

How Translation Shapes Modern Chinese? Corpus-Based Evidence and Theoretical Explanations from the Perspective of Language Contact

Feiyang Sun

School of translation, University of South China, Hengyang 421000, China

Abstract: Translation has long been recognized as a major site of language contact, yet the mechanisms through which translated texts influence the development of modern Chinese remain underexplored. Drawing on a synthesis of corpus-based translation studies over the past three decades, this paper examines how translation has shaped modern Chinese from the perspective of language contact. Evidence from parallel and comparable corpora reveals that translation-induced changes are observable at multiple linguistic levels: lexical (increased frequency and expanded functions of prepositions, adoption of new vocabulary), syntactic (longer attributive structures, more explicit clause connectors, rising dependency distances), and registeral (convergence between translated and original academic prose on some dimensions but divergence on others). Theoretically, the “shining-through” effect of source languages, the normalization tendency of translators, and the gradual replication of frequency patterns from translated to original texts explain the transmission of these features. However, the shaping effect operates within the constraints of the target language system: translation activates and enriches existing grammatical resources rather than fundamentally altering Chinese linguistic typology. The paper concludes that translation functions as a conservative yet cumulative agent of change, accelerating ongoing trends while respecting the structural boundaries and cultural preferences of modern Chinese.

Keywords: Language Contact, Translation, Modern Chinese, Corpus-Based Translation Studies, Shining-Through, Normalization.

1. Introduction

Modern Chinese has undergone substantial changes since the early twentieth century, particularly in its lexicon, syntax, and discourse organization. While social transformation, educational reforms, and the vernacular movement have all contributed to these changes, translation has often been singled out as a primary vector of linguistic innovation. During the May Fourth period, translators and writers consciously introduced foreign grammatical structures and lexical items to enrich the expressive capacity of written Chinese (Zhong & Wang, 2025: 72). Yet for decades, the discussion of translation’s influence on Chinese remained largely impressionistic, relying on isolated examples and intuitive judgments rather than systematic evidence.

The emergence of corpus-based translation studies has fundamentally changed this situation. Large-scale parallel and comparable corpora enable researchers to quantify the differences between translated and original Chinese, to trace diachronic changes across multiple time periods, and to test hypotheses about language contact through translation (Laviosa, 2004). Over the past thirty years, a substantial body of empirical research has accumulated, covering diverse registers such as literary fiction, academic prose, scientific texts, and political discourse. This research provides a solid foundation for addressing a central question: How has translation shaped modern Chinese, and what are the mechanisms and limits of such influence?

This paper synthesizes evidence from sixteen corpus-based studies published between 1991 and 2026. Tracing the developmental trajectory of corpus-based translation studies (CTS), the field has undergone a methodological and theoretical diversification since its inception. Early

foundational work in corpus linguistics, such as the edited volume by Aijmer and Altenberg (1991), established the empirical groundwork for cross-linguistic contrastive analysis, while Laviosa (1998) pioneered the identification of core lexical patterns in narrative prose, directly inspiring subsequent systematic inquiries into translation universals. As quantitative methods matured, researchers began to probe the finer-grained features of “translationese”, with Volansky, Ordan, and Wintner (2015) employing large-scale computational measures to delineate its characteristic fingerprints, and Kruger (2018) applying multivariate statistical models to disentangle the complex conditioning factors behind syntactic explication in translated English. In recent years, a critical turn has emerged, questioning the absolute universality of earlier claims; for instance, Afzaal et al. (2022) scrutinized recent case studies and argued for a more context-dependent interpretation of translation patterns, while Wang and Xin (2024) proposed an integrated analytical framework that synthesizes linguistic, cognitive, and socio-cultural dimensions to better capture the variability of translated language. Specifically within the context of Chinese translation studies, Dai (2021) offered a focused examination of aspect markers, revealing systematic temporal-aspectual divergences between translated and original Chinese texts. Methodologically, Liu and Zhang (2022) provided a comprehensive retrospect of syntactic analysis and corpus development, highlighting both the advances and lingering challenges in feature extraction and tool application. Complementing these empirical and technical pursuits, Flowerdew (2009) adopted a critical pedagogical perspective, reminding practitioners of the epistemological limitations when applying corpus findings to educational practice. Taken together, these diverse yet

interlocking studies constitute the multifaceted academic landscape upon which the present paper is built, offering a rich reservoir of theoretical constructs, empirical methods, and critical insights that inform our subsequent review and synthesis.

Rather than presenting a simple literature review, the paper integrates findings across studies to develop a coherent argument about the nature, mechanisms, and constraints of translation-induced language change. The paper is organized as follows. Section 2 outlines the theoretical framework of language contact and its adaptation to translation studies, introducing key concepts such as the third code, shining-through, normalization, and frequency replication. Section 3 presents corpus-based evidence at the lexical, syntactic, and registeral levels, drawing on multiple studies to show how translated Chinese differs from original Chinese and how these differences have evolved over time. Section 4 discusses the underlying mechanisms that transmit features from translation to original writing, as well as the conditions that constrain such transmission. Section 5 concludes with implications for understanding translation as a force of linguistic change and suggests directions for future research.

2. Theoretical Framework: Language Contact Through Translation

Language contact refers to situations where speakers of different languages interact, leading to mutual influence in vocabulary, sound patterns, grammar, and discourse structures. When contact occurs indirectly through written materials, especially through translations, linguists speak of “indirect language contact” or “contact through translation” (Xu, 2026: 62). Unlike direct contact between bilingual communities living in close proximity, indirect contact is mediated by texts and often involves planned, deliberate borrowing. Translators make conscious or subconscious decisions about how to render source-language features into the target language, and these decisions, when repeated across many texts and many translators, can gradually shift the norms of the target language.

Frawley (1984) introduced the concept of the “third code” to capture the hybrid nature of translated language. Translations are neither pure source language nor pure target language as normally used by native speakers. Instead, they constitute a distinct variety shaped by the dual pressures of faithfulness to the source text and acceptability in the target language. This intermediate status makes translated texts particularly interesting for language contact research because they represent a controlled environment where cross-linguistic influence can be observed before it potentially spreads to original writing.

Baker (1993) initiated the systematic investigation of translation universals – features that recur across translations regardless of language pairs. These include explicitation (the tendency to spell things out rather than leave them implicit), simplification (the tendency to use less complex vocabulary and syntax), and normalization (the tendency to conform to typical target-language patterns). Over the following decade, corpus-based translation studies developed into a coherent paradigm, with researchers building parallel and comparable corpora for multiple language pairs (Laviosa, 2004: 9-11). However, as evidence accumulated, it became clear that these features are not absolute universals that hold for all translations in all contexts. Rather, they are probabilistic

tendencies whose strength depends on factors such as text type, language pair, translation direction, and historical period (Pang, 2024: 132-133).

In response to these findings, researchers have refined the theoretical framework. One influential proposal is to treat translated language as a type of “constrained language” (Kruger & van Rooy, 2016). Translations are produced under multiple constraints: cognitive constraints such as limited working memory and bilingual activation, social constraints such as the need to satisfy readers’ expectations and avoid criticism, and linguistic constraints such as the structural differences between source and target languages. These constraints jointly shape the features of translated texts. The constrained language perspective shifts the focus from searching for absolute universals to understanding how different constraints interact to produce particular patterns in particular contexts.

Hu and Zhu (2023) argued that the term “translation universals” is itself problematic because it carries connotations of absoluteness and scientific inevitability that are not warranted by the empirical evidence. They proposed replacing it with “translation tendencies” or “translation features” and developed a multidimensional model that integrates linguistic, socio-cultural, and cognitive explanations. In their model, translation features arise from the interaction of language-internal factors (such as typological differences between source and target languages), socio-cultural factors (such as translation norms and the relative prestige of languages), and cognitive factors (such as the salience of certain structures and the cognitive load imposed by translation). This multidimensional framework provides a useful foundation for analyzing how translation shapes the target language, because it directs attention to the multiple pathways through which source-language features can influence target-language usage.

Johanson’s (2002) code-copying framework offers another valuable theoretical resource. Johanson distinguishes between global copying, where an entire item or structure is transferred from a model language to a basic language, and selective copying, where only certain features are copied. Crucially, he also distinguishes between frequency copying, where the probability of using a structure that already exists in the target language is increased under model language influence, and structural copying, where a new structure is introduced. Malamatidou (2016) applied this framework to translation, arguing that translators often engage in frequency copying without being fully aware of it. When a translator repeatedly uses a particular Chinese structure because the English source text uses its counterpart, and when this pattern is replicated across many translations, the frequency of that structure in the target language may rise. Over time, original writers exposed to these translations may also increase their use of the structure, leading to permanent language change. This frequency-based mechanism is subtle but potentially powerful because it does not require the introduction of entirely new elements; it merely accelerates ongoing trends.

Pang and Wang (2020) further developed this line of research by examining how explicitation features in English-Chinese translation have influenced vernacular Chinese over time. They found that certain patterns of explicit marking, such as the use of conjunctions and pronouns, became more frequent in original Chinese writing after having first become common in translations. This finding supports the view that translation serves as a site of language contact where

frequency patterns are replicated and then spread to the wider language community.

3. Corpus-Based Evidence of Translation Shaping Modern Chinese

The theoretical framework outlined above predicts that if translation indeed shapes modern Chinese, we should observe systematic differences between translated and original Chinese texts, as well as diachronic shifts in original texts that parallel patterns first established in translations. This section reviews the empirical evidence from three levels of analysis: lexis, syntax, and register.

3.1. Lexical Evidence: The Case of Prepositions

Among the most tangible effects of translation on the Chinese lexicon is the increased frequency and expanded functional range of prepositions. English, as a hypotactic language, relies heavily on prepositions to mark grammatical relations such as time, location, manner, and agency. Chinese, by contrast, is a paratactic language where prepositions are fewer in number and often optional. Many Chinese prepositions have evolved from verbs and retain verbal characteristics, such as the ability to take aspect markers or to function as the main predicate in certain constructions.

Xu (2026) conducted a large-scale corpus study to investigate preposition use across English novels, their Chinese translations, and original modern Chinese novels. The English-Chinese Classics Corpus (ECCC) that he built includes English literary works from the nineteenth and twentieth centuries, their Chinese translations, and comparable original Chinese novels from the same periods. The results showed a clear continuum in preposition frequency. English novels had the highest frequency of prepositions, with an average of 102.44 prepositions per 1,000 words for British novels and 92.77 for American novels. Their Chinese translations showed substantially lower frequencies, averaging 66.01 per 1,000 characters for translated British novels and 75.67 for translated American novels. Original modern Chinese novels had the lowest frequencies, averaging 52.39 per 1,000 characters (Xu, 2026: 68-69). In other words, translated Chinese occupies an intermediate position between English source texts and original Chinese texts.

A repeated-measures ANOVA confirmed that the difference between translated and original Chinese in preposition usage was statistically significant ($F = 7.479, p < 0.05$). Moreover, the main effect of text type was strong, and the interaction between text type and period indicated that the gap between translated and original texts has been changing over time (Xu, 2026: 71-72).

Importantly, Xu also examined diachronic change within original Chinese novels. Novels written after 1949 used more prepositions (54.17 per 1,000 characters) than those written in the Republican era (43.84 per 1,000 characters). This increase suggests that over the course of the twentieth century, original Chinese writing gradually adopted a higher frequency of prepositions, moving in the direction already observed in translated texts. Xu (2026: 73) concluded that this change is best understood as the activation and expansion of existing grammatical resources rather than the creation of entirely new categories. The prepositions that increased in

frequency were already present in Chinese, such as “对”, “对于”, “关于”, “按照”, “根据”, and “通过”. Translation did not introduce new prepositions; it increased the frequency with which writers used the prepositions that already existed.

This finding is consistent with the frequency copying mechanism discussed earlier. Translators, under the influence of English source texts that use prepositions frequently, produce Chinese translations with elevated preposition frequencies. Over time, original writers who read these translations (and who may themselves be translators) internalize the pattern and produce original texts with higher preposition frequencies as well. The change is gradual and cumulative, and it respects the grammatical typology of Chinese.

3.2. Syntactic Evidence I: Attributive Complexity and Europeanized Structures

At the syntactic level, translation-induced changes are observable in several areas, including attributive complexity, clause connectors, and dependency distance. Each of these features has been examined in corpus studies using different language pairs and different methodologies.

One of the most frequently noted features of Europeanized Chinese is the increased use of long, multi-adjective attributive phrases preceding a noun. In traditional Chinese prose, modifiers are typically short, often consisting of single words or brief phrases placed before the head noun. Longer modifiers are usually placed after the head noun, often in the form of separate clauses. However, translations from European languages, where prenominal modification can be quite elaborate, have introduced the practice of stringing multiple adjectives and participial phrases before the noun in Chinese.

Liu (2024) provided a detailed analysis of this phenomenon by examining how Russian participle phrases are translated into Chinese. Russian participle phrases are verbal forms that can function as attributes, often placed after the head noun and carrying semi-predicative meaning. In Chekhov's stories and their three Chinese translations by Ru Long, Shen Nianju, and Feng Jia, Liu found that out of 264 independent participle structures in the source texts, 162 (61.4%) were rendered as prenominal attributives in the Chinese translations (Liu, 2024: 123-124). This conversion strengthens the modifying function of the participle phrase while weakening its predicative meaning. The result is a Chinese sentence with a longer, more complex noun phrase than would normally be found in original Chinese writing.

Liu further noted that this translation strategy leads to what is often called “Europeanized attributive clusters.” For example, a Russian phrase meaning “the ancient, emaciated, impassive nag” was translated with all three adjectives placed before the noun, creating a dense prenominal modification structure that is grammatically acceptable in Chinese but stylistically marked. While such constructions have become more common in modern Chinese, particularly in written registers, they remain more frequent in translated than in original texts. The persistence of this difference suggests that the influence of translation has been significant but not complete; original writers still tend to avoid the most extreme cases of attributive clustering.

Zhong and Wang (2025) took a different approach to the same question. Instead of comparing translations by different translators with original texts by different authors, they compared the translation and original novel of a single author,

Zhou Libo. Zhou first translated Sholokhov's *Virgin Soil Upturned* from English (itself a translation from Russian) and later wrote his own novel *The Hurricane* on a similar theme of rural collectivization. By comparing Zhou's translation with his later creative writing, and by comparing both with contemporary novels by other Chinese writers who did not translate extensively, Zhong and Wang were able to trace the transfer of syntactic features from translation to creation within a single individual's output.

The results showed that the translation exhibited a higher proportion of sentences containing long attributives (4.27% of all sentences) than Zhou's pre-translation writings (3.63%) and than contemporary novels by other authors (3.23%). More importantly, *The Hurricane* itself showed an intermediate proportion (4.06%), higher than other authors' novels and higher than Zhou's own earlier writings but slightly lower than his translation (Zhong & Wang, 2025: 75). This pattern strongly suggests that the act of translating exposed Zhou to a particular syntactic pattern, which he then internalized and carried over into his original writing. The influence did not completely transform his style – his original novel remained somewhat less attributively dense than his translation – but it did shift his usage in the direction of the translation.

Zhong and Wang (2025: 77) also found that Zhou's translation used more explicit clause connectors and had longer average clause lengths than both his pre-translation writings and other writers' novels. The mean clause length in the translation was 9.46 words, compared to 8.22 words in the pre-translation writings and 7.06 words in other contemporary novels. The variance of clause length was also larger, indicating greater syntactic diversity. Features such as the extended use of “和” (and) and “或” (or) to link clauses, as well as the increased frequency of pronouns such as “他” (he) and “她” (she), were carried over from the translation into **The Hurricane**. These findings provide concrete evidence that translation can shape an author's syntactic habits at a relatively deep level, affecting not just individual words but the way sentences are structured.

3.3. Syntactic Evidence II: Dependency Distance and Processing Load

A more recent development in corpus-based translation studies is the application of dependency grammar to measure syntactic complexity and processing difficulty. Dependency grammar represents sentences as networks of relations between words, with each link connecting a head (governor) and a dependent. The distance between a head and its dependent, measured in terms of the number of intervening words, is called dependency distance. Mean dependency distance (MDD) for a sentence or a corpus is the average of these distances. Research has shown that MDD correlates with processing difficulty: longer dependency distances impose greater demands on working memory.

Jiang, Fan, and Wang (2021) applied this method to compare English translations of Chinese government work reports with original English State of the Union addresses. The two text types are comparable in register and length, both being political speeches delivered annually. The Chinese work reports were translated into English by professional translators, while the State of the Union addresses were originally written in English by American presidents and their speechwriters. The research question was whether the English translations differed systematically from native English texts

in terms of dependency distance.

The results were striking. The translated English texts had a significantly higher MDD (3.713) than the original English texts (3.017). After controlling for sentence length by restricting the analysis to sentences between 5 and 40 words, the difference remained substantial (3.333 vs. 2.900). When text length was further controlled by creating matched samples of equal size, the difference persisted (3.429 vs. 2.931). All t-tests confirmed that the differences were statistically significant ($p < 0.001$) (Jiang et al., 2021: 43-44).

The authors interpreted this finding as evidence of source-language interference, specifically the shining-through effect. Chinese has a mean dependency distance of approximately 3.662 (Liu, 2008, cited in Jiang et al., 2021: 44), which is considerably higher than native English's MDD of around 2.5 to 2.9. When Chinese texts are translated into English, the high MDD of the source language influences the target language, resulting in English translations that have longer dependency distances than native English. In other words, the syntactic processing load characteristic of Chinese is carried over into the English translations, making them more demanding to process than original English texts.

While this study examined translation from Chinese into English, the principle is symmetric. When English texts are translated into Chinese, the low MDD of English might be expected to influence Chinese translations, potentially reducing dependency distances compared to original Chinese. Alternatively, the much higher MDD of Chinese might resist such reduction, resulting in translations that still have high MDD but perhaps slightly lower than original Chinese. Direct testing of this hypothesis awaits further research, but the Jiang et al. study establishes the methodological feasibility of using dependency distance as a quantitative measure of translation-induced syntactic influence.

From the perspective of language contact, this finding is significant because it shows that translation can affect not just the choice of words or the order of constituents but the global syntactic structure of sentences. Translators, under the influence of the source language, produce target sentences with different processing profiles than native speakers would produce. If such patterns are replicated across many translations, they could potentially shift the norms of the target language, making longer dependency distances more acceptable over time. This would constitute a fundamental change in the syntactic character of the language, affecting how writers organize information within sentences.

3.4. Registeral Evidence: Convergence and Divergence Across Dimensions

The lexical and syntactic evidence reviewed so far focuses on individual features in isolation. However, language contact through translation can also affect the overall registeral profile of a language – the characteristic clustering of linguistic features that distinguishes, for example, academic prose from fiction or speech from writing. Zhang and Pang (2026) conducted a diachronic multi-dimensional analysis to investigate how translated and original Chinese academic texts have changed relative to each other over nearly a century.

They built a corpus of Chinese translated and original academic texts from four periods: around 1930, 1960, 1990, and 2010. The total corpus size was approximately 3.2 million words. They extracted 59 linguistic features, including traditional lexico-grammatical indices (such as parts of speech and clause structures), logical-semantic relations

(such as causal and contrastive connectors), and features of classical Chinese style (such as embedded monosyllabic words and classical syntactic patterns). Using factor analysis, they identified five functional dimensions.

The first dimension, which they labeled “logical explicitness versus abstract narration,” had high positive loadings for causal relations, concessive conjunctions, and connectors, and high negative loadings for nominalization and past tense. On this dimension, original academic texts initially showed higher logical explicitness than translated texts, meaning they used more explicit connectors. Over time, however, original texts shifted toward abstract narration, while translated texts remained stable. As a result, the two registers converged. By the 2010 period, the difference between translated and original texts on this dimension was no longer statistically significant (Zhang & Pang, 2026: 32).

The third dimension, “precision versus redundancy,” had positive loadings for numerals, classifiers, and proper nouns, and negative loadings for particles (such as “的”, “了”, “吗”). On this dimension, original academic texts initially had higher precision scores than translated texts, meaning they used more precise, information-dense language. Over time, original texts moved toward redundancy, while translated texts remained stable. Again, the two registers converged. The authors suggested that the high frequency of the particle “的” in translated texts may have spread to original writing through frequency replication (Zhang & Pang, 2026: 34). Particles such as “的” are grammatically optional in many contexts; their increased use adds redundancy without adding meaning. The fact that original texts have become more redundant over time, moving closer to the pattern of translated texts, supports the hypothesis that translation-induced frequency changes can spread to original writing.

However, not all dimensions showed convergence. On the second dimension, “argumentative features”, which had high loadings for epistemic modals and conditional conjunctions, translated and original texts diverged. Translated texts showed a stronger and increasing tendency to use explicit argumentative markers, while original texts maintained a lower frequency. Zhang and Pang attributed this to different rhetorical traditions: English academic writing, which heavily influences translated texts, values explicit argumentation and the clear marking of the writer’s stance, whereas traditional Chinese academic writing favors exemplification over explicit argumentation (Zhang & Pang, 2026: 33). This divergence suggests that translation’s influence is not unlimited; when a source-language feature conflicts with a deeply entrenched cultural preference in the target language, it may fail to spread.

On the fourth dimension, “interactivity versus informativeness”, translated texts remained more informative (higher loading on nouns) while original texts became more interactive over time (higher loading on first-person pronouns and emphatics). The two registers diverged after 1960. On the fifth dimension, “resonance versus relevance”, the pattern fluctuated, with periods of convergence followed by periods of divergence. These findings demonstrate that translation’s influence on registeral features is selective and dimension-specific. Some features are readily transferred and lead to convergence; others are resisted and lead to divergence; still others show complex patterns that depend on historical circumstances.

4. Mechanisms and Limits of Translation-Induced Change

The empirical evidence reviewed in the previous section establishes that translation has indeed shaped modern Chinese in measurable ways. This section addresses the mechanisms that transmit features from translation to original writing and the conditions that constrain such transmission.

4.1. Shining-Through and Source-Language Interference

The most direct mechanism is source-language interference, often called shining-through (Teich, 2003). When a translator produces a target text, elements of the source language’s structure, lexical choices, and collocational patterns may be carried over. This can happen consciously, when the translator deliberately preserves a source-language feature for reasons of fidelity, or subconsciously, when the translator’s bilingual competence leads to automatic transfer.

Li, Zhang, and Liu (2011) provided a clear illustration of this mechanism in their comparison of two English translations of *Hongloumeng*. The Yangs, translating for the Foreign Languages Press in Beijing, adopted a literal strategy that preserved many Chinese cultural terms and syntactic patterns. Their translation retained Chinese-specific references, used footnotes extensively to explain cultural concepts, and followed Chinese word order where possible. In contrast, Hawkes adopted a more free translation, paraphrasing cultural terms and restructuring sentences to conform to English expectations. The difference in translation strategy directly affected how much Chinese source-language material shone through into the English target text (Li, Zhang & Liu, 2011: 12-14).

While this study concerns translation from Chinese into English, the same principle applies to translation into Chinese. When English is the source language, its hypotactic structure, frequent use of prepositions, and elaborate pronominal modification patterns can shine through into Chinese translations. The increased frequency of prepositions observed by Xu (2026) and the longer attributive structures observed by Liu (2024) and Zhong and Wang (2025) are manifestations of shining-through. Translators, whether consciously or not, carry over source-language patterns into their target-language output.

The shining-through effect is stronger when the source and target languages are typologically distant. English and Chinese differ substantially in their preferred modes of discourse organization: English favors hypotaxis (explicit marking of relationships with conjunctions and subordinators), while Chinese favors parataxis (implicit relationships signaled by word order and context). This typological distance creates many opportunities for shining-through, because there is no one-to-one correspondence between source and target structures. Translators must make decisions, and those decisions are influenced by the source text’s patterns.

4.2. Normalization and Risk Avoidance

If shining-through were the only mechanism operating, translations would simply be hybrids that mix source and target features without any systematic bias. However, a second mechanism, normalization, pulls translations toward target-language conventions. Normalization refers to the

tendency of translators to use more typical, more standard, and often more formal target-language forms than original writers do (Baker, 1996). This tendency is driven by risk avoidance: translators, aware of their mediating role and the possibility of being criticized for unnatural language, choose forms that are safe, conventional, and unlikely to attract negative attention.

Hu and Li (2025) provided systematic evidence of normalization in both human and machine translation. Using a lectometric approach, they compared the use of formal versus informal lexical and grammatical variants across four registers: news, general prose, academic writing, and fiction. The formal variants included single verbs (e.g., “ascend”) rather than phrasal verbs (e.g., “go up”) and non-contracted negatives (e.g., “is not”) rather than contracted forms (e.g., “isn’t”). The informal variants were the opposites.

The results showed that human translations were significantly more formal than original English texts in all four registers. The effect was strongest in news and fiction, where the difference was largest, but it was present even in academic writing, where formality is already high. Machine translations also showed normalization in news and fiction, but not in general and academic registers, where they did not differ significantly from original English (Hu & Li, 2025: 42-44). The authors interpreted this as evidence that human translators are more sensitive to register-specific norms than current machine translation systems. Humans adjust their language to the expected level of formality for each register, while machines may rely more on general patterns learned from training data.

In the Chinese context, normalization operates alongside shining-through. A translator can simultaneously preserve a source-language structure (shining-through) while using more formal target-language vocabulary (normalization). For example, a translator might preserve the English-like prenominal modification pattern but choose formal Chinese vocabulary items rather than colloquial ones. The net effect on the target language depends on which force is stronger for a given feature in a given register.

4.3. Frequency Replication and the Spread to Original Writing

The mechanisms discussed so far explain how translated texts come to have distinctive features. However, for translation to shape the target language in a lasting way, those features must spread from translated texts to original writing. The key mechanism for this spread is frequency replication (Johanson, 2002; Malamatidou, 2016). Frequency replication does not require the introduction of entirely new structures; it operates on structures that already exist in the target language. When translators use a particular structure more frequently under source-language influence, and when original writers exposed to these translations also increase their use of that structure, the frequency of the structure in the target language rises. This frequency shift is a form of language change that is gradual, cumulative, and often unnoticed by speakers.

Pang (2024) tested this hypothesis across three registers (literary, scientific, academic) and two time periods (1930s and 1990s). She examined five indicators of simplification (high-frequency word coverage, lexical density, type-token ratio, mean word length, and mean segment length) and two indicators of explicitation (connectors and personal pronouns). The results showed that for some features, the differences between translated and original Chinese remained

stable across the two periods, suggesting that frequency replication was weak or absent. For other features, however, the differences diminished over time, indicating convergence. The interaction between register and period was significant for several features, meaning that the spread of features from translation to original writing depends on both the register and the historical context (Pang, 2024: 142-143).

For example, the use of connectors in literary translations remained higher than in original literary works across both periods, showing no convergence. In scientific texts, by contrast, the difference between translated and original texts disappeared in the later period, suggesting that connectors spread from scientific translations to original scientific writing. This register-specific pattern may be explained by different norms: scientific writing values clarity and explicitness, making connectors more acceptable, while literary writing may value parataxis and implicit coherence, resisting the increased use of explicit connectors.

Pang and Wang (2019) reviewed the state of the art in corpus-based translation studies and identified language contact research as one of the most promising frontiers. They noted that researchers are increasingly treating translated texts as a “contact variety” and are investigating the conditions under which translation-induced changes become permanent. These conditions include the intensity and duration of contact (how many translations are produced and read over how long a period), the prestige of the source language and culture (whether speakers are motivated to imitate source-language patterns), the availability of alternative native constructions (whether the target language has a competing way to express the same meaning), and the degree of awareness among writers and editors (whether the change is noticed and either encouraged or resisted).

4.4. The Limits of Translation-Induced Change

Despite the substantial evidence that translation has shaped modern Chinese, it is important not to overstate the extent of this influence. Three types of constraints limit the power of translation to reshape a language.

The first constraint is grammatical typology. Xu (2026: 73) made a crucial observation: although translated Chinese uses more prepositions than original Chinese, the increase is within the range of forms already present in the language. Chinese does not develop new categories of prepositions, nor does it change the fundamental fact that prepositions in Chinese are derived from verbs and still retain some verbal properties. Similarly, the increased use of the passive marker “被” does not transform Chinese into a language with a fully grammaticalized passive voice comparable to English. Translation accelerates existing trends but does not create typologically new structures. This constraint is not absolute – over very long time scales, languages can undergo typological change – but it operates strongly over the century-long period examined in the available studies.

The second constraint is rhetorical tradition. Zhang and Pang (2026: 35) found that on the argumentative features dimension, translated and original academic texts diverged rather than converged. Original Chinese academic writing continues to favor exemplification over explicit argumentation, reflecting a rhetorical tradition that differs from English academic norms. This suggests that translation’s influence is resisted when it conflicts with deeply embedded cultural preferences about how to persuade and inform readers. Chinese writers may be aware of English-style

argumentation but choose not to adopt it because they believe it is less appropriate for Chinese readers.

The third constraint is editorial and institutional gatekeeping. Translations, especially those produced for official or commercial publication, are subject to editing, proofreading, and quality control. Editors may alter translators' choices to conform to institutional norms, and these norms may be more conservative than the norms of translators or original writers. Kruger (2017, cited in Huang & Wang, 2023: 772) argued that many features traditionally attributed to translation may actually result from editorial intervention. In the Chinese context, the Yangs' translation of Hongloumeng was produced for the Foreign Languages Press, a government-funded publisher with a mandate to promote Chinese culture abroad. The institutional context shaped the translation strategy, favoring literalness and footnotes over fluency (Li, Zhang & Liu, 2011: 12-14). Such gatekeeping can either amplify or dampen translation-induced changes, depending on the institution's norms.

5. Conclusion

This paper has examined how translation has shaped modern Chinese, drawing on a synthesis of corpus-based evidence from sixteen studies and theoretical explanations rooted in language contact theory. The findings can be summarized in three main points.

First, translation-induced changes are observable at multiple linguistic levels. At the lexical level, translation has led to increased frequency and expanded functions of prepositions, with translated Chinese occupying an intermediate position between English source texts and original Chinese texts (Xu, 2026). At the syntactic level, translation has introduced longer attributive structures (Liu, 2024; Zhong & Wang, 2025), more explicit clause connectors (Zhong & Wang, 2025), and increased dependency distances (Jiang et al., 2021). At the registeral level, translated and original academic texts have converged on some dimensions (such as abstract narration and redundancy) but diverged on others (such as argumentative features), showing that translation's influence is selective and dimension-specific (Zhang & Pang, 2026).

Second, the mechanisms underlying these changes include shining-through (source-language interference), normalization (risk-driven conformity to target norms), and frequency replication (the gradual spread of usage patterns from translated to original texts). These mechanisms operate simultaneously and sometimes in opposition. The net effect on the target language depends on the relative strength of each mechanism for a given feature in a given register and historical period.

Third, the influence of translation is constrained by the typological boundaries of Chinese, by persistent rhetorical traditions, and by editorial and institutional gatekeeping. Translation activates and enriches existing grammatical resources rather than fundamentally altering the linguistic system. It accelerates ongoing trends but does not create typologically new structures. It can change frequencies and preferences, but it cannot easily override cultural preferences about how to communicate effectively.

The corpus-based approach, which has transformed translation studies over the past three decades (Laviosa, 2004; Huang & Wang, 2023), has made it possible to move beyond anecdotal observations to systematic, quantifiable descriptions of translation-induced language change. The

evidence reviewed in this paper supports the view that translation is a significant agent of change in modern Chinese, but it is a conservative and cumulative agent. Translation does not revolutionize the language overnight; it nudges it gradually in particular directions, and those nudges accumulate over decades and centuries. Understanding this process requires not only more data from more language pairs and registers but also more sophisticated theoretical models that can explain why some features spread and others do not.

Future research should extend the diachronic coverage to earlier and later periods, should compare translation-induced changes with changes arising from other forms of language contact (such as bilingual education and digital communication), and should integrate experimental methods to test the cognitive reality of frequency replication. The increasing availability of large language models also offers new tools for simulating and analyzing language contact scenarios (Xu & Sun, 2025). For now, the evidence clearly indicates that translation does shape modern Chinese, but it does so gradually, selectively, and always within the structural and cultural constraints of the receiving language.

References

- [1] Afzaal, M., et al. (2022). Myth or reality? Some directions on translation universals in recent corpus-based case studies. *Frontiers in Psychology*, 13, 90–100.
- [2] Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. *Corpus Linguistics and Translation Studies*, 234–249.
- [3] Dai, G. R. (2021). A corpus-based study on the use of aspect markers in translated Chinese texts. *Technology Enhanced Foreign Language Education*, (4), 76–82.
- [4] Li, D., Zhang, C., & Liu, K. (2011). Translation style and ideology: A corpus-assisted analysis of two English translations of Hongloumeng. *Literary and Linguistic Computing*, 26(2), 153–166.
- [5] Hu, M., & Zhu, Y. (2023). From translation universals to a multi-dimensional model of translation features. *Technology Enhanced Foreign Language Education*, (5), 13–18.
- [6] Hu, X., & Li, J. (2025). A metrical study of normalization features in human and machine translations across registers. *Technology Enhanced Foreign Language Education*, (2), 37–46.
- [7] Huang, L., & Wang, K. (2023). Developmental stages and frontier trends in corpus-based translation studies. *Foreign Language Teaching and Research*, (5), 764–776.
- [8] Jiang, Y., Fan, L., & Wang, Y. (2021). Syntactic features of translated language based on dependency treebanks. *Foreign Language Education*, (3), 41–46.
- [9] Aijmer, K., & Altenberg, B. (Eds.). (1991). *English corpus linguistics*. Longman.
- [10] Kruger, H. (2018). That again: A multivariate analysis of the factors conditioning syntactic explicitness in translated English. *Across Languages and Cultures*, 1585–1923.
- [11] Laviosa, S. (1998). Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta*, (4), 557–570.
- [12] Liu, D., & Zhang, Z. (2022). Syntactic analysis and corpus development and application: Retrospect and prospect. *Technology Enhanced Foreign Language Education*, (6), 82–88.
- [13] Flowerdew, L. (2009). Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistics*, (3), 393–417.

- [14] Pang, S., & Wang, K. (2019). On the trends in corpus-based translation studies from the perspective of expanding the third code. *Foreign Language Education*, (1), 95–99.
- [15] Pang, S. (2024). A dynamic investigation of translation universals: A multi-genre analysis based on diachronic corpora. *Foreign Languages and Literature*, (5), 132–144.
- [16] Laviosa, S. (2004). Corpus-based translation studies: Where does it come from? Where is it going? *Language Matters: Studies in the Languages of Africa*, (1), 6–27.
- [17] Volansky, V., Ordan, N., & Wintner, S. (2015). On the features of translationese. *Digital Scholarship in the Humanities*, (1), 98–118.
- [18] Wang, G., & Xin, Y. (2024). An analytical framework for corpus-based translation studies. *Humanities and Social Sciences Communications*, (4), 1–12.
- [19] Xu, J., & Sun, M. (2025). Innovations in corpus-based translation research methods driven by large language models. *Foreign Language Teaching and Research*, (3), 456–468.
- [20] Xu, W. (2026). A comparative study of English and Chinese prepositions and their translation from the perspective of language contact: A sampling analysis based on the 'English-Chinese Literary Classics Corpus'. *Journal of Foreign Languages*, (2), 62–73.
- [21] Zhang, Z., & Pang, S. (2026). A diachronic comparative study of stylistic features between translated Chinese academic texts and original Chinese academic texts. *Technology Enhanced Foreign Language Education*, (2), 28–36.