaço coworking*, the word *coworking* gains the meaning of a physical space. This interpretation is supported by the occurrences of the word *coworking* used on its own to refer to the location, as in the sentence *Em um coworking, há espaços específicos para reuniões*, a use found exclusively in PTBR-AUT. The fact that PTBR-TRAD includes only occurrences of *espaço de coworking* suggests that these texts view *coworking* solely as an activity. In the case of PTBR-AUT, the existence of both *espaço coworking* and *espaço de coworking* suggests that authentic texts sometimes view *coworking* as an activity and sometimes as a space. Based on these findings, we recommend that translators take notice of these nuances. When translating the word *workspace*, translators should choose *espaço de coworking* the term clearly refers to the activity, and *espaço coworking* or simply *coworking* when the focus is on the physical space.

### 6.2.3 Nominal collocation with the highest keyness score present mainly in PTBR-TRAD

The nominal collocation with the highest keyness score present mainly in PTBR-TRAD is *estilo de trabalho*, with 36 occurrences (keyness 94.24). With AntPConc, it was found that the source text contained *style of working* or *work(ing) style*. The concordance lines in PTBR-TRAD do not clearly indicate what *estilos de trabalho* are exactly, but suggest that there are specific spaces for each style. For example, meeting rooms seem to be suitable for styles that require exchange of ideas or group brainstorming, while private offices are best suited for those that require silence and privacy.

In PTBR-AUT, *estilo de trabalho* occurs only 3 times and seems related to work models, such as remote, on-site or hybrid. A search for *modo de trabalho* was performed as an alternative, but this word combination also has low frequency in this corpus and is used with the same sense of *work model*. Another possibility that we investigated was *jeito*, hoping to find *jeito de trabalho* or *jeito de trabalhar*. Although neither of these were found, there were instances of *trabalhar do seu jeito*, in contexts similar to *style of working* and *working style*. Thus, when translators render one of these collocations in Portuguese, we recommend using *trabalhar do seu jeito*, with the necessary adjustments in inflections and word forms, if the context refers to the way of performing certain activities, that is, in more reserved or dynamic spaces.

## 6.3 Verbal collocations

This subsection analyses verbal collocations with the highest keyness scores, showing how differences across corpora may stem from source-text influence and highlight the need for translators to consider collocational conventions in the target language.



### 6.3.1 Verbal collocation with the highest keyness score present in both corpora

The verbal collocation with the highest keyness score present in both corpora is *trabalhar em casa*, with 92 occurrences (keyness 189.04) in PTBR-AUT and 143 occurrences (keyness 310.67) in PTBR-TRAD. Although *trabalhar de casa* is not among the items with the highest keyness score, we decided to check its frequency, due to its similarity with the analyzed collocation. This search resulted in 26 occurrences in PTBR-AUT and 35 in PTBR-TRAD. With AntPConc, we found that the English source text that was translated as *trabalhar em casa* and *trabalhar de casa* is *work from home*. Since *work* is usually translated as *trabalhar*, and *home* as *casa*, this is a case of *prima facie* translation. The higher frequency of *trabalhar em casa* and *trabalhar de casa* in PTBR-TRAD suggests that the target text may have been influenced by the source text, causing translators to use these word combinations more often than in authentic Portuguese texts.

We searched for *work from home* in the English subcorpus of the parallel corpus to verify if there were other translations that shed a light on ways of expressing this idea that we could search in PTBR-AUT. We found *fazer home office*, which can be used interchangeably with *trabalhar em/de casa* in many contexts. We decided to check the frequency of *home office* in PTBR-AUT to understand whether it compensates for the lower frequency of *trabalhar em casa* in this corpus compared to PTBR-TRAD. There are 510 occurrences of *home office*, a number that more than compensates for the low frequency of *trabalhar em casa*. Therefore, our suggestion is to translate *work from home* preferably as *fazer home office* and sparingly as *trabalhar em/de casa*.

### 6.3.2 Verbal collocation with the highest keyness score present mainly in PTBR-AUT

The verbal collocation with the highest keyness score present mainly in PTBR-AUT is *alugar sala*, with 61 occurrences (keyness 109.09). The concordance lines show that this collocation is mainly nested with other collocations, in *alugar sala privativa* and *alugar sala de reunião*. In PTBR-TRAD, *alugar sala* occurs only 3 times. To find out if the translated texts use another verb, we searched the node word *sala* using Word Sketch in this corpus and found 15 occurrences of *reservar*, some of which are *reservar sala de reunião*, which is a similar use to *alugar sala* in PTBR-AUT.

With AntPConc, we found that the English excerpts translated as *reservar sala* contained *book room* and *reserve room*. The translation of *reserve* as *reservar* is performed using a cognate word, while the translation of *book* as *reservar* may be seen as *prima facie*. We believe this is the case because *book* is widely used in the collocation *book a room* in the hospitality field, which is usually expressed as *reservar um quarto*. It is possible that this well-known hospitality collocation has influenced the translators' choice in the coworking context, causing them to employ this common translation for *book*. Therefore, when encountering *book room* and *reserve room* in articles published in coworking blogs, we recommend the use of *alugar sala*.

### 6.3.3 Verbal collocation with the highest keyness score present mainly in PTBR-TRAD

The verbal collocation with the highest keyness score present mainly in PTBR-TRAD is *retornar ao escritório*, with 37 occurrences (keyness 93.01). With AntPConc, we found that the source text contained *return to the office* in most instances. Although *retornar ao escritório* is a somewhat free

word combination because there is also the possibility of using *voltar ao escritório* (also with 37 occurrences in this corpus), we decided to consider it as a collocation due to its high keyness score, indicating that it has a prominent presence in translated articles published in coworking blogs. The concordance lines reveal that these collocations were used after the covid-19 pandemic, when professionals were considering working in offices once again.

In PTBR-AUT, there are only three occurrences of *retornar ao escritório*. The concordance lines of the verb *retornar* in PTBR-AUT show that, in this corpus, it is common to write about returning not to a physical space, such as an office, but to something less concrete, such as *voltar ao trabalho presencial*, *às atividades*, and *à rotina*. Thus, our recommendation to translate *return to the office* is to consider whether these less concrete options are possible given the context.

## 6.4 Adverbial collocations

Next we present the adverbial collocations with the highest keyness scores, highlighting differences in frequency and usage across corpora and showing how source-text influence and genre conventions may shape translation choices.

### 6.4.1 Adverbial collocation with the highest keyness score present in both corpora

The adverbial collocation with the highest keyness score present in both corpora is *trabalhar remotamente*, with 22 occurrences (keyness 54.91) in PTBR-AUT and 96 occurrences (keyness 239.66) in PTBR-TRAD. AntPConc reveals that the English source text for the translations in the bilingual corpus is *work remotely* in most cases. The concordance lines suggest that the contexts of use are similar in both corpora: digital nomads, who can work from anywhere; types of work that lend themselves to the remote model; and tips for managing teams that work remotely. If the contexts are similar, why is the frequency much higher in the translated texts?

We investigated whether there were alternatives with the same meaning in PTBR-AUT that would compensate for the lower frequency of *trabalhar remotamente* in this corpus. We found *trabalhar a distância*, with 10 occurrences in PTBR-AUT and only 2 in PTBR-TRAD, and *trabalhar + coworking* (*trabalhar em um coworking*, *trabalhar em ambientes de coworking*, etc.), with 135 occurrences in PTBR-AUT and only 5 in PTBR-TRAD. We realize that *trabalhar em um coworking* is more specific than *trabalhar remotamente* since *remotamente* does not specify the location. However, we interpret that the higher frequency of *trabalhar remotamente* in PTBR-TRAD and of *trabalhar em um coworking* and *trabalhar a distância* in PTBR-AUT reveal different approaches to the same topic. Therefore, we recommend that translators consider translating *work remotely* as *trabalhar em um coworking* when they deem adequate, encouraging people to book these spaces.

### 6.4.2 Adverbial collocation with the highest keyness score present mainly in PTBR-AUT

The adverbial collocation with the highest keyness score present mainly in PTBR-AUT is *bem localizado*, with 47 occurrences (keyness 96.53), with only 2 occurrences in PTBR-TRAD. There is no doubt that international coworking companies also need to consider the location of their spaces.





Therefore, we tried to identify whether the translations used another way of expressing this idea, possibly influenced by the source text words.

With AntPConc, we checked the source text for *bem localizado* in the bilingual corpus and found that one instance in English used *well-positioned* and the other used *perfectly positioned*. Using these options as clues of English excerpts that might have influenced the translations, we searched for *position\** using AntPConc and found several structures to refer to the location of the offices: *occupies a prime position*, *enjoys a premium position*, *it's ideally positioned*, among others. To render all of them in Portuguese, the translators chose the verb *posicionar*, with the appropriate inflection in each instance.

As the translations seemed to be influenced by the cognate words *position-posição*, we also searched *localized*, believing that the translations might have also been influenced by its cognate, resulting in *localizado* in the Portuguese texts. That is indeed what happened: we found three occurrences of *conveniently located* translated as *convenientemente localizado*.

These searches show that when the words *located* or *positioned* appear in the source text, the translation seems to be influenced by them and applies the cognate words *localizado* and *posicionado*. While the influence is positive in the first case, since it results in a convention form in the target language, the second case demonstrates negative influence because it results in an atypical option in the genre in Portuguese.

Moreover, we emphasize that authentic Portuguese texts showed less variation of collocations in this context, being *bem localizado* the most prevalent. Thus, it is advisable for translators to consider the option *bem localizado* even if the source text contains *occupies a prime position*, *is perfectly positioned* or *conveniently located*. Another option would be to modify the way of expressing the idea and choosing *localização privilegiada*, which also occurs with some frequency in PTBR-AUT (33 times).

### 6.4.3 Adverbial collocation with the highest keyness score present mainly in PTBR-TRAD

The adverbial collocation with the highest keyness score present mainly in PTBR-TRAD is *mais propenso*, with 21 occurrences (keyness 45.99). With AntPConc, we found that the source text contained *be more likely to* in most instances. The concordance lines reveal that this word combination is used to describe participants' responses in surveys, as in *Uma pesquisa mostra que 33% das pessoas que enfrentam deslocamentos longos (mais de 60 minutos por trecho) estão mais propensas a apresentar quadros depressivos*.

In PTBR-AUT, there are only two occurrences of this combination, also used in the context of surveys. We did not find another collocation in PTBR-AUT that expressed a similar idea, such as *mais chances* or *maior probabilidade*. The concordance lines for *pesquisa* in authentic texts suggest that these do not discuss how likely people are to do something. Instead, they present the findings more categorically, as in *A pesquisa revelou ainda que 74% gostavam de trabalhar em casa*. Therefore, we recommend that translators consider whether it is possible to modify the presentation of the data, avoiding rendering in the Portuguese texts the idea expressed by *more likely to*.

## 7. Concluding remarks

This descriptive and comparative study offers important findings and interpretations regarding the use of collocations in translated articles published in coworking blogs. Some collocations, such as *trabalho remoto* and *espaço de coworking*, were found in both corpora (PTBR-AUT and PTBR-TRAD), while others appeared exclusively or predominantly in one of them.

The differences identified are due to several factors, revealing the complexity involved in translating these texts. In some cases, the variation does not refer to how a concept is expressed, but rather to the fact that one of the corpora addresses a particular topic and the other does not. We do not interpret this as a deviation from collocational conventions. Although this aspect was not the focus of our analysis, we deemed it important to note.

In other cases, we observed that the differences between the collocations identified in the authentic and translated corpora stem from translations that mirror the English source text, resulting in atypical word combinations (such as *escritório privativo*, influenced by *private office*). This occurs because cognates and *prima facie* translations are more readily recalled than alternative expressions, which is in line with findings from Hansen-Schirra, Nitzke and Oster's (2017) study on cognates.

Less frequently, we observed differences in how certain terms are conceptualized across the two Portuguese corpora. This is exemplified by the term *coworking*, which may refer to the physical space or to the activity performed within it. This distinction was evident in the use of *espaço coworking* in authentic texts, a usage not found in the translated texts. It should be noted that this study did not have access to some specific information, such as the date of publication for the texts since not all blogs make this information available, which may have influenced the linguistic patterns identified. Another missing information was the metadata about the authorship of the texts.

Regarding the possibilities for future research, we believe that it would be productive to replicate this study using more established textual genres, which would allow for the compilation of larger corpora, which could provide us with more patterns and generalizable findings. Another aspect that can be investigated is the correlation between the level of combinatorial restriction, perhaps based on Nesselhauf's (2003) classification, and the deviations from convention in translation.

The considerations presented regarding the differences in collocation use across the corpora may help raise awareness among professional and trainee translators, translation instructors, and translation agencies about the importance of producing word combinations that are conventional in the target language in a given genre. In addition to acquiring theoretical knowledge about this phraseological unit, we recommend that translators learn about frequency scores and association measures, and learn to search for collocations in corpora tools to find conventional solutions for their translations.

Although the analysis was performed on corpora of articles published in coworking blogs and focused on collocations specific to this genre, the considerations presented – as well as the procedures for extracting and analysing collocations –, can be applied to other genres. In sum, we emphasize the importance of comparing translated corpora with authentic texts to identify deviations from collocational conventions, and we highlight the importance of consulting authentic corpora to find conventional collocation in the target language.



## References

- Anthony, L. (2017). *AntPConc* (1.2.1). [Software]. <https://www.laurenceanthony.net/software>
- Baker, M. (1993). Corpus Linguistics and Translation Studies: Implications and Applications. In M. Baker, G. Francis & E. Tognini-Bonelli (Eds.), *Text and Technology: In Honour of John Sinclair* (pp. 233–250). John Benjamins. <http://dx.doi.org/10.1075/z.64.15bak>
- Baker, M. (1995). Corpora in Translation Studies: An Overview and Some Suggestions for Future Research. *Target International Journal of Translation Studies*, 7(2), 223–243. <https://doi.org/10.1075/target.7.2.03bak>
- Baker, M. (1996). Corpus-based Translation Studies: The Challenges that Lie Ahead. In H. Somers (Ed.), *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager* (pp. 175–186). John Benjamins.
- Baker, M. (2004). A Corpus-based View of Similarity and Difference in Translation. *International Journal of Corpus Linguistics*, 9(2), 167–193. <https://doi.org/10.1075/ijcl.9.2.02bak>
- Berber Sardinha, T. (2004). *Linguística de Corpus*. Manole.
- Bernardini, S. (2016). Discovery Learning in the Language-for-translation Classroom: Corpora as Learning Aids. *Cadernos de Tradução*, 36(1), 14–35. <https://doi.org/10.5007/2175-7968.2016v36nesplp14>
- Biber, D. (1993). Representativeness in Corpus Design. *Literary and Linguistic Computing*, 8(4), 243–257. <https://doi.org/10.1093/lc/8.4.243>
- Davies, M. (2012-2019). *Corpus do Português: NOW*. <http://www.corpusdoportugues.org/now/>
- Farkas, A. (2019). *LF Aligner* (4.21) [Software]. <https://sourceforge.net/projects/aligner/>
- Firth, J. (1957). A Synopsis of Linguistic Theory, 1930-1955. *Studies in Linguistic Analysis*, 10–32.
- Goldschmidt, M. A., Rodrigues, R. R., & Degasperi, M. H. (2017). O uso de corpora comparáveis bilíngues como subsídio para a tradução de colocações. *Belas Infêis*, 6(1), 9–24. <https://doi.org/10.26512/belasinfeis.v6.n1.2017.11416>
- Grosjean, F. (1989). Neurolinguists, Beware! The Bilingual is Not Two Monolinguals in One Person. *Brain and Language*, 36(1), 3–15. [https://doi.org/10.1016/0093-934x\(89\)90048-5](https://doi.org/10.1016/0093-934x(89)90048-5)
- Hansen-Schirra, S., Nitzke, J., & Oster, K. (2017). Predicting Cognate Translation. In S. Hansen-Schirra, O. Czulo & S. Hofmann (Eds.), *Empirical Modelling of Translation and Interpreting* (pp. 3–22). Language Science Press.
- Hausmann, F. J. (1989). Le dictionnaire de collocations. In F. J. Hausmann, O. Reichmann, H. E. Wiegand & L. Zgusta (Eds.), *Wörterbücher. Dictionaries. Dictionnaires: Ein internationales handbuch zur lexicographie. An International Encyclopedia of Lexicography. Encyclopédie internationale de lexicographie* (pp. 1010–1019). Walter de Gruyter.
- Hoey, M. (2005). *Lexical Priming: A New Theory of Words and Language*. Routledge.
- Jiménez-Crespo, M. A. (2013). Crowdsourcing, Corpus Use, and the Search for Translation Naturalness. *Translation and Interpreting Studies*, 8(1), 23–49. <https://doi.org/10.1075/tis.8.1.02jim>
- Kilgarrieff, A., Baisa, V., Bušta, J., Jakubiček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten Years On. *Lexicography*, 1(1), 7–36. <https://doi.org/10.1007/s40607-014-0009-9>
- Lamparelli, A. H. (2007). *A naturalidade na tradução: quem garante?* [PhD Dissertation]. Universidade de São Paulo. <https://doi.org/10.11606/d.8.2007.tde-04122007-102315>
- Leech, G. (2007). New Resources, or Just Better Old Ones? The Holy Grail of Representativeness. In M. Hundt, N. Nesselhauf & C. Biewer (Eds.), *Corpus Linguistics and the Web* (pp. 133–149). Rodopi. [https://doi.org/10.1163/9789401203791\\_009](https://doi.org/10.1163/9789401203791_009)

- López-Rodríguez, C. I. (2016). Using Corpora in Scientific and Technical Translation Training: Resources to Identify Conventionality and Promote Creativity. *Cadernos de Tradução*, 36(1), 88–120. <https://doi.org/10.5007/2175-7968.2016v36nesp1p88>
- Mauranen, A. (2007). Universal Tendencies in Translation. In G. Anderman & M. Rogers (Eds.), *Incorporating Corpora: The Linguist and the Translator* (pp. 32–48). <https://doi.org/10.2307/jj.27710861.8>
- Nesselhauf, N. (2003). The Use of Collocations by Advanced Learners of English and Some Implications for Teaching. *Applied Linguistics*, 24(2), 223–242. <https://doi.org/10.1093/applin/24.2.223>
- Nord, C. (2018). *Translating as a Purposeful Activity: Functionalist Approaches Explained*. 2nd ed. Routledge.
- Oliveira, B. M. (2017). Interferência e naturalidade no par português-espanhol: línguas próximas, contraste e ensino de tradução. *Caracol*, 14, 130–171. <https://doi.org/10.11606/issn.2317-9651.v0i14p130-171>
- Pastor, G. C. (1996). *Manual de fraseología española*. Gredos.
- Putranti, A. (2018). Modulation: A Translation Method to Obtain Naturalness in Target Language Texts. *Journal of Language and Literature*, 18(1), 98–101. <https://doi.org/10.24071/joll.v18i1.1115>
- Tagnin, S. E. O. (2005). O humor como quebra da convencionalidade. *Revista Brasileira de Lingüística Aplicada*, 5(1), 247–257. <https://doi.org/10.1590/s1984-63982005000100013>
- Tagnin, S. E. O. (2013). *O jeito que a gente diz: combinações consagradas em inglês e português*. Disal.
- Tagnin, S. E. O., & Teixeira, E. D. (2004). Lingüística de Corpus e Tradução Técnica: relato da montagem de um corpus multivarietal de culinária. *TradTerm*, 10, 313–358. <https://doi.org/10.11606/issn.2317-9511.tradterm.2004.47184>
- Woba. (2023). *Censo Coworking: uma análise Woba do mercado brasileiro*. Woba. <https://blog.woba.com.br/censo-coworking-2023/>

## Notes

## Authorship contribution

**Conceptualization:** P. H. Freitag

**Data collection:** P. H. Freitag

**Data analysis:** P. H. Freitag

**Results and discussion:** P. H. Freitag

**Supervision:** R. R. Rebechi

**Review and editing:** P. H. Freitag & R. R. Rebechi

## Research dataset

The research data analyzed in this paper is part of the doctoral dissertation defended by the first author at Universidade Federal do Rio Grande do Sul (UFRGS) in 2025. The data was collected and analyzed as part of the dissertation project. The full dissertation text has not yet been made available at the university's online repository.

## Funding

Not applicable.

## Image copyright

All images are screen captures of the specified software as indicated in their captions. The data displayed in the images is derived from the study corpora.

## Approval by ethics committee

Not applicable.



## Conflicts of interest

Not applicable.

## Data availability statement

The data from this research, which are not included in this work, may be made available by the author upon request.

## License

The authors grant *Cadernos de Tradução* exclusive rights for first publication, while simultaneously licensing the work under the Creative Commons Attribution ([CC BY](#)) 4.0 International License. This license enables third parties to remix, adapt, and create from the published work, while giving proper credit to the authors and acknowledging the initial publication in this journal. Authors are permitted to enter into additional agreements separately for the non-exclusive distribution of the published version of the work in this journal. This may include publishing it in an institutional repository, on a personal website, on academic social networks, publishing a translation, or republishing the work as a book chapter, all with due recognition of authorship and first publication in this journal.

## Publisher

*Cadernos de Tradução* is a publication of the Graduate Program in Translation Studies at the Federal University of Santa Catarina. The journal *Cadernos de Tradução* is hosted by the [Portal de Periódicos UFSC](#). The ideas expressed in this paper are the responsibility of its authors and do not necessarily represent the views of the editors or the university.

## Section editors

Andréia Guerini – Willian Moura

## Technical editing

Alice S. Rezende – Ingrid Bignardi – João G. P. Silveira – Kamila Oliveira

## Article history

Received: 26-10-2024

Approved: 01-03-2025

Revised: 03-05-2025

Published: 06-2025

