

ARTICLE

Lexical Features in the Portuguese Translations of Contemporary Political Discourse: A Corpus-Based Analysis

Wenying Luo ^{1,2} , Xiang Zhang ^{3*} 

¹ School of European Languages and Cultures, South China Business College of Guangdong University of Foreign Studies, Guangzhou 510545, China

² Department of Portuguese, University of Macau, Macao SAR 519000, China

³ Faculty of Languages and Translation, Macao Polytechnic University, Macao SAR 999078, China

ABSTRACT

The characteristics of translations result not only from the interaction between linguistic systems but also from inherent elements of the translated text. This study emphasizes the empirical analysis of translation within the Sino-Portuguese translation context, specifically focusing on contemporary political discourse. The objective of this research is to examine the lexical features in the Portuguese translation of representative contemporary political texts through a comparable Sino-Portuguese corpus. The corpus comprises source texts in Chinese and their corresponding Portuguese translations, totaling 127,071 tokens in Chinese, 204,064 in Portuguese, and a reference corpus with original political texts (185,959 tokens). The comparative analysis revealed distinct lexical differences and translation strategies. A reduction in lexical richness was observed in the Portuguese translation compared to the original Chinese text, suggesting a post-translation simplification, alongside a higher lexical density in the Chinese version. The presence of connectors was more noticeable in the translation, adapting logical relations from Chinese. Nominalized structures prevailed in the Portuguese translation, while the Chinese version emphasized verbs. Chinese modal verbs were systematically replaced with diverse grammatical resources in the Portuguese translation, avoiding repetitions and conforming to the language's conventions. Finally, translating institutional discourse necessitates a balance between semantic preservation and typological conformity, with implications for translator training and the analysis of transcultural discourses.

*CORRESPONDING AUTHOR:

Xiang Zhang, Faculty of Languages and Translation, Macao Polytechnic University, Macao SAR 999078, China; Email: zhangxiang@mpu.edu.mo

ARTICLE INFO

Received: 28 June 2025 | Revised: 8 July 2025 | Accepted: 15 July 2025 | Published Online: 19 August 2025

DOI: <https://doi.org/10.30564/fls.v7i8.10785>

CITATION

Luo, W., Zhang, X., 2025. Lexical Features in the Portuguese Translations of Contemporary Political Discourse: A Corpus-Based Analysis. *Forum for Linguistic Studies*. 7(8): 938–948. DOI: <https://doi.org/10.30564/fls.v7i8.10785>

COPYRIGHT

Copyright © 2025 by the author(s). Published by Bilingual Publishing Group. This is an open access article under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License (<https://creativecommons.org/licenses/by-nc/4.0/>).

Keywords: Chinese-Portuguese Translation; Lexical Features; Corpus; Political Discourse Translation

1. Introduction

This study aims to construct a comparable corpus of contemporary Chinese institutional discourse and its Portuguese translation, employing corpus analysis to explore the lexical characteristics of Chinese-Portuguese political text translation. The source texts comprise a representative collection of official documents and speeches originally published in 2020, spanning the period from late 2017 to early 2020. These texts reflect key concepts in public governance across 92 written and spoken works, including policy articulations and diplomatic communications. The Portuguese translation, produced by professional translators at an authoritative publishing house, was released in 2021.

In the realm of political texts, systematic studies on translational lexical characteristics in Chinese-Portuguese remain scarce, particularly those employing integrated quantitative-qualitative methodologies to conduct three-dimensional analysis encompassing lexical frequency, vocabulary diversity, and semantic-pragmatic differences. The unique nature of this genre – with its distinctive linguistic style and strict requirements for translational accuracy and normativity — makes contrastive analysis relevant. Such an investigation not only unveils lexical usage patterns in bilingual political contexts but also provides theoretical insights for translational practice in this specialized domain. Specifically, this research will focus on lexical frequency, vocabulary diversity, and semantic-pragmatic differences of specific terms in political texts of both languages. Through contrastive analysis, the aim is to reveal the linguistic strategies and characteristics employed in Chinese and Portuguese to express political concepts and perspectives. Key questions include: Are there significant lexical differences between the original Chinese and the Portuguese translation? What factors determine these disparities? What implications do these lexical particularities bring to translational practice?

The article's structure comprises six parts: 1) Introduction and contextualization of the research; 2), 3) Theoretical review of corpus-based translation studies, with a special focus on political texts, establishing an analytical framework associating lexical characteristics with transla-

tional universals and the translator's style; 4) Presentation of corpus construction principles, and application of a mixed method integrating lexical metrics, syntactic complexity, and keyword analysis; 5) The analytical section, which systematically investigates six dimensions: 5.1) lexical diversity and 5.3) density unveiling domestication trends; 5.2) vocabulary length distribution; 5.4) analysis of frequent words and 5.5) key terms, with a special focus on the Portuguese modal verb “dever” (must/should) and the use of Portuguese personal pronoun in the first-person plural “nós” (we/us); 6) Conclusions.

2. Corpus-Based Translation Studies

The methods and techniques of corpus linguistics have had a profound impact on translation studies, particularly in theoretical descriptions and applied research^[1]. Baker emphasizes that corpora like the Translational English Corpus (TEC) enable empirical investigation of translators' linguistic behavior across variables such as source language, genre, and individual style, moving beyond abstract theorizing^[2]. The author further emphasizes that the construction of large corpora provides researchers with unique opportunities to observe and analyze translational phenomena, proposing three categories of corpora for translation studies: parallel corpus, comparable corpus, and multilingual corpus. Based on comparable corpora, Baker identified universal features such as explicitation, simplification, and preference for conventional 'grammaticality'^[3]. Subsequently, scholars like Laviosa validated these hypotheses by comparing lexical usage patterns between original and translated texts in English, arguing that “translation universals” can be influenced by language pairs and text types. This data-driven approach established a crucial methodological pathway for corpus-based translation studies^[4].

Analyzing translations from German to English in comparable corpora, Kenny identified a tendency towards the preference for conventional grammaticality in the use of creative vocabulary by translators, although the style of specific authors can still be preserved^[5]. This suggests that the act of translation is simultaneously conditioned by norms and

marked by the translator's subjectivity^[6]. Carmen Dayrell's research highlights the relevance of corpora in translation studies, discussing not only their types but also their theoretical and pedagogical applications^[7]. Dayrell operationalizes the framework in her study of Brazilian Portuguese, designing a comparable corpus (translated vs. non-translated texts) to test simplification through collocational diversity. Her findings show that translated texts exhibit reduced collocational range (fewer unique collocates per node) and stronger reliance on high-frequency collocates, supporting Baker's simplification hypothesis^[3, 8].

Within the scope of Chinese-Portuguese translation corpora, Wang and Leal's study demonstrated, through concrete cases, the application of parallel corpora in computer-assisted translation, proving its value in enhancing translational efficiency and quality^[9]. Meanwhile, Hu and Wang's research, based on a Chinese-Portuguese parallel corpus, analyzed translation patterns for the Chinese negative structure “不 bù / 没 méi + verb” in Portuguese, revealing fundamental transfer patterns of negative structures between the two languages^[10].

In corpus-based translation studies, the analysis of lexical characteristics is prevalent. Laviosa noted a systematic trend towards lexical simplification and paradoxical syntactic complexity in translated texts when comparing corpora of original and translated English texts^[4]. Analyzing a Chinese-English literary parallel corpus, Hu observed that translations exhibit less lexical variance, lower lexical density, a higher prevalence of high-frequency words, decreased use of concrete nouns, and excessive repetition of common terms compared to originals^[11]. The construction of large-scale parallel corpora like UM-Corpus further enables quantitative lexical analysis across domains, revealing significant disparities in vocabulary size between English and Chinese, with Chinese translations exhibiting higher lexical repetition rates^[12].

Utilizing Systemic Functional Linguistics (SLF) approaches and corpus analysis, Fleuri et al. investigated the frequency and role of “translator” and “tradutor” (translator) in English-Portuguese texts, identifying similarities in usage patterns between the terms across different genres^[13]. Similarly, based on a parallel corpus, Cruz studied the translation of specialized terminology in the export and agriculture sectors, emphasizing the relevance of lexical analysis for

technical translations^[14]. Conversely, Zidan and Yagi analyzed the translation of the pragmatic marker “我认为” wǒ rènwéi (I believe) in 19th-century English literature through a parallel corpus, revealing semantic-functional divergences between distinct linguistic and cultural contexts^[15].

3. Corpus-Based Translation Studies of Political Texts

In the domain of political text translation studies, the analysis of lexical characteristics holds significant importance. Leandro Parra Rodrigues Cruz's study illustrated the use of parallel corpora to analyze the translation of specialized terminology in specific domains. It was found that political texts in English and Portuguese share frequent use of connectives as a striking feature, with Portuguese particularly reinforcing logical explicitation through increased connectives, reflecting differences in linguistic demands of argumentative texts, especially in the translation process^[14]. These findings align with Nunes' corpus-based analysis of the Corpus de Língua Portuguesa em Tradução, which revealed statistically significant overuse of conjunctions (e.g., *mas*, **e**, *porém*) in political speeches (*discurso político*) and reviews (*resenha*)—both argumentative genres—compared to other text types. This corroborates the hypothesis that argumentative texts explicitly mark logical connections through cohesive devices^[16]. These findings underscore the value of corpus methodologies in translation studies.

Chen and Cui, in their analysis of lexical characteristics of Government Work Reports, identified in English translations: higher average vocabulary extent, elevated lexical density, accentuated textual complexity, limited variation in word usage, predominance of nouns among high-frequency terms, and significant occurrence of preference for conventional grammaticality^[17]. Similarly, Yu's research on lexical characteristics in governmental documents pointed out, in English translations, the coexistence of linguistic complexification and simplification trends^[18]. Li and Pan's study employed a corpus-based methodology to examine the operational mechanisms of ideological manipulation in political translation^[19]. Huang's study demonstrates that while the use of subject personal pronouns exhibits a declining trend in English-Chinese translation, their frequency remains significantly higher than in native Chinese texts, attributable to the

dual causation of source language transfer and comparable explicitation^[20]. A parallel perspective is validated in Pan, Fu, and Li's analysis of CPC Reports translations (1990s–2010s), which reveals systematic micro-level variations in personal pronouns. The study observes a rising trend in first-person plural (“we”) and second-person pronouns within English translations, alongside a semantic shift in collocational patterns: “we + modal verbs” increasingly encodes inclination rather than obligation. These lexical alterations construct more symmetrical reader relationships, reflecting ideological evolution in translators' mediation strategies^[21].

From these studies, it is observed that corpus-based translation studies have revolutionized research by offering insights into translational phenomena through the analysis of large text collections. From universal features to translation patterns, this approach provides a robust framework for understanding translation processes. The findings on lexical characteristics, stylistic preferences, and specialized terminologies underscore the importance of data-driven analysis. Transitioning from these insights, the upcoming section will explore the methodological aspects.

4. Methodology

4.1. Building a Corpus

The corpus construction process followed established methodologies in corpus linguistics^[22, 23], utilizing WordSmith Tools 9.0 and the corpus software AntConc 4.2.4 to analyze and explore a self-built corpus. Firstly, the original Chinese version of 92 political texts was collected and identified, containing 127,071 tokens (abbreviated as PDCh), along with the Portuguese translation, consisting of 204,064 tokens (abbreviated as PDPt). Additionally, as a comparative reference, 90 official speeches by the Brazilian president from October 2017 to February 2020 were collected. The texts of these speeches were collected from the “Discurso” (Speeches) section of the Biblioteca da Presidência da República (Library of the Presidency of the Republic) website, forming a corpus of Portuguese speeches with 185,959 tokens (abbreviated as DP). All three datasets are official texts and were stored in Word document format.

Subsequently, WPS and ROST CM6 were used to clean and segment the texts. Excessive blank lines and illegible characters were manually removed from all texts and saved

as .txt documents. Following the cleaning process, TreeTagger was utilized for morphosyntactic tagging.

Finally, WordSmith Tools 9.0 was used to identify full word density statistics and generate keyword lists, while AntConc 4.2.4 was used to observe keyword patterns.

4.2. Data Analysis Methods

This study utilizes WordSmith and AntConc to conduct quantitative and qualitative analyses of the lexical characteristics of political texts translated into Portuguese. The investigation focuses on two main aspects: firstly, comparing the original text with its translation, using PDPt as the observed corpus and PDCh as the reference corpus; secondly, comparing texts in the same language (Portuguese), i.e., the translated text (PDPt) with original texts (DP), using PDPt as the observed corpus and DP as the reference corpus. The following parameters are compared: Standardized Type-Token Ratio (STTR), lexical density, average word length, average sentence length, and keywords.

This mixed methodological approach promises a detailed and insightful analysis of the lexical nuances present in political texts across languages, shedding light on translation dynamics and linguistic adaptations within this specialized domain.

5. Results and Analysis

5.1. Lexical Diversity

A token refers to the total number of word occurrences in a corpus, while a type represents unique and distinct lexical forms. Lexical diversity is measured by the Type-Token Ratio (TTR) and the Standardized Type-Token Ratio (STTR), where higher values indicate greater vocabulary richness, and lower values suggest less diversity^[14, 15]. As can be seen from the obtained data (**Table 1**), the TTR (8%) and STTR (48.34%) of PDCh are significantly higher than those of PDPt (TTR = 7%; STTR = 43.83%) and DP (TTR = 6%; STTR = 42.57%). The standard deviation of STTR indicates that the lexical distribution fluctuation in PDCh is less than that of the PDPt and DP corpora.

This demonstrates that the original text in Chinese has a greater vocabulary richness than the translation, with an increase in lexical repetition in the translated text, aligning with

the simplification described by Baker and Laviosa in the translation process—repetition of high-frequency words or the use of generic expressions to reduce complexity^[2, 3]. The higher

standard deviation in both translated and Portuguese reference texts suggests significant fragmentary differences in lexical usage between political texts in Portuguese and Chinese.

Table 1. Comparison Among the Corpora PDCh, PDPt, and DP (Words).

	PDCh	PDPt	DP
Type	10,231	14,794	11,600
Token	127,064	202,639	185,959
TTR	8%	7%	6%
STTR	48.34%	43.83%	42.57%
STTR std.dev	52.3	56.36	56.62
Mean Word Length (in Characters)	1.81	5.23	4.84

5.2. Word Length

The average word length in PDCh is 1.81 (counting the number of Chinese characters), which aligns with the characteristic of the Chinese vocabulary predominantly comprising words of one or two characters (such as “经济” jīngjì (economics), a two-character word). The average word length in PDPt significantly increases to 5.23, surpassing the control text (DP, 4.84).

Figure 1 reflects how the average word length is related to the typological characteristics of the language and lexical choice strategies in the translation process. This difference originates from two mechanisms: firstly, the inflectional morphology of Portuguese (such as verb conjugation and noun

pluralization) naturally leads to an increase in word length; for example, the Chinese expression “深化改革” (shēnhuà gǎigé, deepen reforms) will be translated in Portuguese as a conjugated verb (e.g., “aprofundamos a reforma”). Secondly, to preserve the semantic integrity of the original text during translation, compounds or explanatory phrases are used, such as the high-frequency word “可持续发展” (kě chíxù fāzhǎn, sustainable development), further increasing the vocabulary length. The word length in the control text (DP), which is 4.84, reflects the characteristics of natural Portuguese vocabulary. From the perspective of word length, translation tends to become more complex, also confirming the coexisting phenomenon of simplification and complexification observed in Yu’s study^[14].

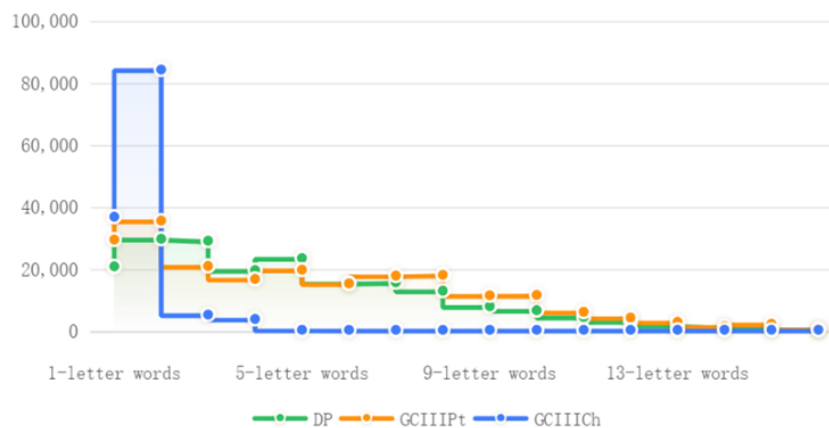


Figure 1. Word Length of the Corpora PDCh, PDPt, and DP.

5.3. Lexical Density

As a comparative reference, when quantifying differences in lexical density and distribution of grammatical

classes between PDPt and DP, it was also found that there are divergences in the information encoding mechanisms between Chinese political texts translated into Portuguese and original Portuguese political texts. As shown in **Ta-**

ble 2, in terms of lexical density, DP presents a significantly higher proportion of full words (69%) compared to the translation (62%), reflecting a phenomenon of lexical dilution in translated texts when compared to the originals. This phenomenon can be attributed to an interlinguistic mediation mechanism — that is, the need to reconstruct implicit paratactic relationships between Chinese phrases through the addition of functional words, such as the connector “e” (and) in Portuguese.

Table 2. Comparison between the Corpora PDPt and DP.

	PDPt	DP
Lexical Density	62%	69%
Nouns	29%	25%
Verbs	15%	19%

In terms of the distribution of grammatical classes, the nominalization trend in the translation (proportion of nouns: 29%) and the characteristic of verbal reinforcement in the original texts (proportion of verbs: 19%) form a striking difference. The high-frequency nominalized structures in PDPt (for example, “深化改革”, *shēnhuà gǎigé*, translated as “aprofundamento da reforma” in Portuguese) reflect, on the one hand, typological differences between Chinese and Portuguese leading to translational transfer — the conver-

sion of typical Chinese verb-object structures into nominal phrases in Portuguese; on the other hand, the strategy of de-agentization in political discourse translation — transforming dynamic processes into static concepts to reinforce the objective inevitability of policy implementation.

5.4. High-Frequency Words

To better unveil the linguistic characteristics of the Portuguese translation, a comparison was made between the 15 most frequent words in PDCh and PDPt (**Table 3**), identifying stark differences in expressive modes between Chinese and Portuguese. In the Chinese text, the most common functional words are: 1) “的” (Portuguese translation: *de*) (5786 occurrences), used as a connector between modifier and nominal core in structures of long modification (e.g., “发展的目的”, Portuguese translation: “objetivo *do* desenvolvimento”); 2) “和” (Portuguese translation: *e*) (5786 occurrences), utilized in coordinated clauses (e.g., “理论和实践”, Portuguese translation: “teoria *e* prática”), reinforcing relationships through coordinated repetition; 3) “了” (687 occurrences), an aspectual particle indicating action completion while simultaneously emphasizing results (e.g., “取得了历史性成就”, Portuguese translation: “*tendo obtido* êxitos históricos”).

Table 3. List of High-Frequency Words.

N	Type	Freq	Type	Freq	N	Type	Freq	Type	Freq
1	的	5786	e	12046	16	政治	601	na	1420
2	和	2259	de	9872	17	制度	600	no	1246
3	要	1735	a	9073	18	改革	600	se	1243
4	是	1416	o	5665	19	为	583	desenvolvimento	1223
5	发展	1396	do	4583	20	社会主义	561	como	1013
6	中国	1107	da	3281	21	经济	507	devemos	962
7	在	1004	os	3162	22	年	481	ao	958
8	人民	942	para	2501	23	对	475	é	957
9	我们	858	as	2459	24	工作	467	dos	912
10	党	817	em	2349	25	把	465	das	872
11	国家	809	que	2103	26	问题	463	mais	835
12	建设	736	com	1752	27	上	447	à	782
13	坚持	716	um	1609	28	推进	447	povo	765
14	新	687	uma	1576	29	全面	444	nos	720
15	了	648	partido	1430	30	不	424	sistema	692

In the Portuguese translations, high-frequency words are predominantly prepositions and articles such as “de” (9872) and “a” (9073), primarily used to mark grammatical relations. Although the Portuguese conjunction “e”

(equivalent to Chinese “和”) appears 12,046 times—more than five times the frequency observed in PDCh—, this reflects an adaptation to the phrasal structure of Portuguese: in translation, Chinese structures of emphatic repetition are

streamlined, and connectors are added for grammatical conformity (e.g., “勇于变革、勇于创新”, Portuguese translation: “atrever-se a buscar transformações e inovações”).

This contrast highlights two ways of conveying information in the two languages: Chinese communicates information through short repeated phrases and coordinated terms, resulting in concise yet redundant sentences; whereas Portuguese necessitates prepositions and articles to construct relatively complex sentences. This difference illustrates the translational balance between preserving the original meaning and adhering to the linguistic rules of the target language.

5.5. Words Keyword Analysis

The full words in the corpus are directly related to the text’s theme, influenced and constrained by the original content^[3, 11, 14]. Therefore, this study will not analyze the specific lexical collocation characteristics of the Portuguese translation. Among other keywords, there is a notably high frequency of the Portuguese modal verb “dever” (must/should) conjugated in the first-person plural “nós” (4613 occurrences). The following section takes this as a starting point, focusing on the Portuguese modal verb “dever” and its usage in the first-person plural.

5.5.1. Modal Verb “Dever”

Neves emphasizes that the deontic modality of “dever” encompasses obligation and command, whereas the epistemic modality expresses simple necessity^[24]. **Table 4** below presents the frequency of occurrence of the modal verb “dever” in PDPt, as calculated through AntConc.

Table 4. Frequency of Modal Verb “Dever” in the Corpus PDPt.

Type	Freq	Type	Freq
devem	439	deva	2
devemos	343	deveremos	2
deve	184	deverá	1
deveres	23	devia	1
dever	4		
Total			999

For high-obligation modal verbs in Chinese — “必须 (bì xū)” (must/have to), “要 (yào)” (need to), and “应 (yīng)” (should/shall/ought to) — direct translation strategies were employed. For example:

Original text:

全党同志/一定/要/永远/与/人民/同呼吸/、/共命运/、/

心连心,[...].

Pin yin:

Quán dǎng tóngzhì yīdìng yào yǒngyuǎn yǔ rénmin tóng hūxī、gòng mìngyùn、xīn liánxīn,[...].

Translation in PDPt:

Todos os camaradas do PCCh devem sempre respirar o mesmo ar com o povo, compartilhar com ele a mesma sorte, manter uma afinidade mútua[...].

However, compared to the total frequency of “要 (yào)” (Freq = 1735), “必须 (bì xū)” (Freq = 305), and “应 (yīng)” (Freq = 11), it is evident that translators employed multiple strategies to avoid excessive repetitions of the modal verb “dever”. For example:

a) Nominalization

Original text:

新时代/要/有/新气象/、/更/要/有/新作为。

Pin yin:

Xīn shídài yào yǒu xīn qìxiàng, gèng yào yǒu xīn zuòwéi.

Translation in PDPt:

Nova era, novas iniciativas e novas conquistas.

Here, the modal verb “要 (yào)” is omitted: instead of translating “要有 (yào yǒu, translated as “must have”)” or “更要有 (gèng yào yǒu, translated as “should more importantly have”)” literally, phrases of coordinated nominal groupings are chosen (novas iniciativas, novas conquistas, translated as “new initiatives, new achievements”). Additionally, a nominal conversion is performed, transforming the dynamic expression “有新作为 (yǒu xīn zuòwéi, translated as “make new achievements”)” (ter novas ações) into a static noun (novas conquistas, translated as “new achievements”), reducing agentivity and focusing on the result.

b) Conversion to neutral expressions

Original text:

要/加强/国家安全系统/党的建设 [...].

Pin yin:

Yào jiāqiáng guójiā ānquán xìtǒng de dǎng de jiànshè[...].

Translation in PDPt:

É necessário fortalecer o Partido e seus trabalhos nos departamentos de segurança nacional[...].

The imperative “要 (yào)” is converted into the neutral expression “É necessário” (it is necessary). This conversion operates in two ways: attenuating the direct command into a

statement of necessity, reducing the authoritative tone, and maintaining the verbal core (“fortalecer”) in a neutral grammatical structure, balancing the register. This dual adaptation preserves the original political strength while aligning with the Portuguese rhetorical tradition, which favors objective argumentation in official texts.

c) Modal Explication

Original text:

越是/环境/复杂/, /我们/越是/要/坚持/改革开放/不动摇/, /绝/不会/回到/关起门来/搞/建设/的/老路/上去/。

Pin yin:

Yuè shì huánjìng fùzá, wǒmen yuè shì yào jiānchí gǎigé kāifàng bù dòngyáo, jué bù huì huídào guān qǐ mén lái gǎo jiànshè de lǎolù shàng qù.

Translation in PDPt:

Quanto mais complexa for a situação, mais firmes seremos na reforma e abertura. Nunca voltaremos, de jeito nenhum, ao velho caminho de desenvolvimento a portas fechadas.

The translation adapts the meaning through multiple grammatical resources. It explicates the obligation through the future tense “seremos” (we will be) + the adjective “firmes” (firm), maintaining the decisive intensity of the original and temporally marking the action. Furthermore, it replicates the logical structure “越 (yuè)... 越 (yuè)...” (quanto mais... mais..., translated as “the more... the more...”) to reinforce political inevitability. Additionally, it enhances syntactic dynamism with prepositional phrases (“na reforma”), transforming static verbs (“坚持 (jiānchí)”, translated as “to insist”) into continuous actions (“firmes na reforma”, translated as “firm in the reform”). This stratification of strategies transplants the semantic authority of Chinese while optimizing the acceptability of the discourse through the grammatical resources of Portuguese.

5.5.2. The Expression “Nós (We/Us)” in the Portuguese Translation

From **Table 5**, it can be observed that the expression “我们 (wǒmen)” in Chinese and “Nós” in Portuguese present significant quantitative differences across the corpora. There were 858 occurrences in the PDCh corpus, 4613 occurrences in the PDPt corpus, and 5202 occurrences in the control corpus DP.

Table 5. Frequency of “Nós” in the Corpora.

Corpus	Form(s)	Frequency
PDCh	我们	858
PDPt	Verb(-mos), nosso(s), nossa(s)	4613
DP	Verb(-mos), nosso(s), nossa(s)	5202

PDCh extensively uses structures without explicit subjects, implying the collective agent through context (as in “persistir no enfoque centrado no povo”, translated as “persisting in the people-centered approach”). This discursive pattern is rooted in the paratactic characteristics of Chinese grammar and the collectivist tradition of rhetorical construction. On the other hand, PDPt, constrained by the morphosyntactic rules of the target language, requires the retrieval of the subject through verbal inflections (such as the first-person plural suffix “-mos”) or explicit pronouns (“nós” - we/us), resulting in increased frequency of these markers in the translation. For example:

Original text:

脱贫/攻坚/力度/之/大/, /规模/之/广/, /影响/之/深/, /前所未有的/, /取得/了/决定性/进展/, /显著/改善/了/贫困/地区/和/贫苦/群众/生产/生活/条件/, /谱写/了/人类/反贫困/历史/新篇章/。

Pin yin:

Tuōpín gōngjiān lìdù zhī dà, guīmó zhī guǎng, yǐngxiǎng zhī shēn, qiánsuǒwèiyǒu, qǔdé le juédìngxìng jìnzhǎn, xiǎnzhù gǎishàn le pínkùn dìqū hé pínkǔ qúnzhòng shēngchǎn shēnghuó tiáojiàn, pǔxiě le rénlei fǎn pínkùn lìshǐ xīn piānzhāng.

Translation in PDPt:

Trata-se de uma batalha sem precedentes em termos de intensidade, escala e influência. Mas conseguimos progressos decisivos, melhorando significativamente as condições de trabalho e de vida das áreas pobres e da população necessitada e escrevendo um novo capítulo na história do combate à pobreza.

In PDCh, subjectless clauses are predominant. In the analyzed sentence, the outcome is stated directly (“取得了决定性进展 (qǔdé le juédìngxìng jìnzhǎn)”, Portuguese translation: (nós) conseguimos progressos decisivos, English translation: we achieved decisive progress), with the subject (like “nós” - we or “China” implicit in the context). In the PDPt translation, the verbal inflection (“conseguimos”,

translated as “we managed”) marks the collective agent, complying with both the grammatical norms of Portuguese and restructuring the rhetorical framework of accountability attribution. Comparisons with native Portuguese political texts (DP) reveal that, although they also employ the first-person plural with high frequency, the translation of Chinese texts shows significantly higher usage rates than DP, indicating a phenomenon of hyperexplicitation^[14]. This highlights that the explicitation of the subject in Chinese-Portuguese political translation constitutes a process of double adaptation: on one hand, it adheres to mandatory transformations arising from typological differences between languages; on the other hand, it involves strategic rhetorical choices^[2, 4, 14].

6. Conclusions

The findings of this study, based on the analysis of a Chinese-Portuguese parallel corpus and a reference corpus of original Portuguese political discourses (DP), revealed significant lexical patterns and translational strategies in the translation of political texts. Firstly, it was observed that the original Chinese text (PDCh) exhibited a higher lexical richness (TTR = 8%, STTR = 48.34%) compared to the Portuguese translation (PDPT, TTR = 7%, STTR = 43.83%) and original Portuguese texts (DP, TTR = 6%, STTR = 42.57%), indicating a tendency towards lexical simplification in translations, as theorized by Baker^[2] and Laviosa^[3].

In terms of lexical density, the original Portuguese texts (DP) stood out with 69% full words, surpassing the translation (PDPT, 62%), suggesting an informational dilution in translated texts attributed to the insertion of connectives (e.g., “e”) to articulate implicit logical relations from Chinese. In the distribution of grammatical classes, the translation favored nominalized structures (29% nouns), converting Chinese verb-object constructions into static phrases (e.g., “aprofundamento da reforma” – “deepening of the reform”), while the original Portuguese texts emphasized verbs (19%), creating action-oriented rhetorical chains.

The analysis of high-frequency words revealed deep contrasts: Chinese relied on functional particles (“的”, “和”, “了”) for conciseness and emphasis through repetition, while Portuguese depended on prepositions (“de”, “a”) and articles for syntactic complexity. Additionally, translators adopted multifaceted strategies to avoid overloading the modal verb

“dever”, including nominalization, conversion to neutral expressions (“É necessário”, “it is necessary”), and modal explicitation through verb tenses and comparative structures (“quanto mais... mais...”, “the more... the more...”).

This study, through a systematic comparison between original Chinese political texts (PDCh), their Portuguese translations (PDPT), and native Portuguese political texts (DP), reveals dynamic patterns of linguistic modulation in the translation of political documents. The findings demonstrate a “double adaptive adjustment” in the translation process: on one hand, Chinese paratactic grammar, repetitive high-frequency structures, and subject-less clause patterns are inevitably reconstructed in Portuguese as long hypotactic periods, explicit connections, and subject reinforcement; on the other hand, to balance political expression authority with target language acceptability, translations adopt adaptation strategies, meaning that the directive intensity is attenuated through nominalization, and action orientation is preserved through temporal explicitness and complementing logical relations with functional words. These strategies reflect both the restrictions of Sino-Portuguese typological differences and active rhetorical choices in the intercultural dissemination of political discourse.

The research further validates the coexistence of simplification and complexification in translation: at the lexical level, translated texts show a simplifying trend due to the repetition of frequent words and vocabulary dilution (e.g.: STTR values drop by approximately 4.5% compared to the originals); whereas at the syntactic level, to preserve the Chinese structure while maintaining the informative density of short sentences, translations generate complexification through the chaining of subordinate clauses (average sentence length of 57.77) and nominalization (nouns, 29%). This essential contradiction essentially manifests the function of translation as a “cultural regulator”, seeking a dynamic balance between fidelity to the source content and conformity to target language habits.

This research contributed to corpus-based translation studies by offering empirical validation and nuanced insights into the dynamics of simplification and complexification within political translation, specifically within the underexplored Chinese-Portuguese language pair. It goes beyond mere confirmation, elucidating the linguistic mechanisms behind these phenomena through the concept of “double adap-

tive adjustment". The study enhances theory development by highlighting translators' strategic choices in navigating linguistic differences and rhetorical demands, emphasizing their active role in negotiating meaning beyond norm adherence. Additionally, it provides detailed insights into the lexical features of translated Chinese political texts into Portuguese, filling a critical gap in the literature.

Acknowledging its contributions, the study also recognized certain limitations. The scope of the corpus, primarily focused on political speech and specific sub-corpora, may benefit from broader representation across various Portuguese political text genres to enhance generalizability. While the analysis highlights essential lexico-grammatical patterns^[25], future research could integrate higher-level analyses to incorporate pragmatic, discursive, and ideological dimensions shaping political translation. Furthermore, individual translator styles and decision-making processes warrant further exploration, as does a reception-focused perspective to assess how these translations are perceived by Portuguese readers. The study suggested promising future research directions to address these limitations, emphasizing expanded corpus studies, multidimensional analyses integrating discourse frameworks, reception studies, translator cognition insights, comparative genre analysis, and resource development for Chinese-Portuguese political translation. These efforts promise to deepen our understanding of intercultural political communication through the lens of translation dynamics.

Author Contributions

Conceptualization, W.L. and X.Z.; methodology, W.L. and X.Z.; software, W.L. and X.Z.; validation, W.L. and X.Z.; formal analysis, W.L. and X.Z.; investigation, W.L. and X.Z.; resources, W.L. and X.Z.; data curation, W.L. and X.Z.; writing—original draft preparation, W.L. and X.Z.; writing—review and editing, W.L. and X.Z.; visualization, W.L. and X.Z.; supervision, X.Z.; project administration, X.Z.; funding acquisition, X.Z. All authors have read and agreed to the published version of the manuscript.

Funding

This work is supported by the research fund of Macao Polytechnic University (Research Project No. RP/ESLT-

02/2021).

Institutional Review Board Statement

Not applicable. This study did not involve human or animal participants but analyzed publicly available textual data.

Informed Consent Statement

Not applicable. This study analyzed publicly available political discourse and did not involve human participants or personal data requiring consent.

Data Availability Statement

The political texts analyzed in this study are sourced from publicly available transcripts and official translations. Extracted lexical datasets and annotation guidelines are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflict of interest.

References

- [1] Baker, M., 1993. Corpus Linguistics and Translation Studies: Implications and Applications. In: Baker, M., Francis, G., Tognini-Bonelli, E. (eds.). *Text and Technology: In Honour of John Sinclair*. John Benjamins Publishing Company: Amsterdam, Netherlands. pp. 233–250. DOI: <https://doi.org/10.1075/z.64.15bak>
- [2] Baker, M., 1999. The role of corpora in investigating the linguistic behaviour of professional translators. *International Journal of Corpus Linguistics*. 4(2), 281–298. DOI: <https://doi.org/10.1075/ijcl.4.2.05bak>
- [3] Baker, M., 1995. Corpora in translation studies: an overview and some suggestions for future research. *International Journal of Translation Studies*. 7(2), 223–243. DOI: <https://doi.org/10.1075/target.7.2.03bak>
- [4] Laviosa, S., 1998. Corpus-based Translation Studies: Theory, Findings, Applications. In: Brill: Leiden, Netherlands. DOI: https://doi.org/10.1163/9789004485907_005
- [5] Kenny, D., 2001. *Lexis and Creativity in Translation: A Corpus-based Study*. St. Jerome Publishing: Manchester, UK.
- [6] Baker, M., Somers, H., 1996. *Corpus-based transla-*

- tion studies: the challenges that lie ahead. In: Somers, H. (ed.). *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager*, Vol. 18. John Benjamins Publishing Company: Amsterdam, Netherlands. pp. 175–186. DOI: <https://doi.org/10.1075/btl.18.17bak>
- [7] Dayrell, C., 2015. The use of corpora for translation studies. *Translation Journal*. 2022(34). DOI: <https://doi.org/10.17771/pucrio.tradrev.25346> (in Portuguese)
- [8] Dayrell, C., 2007. A quantitative approach to compare collocational patterns in translated and non-translated texts. *International Journal of Corpus Linguistics*. 12(3), 375–414. DOI: <https://doi.org/10.1075/ijcl.12.3.04day>
- [9] Wang, T., Leal, A., 2020. The use of parallel corpora as a resource in computer-assisted translation: A case study of Portuguese-Chinese judicial sentence translation. *Letters Today*. 55(4). DOI: <https://doi.org/10.15448/1984-7726.2020.4.38339> (in Portuguese)
- [10] Hu, Z.H., Wang, S.Y., 2023. On the Translation of the Chinese “Bu (Mei) + Verb” Structure into Portuguese Based on a Chinese-Portuguese Parallel Corpus. *Translation Notebooks*. 43(3), 233–271. DOI: <https://doi.org/10.5007/2175-7968.2023.e87041> (in Portuguese)
- [11] Hu, X.Y., 2007. A Corpus-Based Study of Lexical Features in Translated Chinese Fiction. *Foreign Language Teaching and Research*. 39(3), 214–220. DOI: <https://doi.org/10.3969/j.issn.1000-0429.2007.03.008> (in Chinese)
- [12] Tian, L., Wong, D.F., Chao, L.S., et al., 2014. UM-Corpus: a large English-Chinese parallel corpus for statistical machine translation. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, Reykjavik, Iceland, May 2014; pp. 1837–1842.
- [13] Fleuri, L., Vasconcellos, M.L., Pagano, A., 2009. The Representation of the “Translator” Participant in Translators Through History and Os Tradutores na História. *Translation Notebooks*. 2(24), 159–191. DOI: <https://doi.org/10.5007/2175-7968.2009v2n24p159> (in Portuguese)
- [14] Cruz, L.P.R., 2014. A Study on the Translation of Specialized Collocations in the Export and Agribusiness Sectors Based on a Parallel Corpus [Master’s thesis]. São Paulo State University: São José do Rio Preto, Brazil. p. 142. (in Portuguese)
- [15] Zidan, B.A., Yagi, S.M., 2025. A parallel corpus analysis of the pragmatic marker “I think”. *Theory and Practice in Language Studies*. 15(3), 706–715. DOI: <https://doi.org/10.17507/tpsls.1503.05>
- [16] Nunes, L.P., 2020. Cohesive devices in the Portuguese Language Corpus in Translation: Investigating conjunctions in a contrastive perspective within a text typology. *Journal of Language Studies*. 28(1), 13–40. DOI: <https://doi.org/10.17851/2237-2083.28.1.13-40> (in Portuguese)
- [17] Chen, J.S., Cui, Y.N., 2010. A Corpus-Based Study of Lexical Features in the English Translation of China’s “Government Work Report”. *Contemporary Foreign Languages Studies*. (6), 6. (in Chinese)
- [18] Yu, H., 2016. A Corpus-Based Investigation of the “Simplification” Trend in Government Document Translation: A Case Study of the English Version of the White Paper “China’s National Defense in 2010”. *Foreign Languages Research*. 33(3), 79–86. (in Chinese)
- [19] Li, T., Pan, F., 2021. Reshaping China’s image: a corpus-based analysis of the English translation of Chinese political discourse. *Perspectives*. 29(3), 354–370. DOI: <https://doi.org/10.1080/0907676x.2020.1727540>
- [20] Huang, L., 2008. Explication of Subject Personal Pronouns in English-Chinese Translation: A Corpus-Based Study. *Foreign Language Teaching and Research*. 40(6), 454–459. (in Chinese)
- [21] Pan, F., Fu, Y., Li, T., 2024. Institutional translators’ mediation of CPC Work Reports diachronically through personal pronouns: a corpus-based discourse analysis approach. *Perspectives Studies in Translatology*. 32(5), 924–941. DOI: <https://doi.org/10.1080/0907676X.2023.2194548>
- [22] Xu, J., 2023. *Research methods in corpus linguistics*. Foreign Language Teaching and Research Press: Beijing, China.
- [23] Scott, M., 1998. *Manual of Wordsmith (version 3)*. Oxford University Press: Oxford, UK.
- [24] Neves, M.H.D.M., 2000. The Polysemy of Modal Verbs. Or: On Ambiguities. Available from: <http://hdl.handle.net/11449/107772> (cited 10 March 2025). (in Portuguese)
- [25] Raeisi, M., Dastjerdi, H.V., 2019. Lexico-grammatical analysis of native and nonnative abstracts based on Halliday’s SFL model. *Theory and Practice in Language Studies*. 9(11), 1388–1395. DOI: <http://dx.doi.org/10.17507/tpsls.0911.03>