frontiers | Frontiers in Language Sciences

TYPE Original Research
PUBLISHED 12 January 2026
DOI 10.3389/flang.2025.1712405

OPEN ACCESS

EDITED BY
Mark Antoniou,
Western Sydney University, Australia

REVIEWED BY
Juqiang Chen,
Shanghai Jiao Tong University, China
Muhammad Guruh Nuary,
Universitas Muhammadiyah Tangerang, Indonesia

*CORRESPONDENCE
Yanmeng Liu
✉ liuyanmeng@nwpu.edu.cn

RECEIVED 24 September 2025
REVISED 04 December 2025
ACCEPTED 08 December 2025
PUBLISHED 12 January 2026

CITATION
Liu Y (2026) Translation as language contact: a multidimensional perspective of syntactic variation. *Front. Lang. Sci.* 4:1712405. doi: 10.3389/flang.2025.1712405

Translation as language contact: a multidimensional perspective of syntactic variation

Yanmeng Liu*

School of Foreign Studies, Northwestern Polytechnical University, Xi'an, China

Translators produce translations via consistent contact with at least two languages. Contact with different languages is a main factor influencing language changes at all levels. However, quantitative exploration of syntactic variation in translation as a language contact event remains underexplored. This study explains translation changes through the lens of language contact and quantifies the influence of language contact on translation syntax in fiction and non-fiction genres. Univariate and multivariate analyses were conducted to examine linguistic variations at the syntactic level in Chinese-English translations by two groups of undergraduate student translators (Grade 4, with a higher degree of language contact; and Grade 3, with a lower degree), with expert translations serving as a benchmark for acceptable language changes. The results reveal four primary dimensions of syntactic variation: (1) structure diversity, (2) production length, (3) sentence complexity, and (4) information fragmentation. Genre-specific effects were observed, with translations in both genres exhibiting statistically significant syntactic changes across four dimensions, while fiction translation insignificant regarding the dimension of structure diversity. The findings imply that contact-induced linguistic changes amplify simplification and explicitation in translation, and manifest more prominently in genres oriented toward information transfer than in those prioritizing stylistic distinctiveness and immersive storytelling. These findings suggest that language contact offers a valuable lens for understanding linguistic changes in translation and imply that genre and language change mechanism are important factors in this perspective.

KEYWORDS

corpus, dimensions, language contact, syntax, translation

1 Introduction

Translation is a special language produced by translators with intense interactions between source and target languages. Translation has been regarded as a language in previous studies noted as "translated language" (Kruger and Rooy, 2012; Tirkkonen-Condit, 2002), and "constrained language" (Baker, 1993, 1996; Toury, 1995; Chen et al., 2024; Kajzer-Wietrzny and Ivaska, 2020). Studies have explored universal features that set translation apart from other types of languages, such as non-translations (Chesterman, 2004; Glynn, 2021). As an output of translators contacting with source and target languages intensively, translation is a language type distinctive from other types. A language is considered to experience continuous change and dynamic development influenced by speakers' interactions with two or more languages, also known as language contact (Matras, 2020; Thomason, 2022). In the context of translators interacting with source and target languages, language contact occurs and provides insight to explain language changes in translations. However, analysis of language contact's impact on translations is scarcely tapped.

The current study sets out to identify the characteristics of language contact-induced changes in translations. The author selected the linguistic variations in translations with different contact degrees that conform to such characteristics to quantify the impact of language contact on translations, focusing on syntactic features. Quantitative analyses, specifically multidimensional analyses, are conducted to explore a possible way to analyze language contact in translations. By doing so, this study reached the aims of formulating explanatory dimensions for the influence of language contact on translations from a quantitative perspective and further expanding the scope of language contact research into translation studies as informed by the view that translation is a language on its own.

2 Literature review

2.1 Common ground: language contact and translation studies

Language contact has received considerable scholarly attention since its launch in 1953 (Weinreich, 1953), and the common ground shared with translation makes it an alternative perspective in studying language changes in translations. Both keep interacting with different languages in large or small communities. Language contact refers to interactions of different languages, resulting in linguistic changes (Sankoff, 2004). Translation is inherently a contact phenomenon. Interactions with different languages are intense in translation activities, as translators continue contacting the source and target languages before, during, and after translating. Before translating, translators obtain bilingual or multilingual competence with the help of materials in languages involved in the translations, as a prior condition for translation (Salamah, 2021). During the translation process, translators transfer the information in the source language into the target language, actively interacting with both languages (Lörscher, 2005). After translating, translators revisit the text in the source language to modify their translations in the target language through editing and revision (Mellinger, 2018). Intense contact with different languages makes translation a language produced under the influence of language contact. What needs to emphasize here is that even though typical studies on language contact focus the interactions that drive large-scale changes in social linguistic environments, language contact operates at small communities as well. The classic foundational proposal by Uriel Weinreich (1953, p. 1) defines language contact as "the use of two languages by the same persons in alternation," which directly includes small-group bilingualism. Contact-induced language changes can be caused by any degree of bilingualism, even within small communities (Thomason and Kaufman, 2001). This study views translation as a language under the influence of language contact in small communities, for example, students in training programs.

Both translation and language contact involve bidirectional transmissions of linguistic elements. Language contact studies identify several types of linguistic element transfer induced by language contact. For example, borrowing (including lexical and structural borrowing, referring to speakers' use of single lexical units or syntactic or grammatical structures of one language into another) (Comrie, 2000; Durkin, 2020; Winford, 2010), code-switching (speakers use both languages simultaneously) (Adamou and Shen, 2019; Verschik, 2021), and pidginization (Huber, 2020; Hymes, 2020; O'Neil, 2019). The influence of language contact on the involved languages is bidirectional intrinsically (Poplack, 2020; Thomason and Kaufman, 2001; Winford, 2003), resulting in reciprocal impacts on vocabulary, structure, and phonology. Such linguistic elements push forward communication and influence among different cultures, leading to shared concepts, practices, and expressions (Hickey, 2010; Matras, 2020). In translation, linguistic elements are transferred from source language to target language, which is investigated as interference (Mauranen, 2004; Xiao, 2015) or shining-through of the source language (Teich, 2003), and foreignization (Paloposki, 2011; Van Poucke, 2012). The translation can also conform to linguistic norms in the target language, which is referred to as normalization (Baker, 1993, 1996; Toury, 1995), adaptation (Hu, 2003; Houlind, 2001), or domestication (Wang, 2014). No matter whether translators choose to keep norms in the source language or adapt to the target language to facilitate cross-cultural communication (Pym, 2004), translation is similar to language contact in a sense of transmitting linguistic elements bidirectionally.

Both translation and language contact can be investigated as a result of prolonged exposure to or interaction with diverse languages. Language contact investigates the effects of contact over time, resulting in observing changes as a consequence of sustained contact with diverse languages in this field. Carter et al. (2020) tracked phonetic and prosodic influences of Spanish on English among second-generation Miami-born Latinx speakers. Significant vocalic differences were observed in their prosodic rhythm and vowel quality. Hennecke and Wiesinger (2023) tracked the code-switching and calquing regarding multi-word units of Spanish-English bilinguals in Arizona during a time span of three-generation speakers. Pešková (2024) studied yes/no questions in Guaran-Spanish bilinguals in Paraguay of contact between the two languages since the 16th century and found into national and syntactic innovations apparent in Paraguayan Spanish that could be attributed to language changes brought by contact with Guarani. Translation studies can also be framed through the lens of long-term cumulative cross-linguistic interaction. Pang and Wang (2025) investigated evolution of register features in translated Chinese from 1919 to 2019, uncovering stable linguistic traits over time with stage-wise variations. Shen and Kotze (2025) observed changes in translated English in diplomatic discourse from 1946 to 2022, finding that such changes tend to lag behind the changes of untranslated English. Collectively, they underscore a parallel between language contact and translation studies regarding investigations of accumulative multilingual interactions and enduring impact on language changes over time.

2.2 Introduction of language contact into translation studies

As already observed, language contact and translation share elements that bridge the two and lay a theoretical foundation for introducing language contact into translation. When language

contact was proposed earlier, Weinreich underscored the role of translators as linguistic transfer agents (Weinreich, 1953). This represents a consensus in translation studies with scholars highlighting interlingual translation as a conduit for linguistic and cultural borrowing (Jakobson, 1959). Follow-up scholars dug deeper into the field. Becher et al. (2009) investigated English-German translations in popular science and economic genres to examine whether translation as a case of language contact can cause convergence and divergence between the source and target languages. In their study, English and German convergence were found in the use of concessive conjunctions and divergence in that of modal verbs. Bisiada (2013) compared frequencies of four causal conjunctions "because," "as," "since" and "for" in translated German from English and non-translated German, observing a trend from hypotaxis to parataxis in non-translated German as a result of contacting English through English to German translations. Malamatidou (2016) acknowledged translation is a significant site of language contact and captured how passive voice is used in translated Greek from 1990 to 2010 and related the diachronic changes to the increased frequency of passive voice in non-translated Greek. Language changes of explicitness was considered because of contact and measured with the adversative conjunction as an indicator by Pang and Wang (2020) to investigate English-Chinese translations' impact on original Chinese. Their examination in three sampling periods in the twentieth century showed that translated Chinese changed with original Chinese regarding adversative conjunction frequencies.

Existing studies enhance the integration of language contact and translation studies, shedding light on investigating translation with an alternative perspective of language contact. However, most studies treat translation as a carrier or a channel of language contact and focus on the impact of translation on the non-translated language. Studies in this domain and those with translation as a language itself, investigating the impact of contact with two different languages on it, are unbalanced. Moreover, as reviewed in the previous section, language contact-induced changes can occur at various language levels. Existing research mainly focuses on lexical-level language changes in their translation studies from a language contact perspective, leaving changes at other linguistic levels such as the syntactic level, in need of research. Furthermore, studying an individual linguistic feature, like most studies, to examine contact-induced language changes is insufficient. Language is a complex system, with multiple linguistic features functioning together to realize communication (Biber, 1991). Therefore, considering interrelations among multiple linguistic features, multivariate analysis was introduced to translation studies (De Sutter and Lefer, 2020; Kajzer-Wietrzny and Ivaska, 2020; Kruger, 2019). Multivariate analysis of multiple linguistic features helps uncover underlying dimensions to explain language variations.

In this study, language contact-induced syntactic changes in translations refer to modifications in the syntactic features of translations that arise directly from translators' sustained contact with source and target languages. These changes are distinguishable from other translation-related adjustments by two core characteristics: rooted in prolonged interactions and community acceptance. Firstly, contact-induced syntactic changes are driven by translators' prolonged interactions with source and target languages. Time and interaction are essential for language contact to induce changes to a language (Thomason, 2020). For example, Jon-And and Aguilar (2019) observed that it took seven years of interactions between the speakers with Portuguese as their first language and second language in Maputo, Mozambique, to introduce novel linguistic features into the language frequently. Thus, syntactic features in translations are identified as those changes due to the contact under prolonged interactions with source and target languages. Besides, contact-induced changes are shaped by community-level spread and acceptance. Spreading and using the changes in a community are necessary to affect the language as a whole (Matras, 2020). "Community" here can be illustrated from two sides. One is that such changes are caused by groups of people involved in contact with different languages, rather than by any individual. The other is that the changes should be accepted by groups of people, merging the changes into the original languages, and being used by communities. Accordingly, changes in linguistic features in translation can be identified as contact-induced changes provided that the changes are found and accepted in translations in large quantity.

Degree of language contact is a multi-faceted construct for encompassing a series of key factors that collectively influence the intensity of language contact. The foundational elements of language contact degree have long been validated in contact linguistics: Thomason and Kaufman (2001) identified bilingual type, contact frequency/duration, social relations, and group integration as primary determinants of contact-induced changes. Winford (2003) expanded this framework by emphasizing typological distance and acquisition contexts as critical modulating variables, with subsequent support from scholars (Muysken, 2013). This inherent complexity poses a challenge for controlling the degree of language contact to detect contact-induced language changes. This study prioritizes the selection of research objects that allow for the systematic regulation of confounding factors.

Therefore, this study operationalizes contact-induced syntactic changes in translation as sustained, community-endorsed syntactic variations emerging from interactions with source and target languages. Building on the defining characteristics of such changes, this study investigates the impact of language contact on translation syntax. Through proper object selection, the analysis is anchored in the dimensional frameworks derived from multivariate analysis of the identified syntactic features, enabling empirical quantification of contact-induced syntactic changes in translation.

3 Methods

3.1 Sampling criteria

This study selects Chinese-English translations by undergraduate English majors in China as research objects. By selecting 3rd- and 4th-grade English majors, this study neutralizes most confounding factors inherent to the degree of language contact and retains one additional year of training as a proxy to contact difference between higher (4th grade) and lower (3rd grade) language contact degrees. This design allows the study

to attribute observed linguistic variation in translations directly to the results of one-year cumulative contact with source and target languages, operationalized through contact duration.

The group exhibits homogeneous language proficiency. On one hand, all students have passed the Test for English Majors-Band 4 (TEM-4), a benchmark validated to indicate solid English foundational skills (Jin and Fan, 2011). On the other, for advanced second language learners, their second language proficiency will enter a stable period as their learning progresses (Han, 2004); the development of translation competence follows a similar pattern, where a plateau effect occurs when learners reach an advanced stage (Göpferich, 2015). In the four-year training system for English majors in China, 3rd- and 4th-grade students are in this stable stage: their English language proficiency and translation skills are relatively mature, eliminating language proficiency as a confounding variable. Their ages are concentrated between 20 and 24, with balanced demographics, further minimizing between-group variability. As native Chinese students in campus, their language contact is also socially constrained: bilingual interactions in intimate social networks are rare, with contact primarily occurring in structured education settings. English major programs in Chinese universities follow highly uniform curricula (Salamah, 2021), so 3rd- and 4th-grade students share identical teaching standards, course objectives, and training model. Meanwhile, in such settings, they feature intense interactions with source and target languages. High intensity of contact in a training context can lead to observable language changes in student translations. Collectively, these attributes validate the sample selection, which helps isolate contact duration (the one-year gap between grades) as variable to observe contact-induced language changes in translations by students with higher contact degree compared to those with lower one.

3.2 Corpus compilation

A corpus populating Chinese-English translations by undergraduate students in China with three-year (student with lower language contact degree, shortened as L) and four-year (student with higher language contact degree, shortened as H) trainings is compiled (271,693 tokens). The source of the student samples was an established corpus, Parallel Corpus of Chinese EFL Learners, which exceeds two million tokens incorporating Chinese-English translations from students at 13 universities in China (Wen and Wang, 2009). To mitigate the potential impact of regional disparities in prior English education of students, the universities covered diverse regions across China. The students are L1 Chinese, proficient in English, and have trained for three and four years at university level. A uniform level of translation difficulty was ensured by designing comparable translation tasks for students. The genres of student translations are further divided into two categories: non-fiction (magazines, news reports, and popular science articles), and fiction (novels, and stories). Along with the student translation samples, this study also collected Chinese-English translations by expert translators as a benchmark of community acceptance. Materials for expert translators (shortened as E) were excerpts of published fiction (a list of eight published fiction works is attached in Supplementary material), and non-fiction translations from http://language.chinadaily.com.cn/. The website was accessed for Chinese-English expert translations, because this platform provides authorized and free language resources for the public with rich topics such as popular science articles, news reports, magazines. Two hundred excerpts of expert translations were randomly collected from this website as non-fiction E samples. All the translation samples were categorized as fiction (shortened as F) and non-fiction (shortened as NF) genres, and built in comparable sizes, as shown in Table 1.

TABLE 1 Corpus details.

Category	L Translations		H Translations		E Translations	
Genre	F	NF	F	NF	F	NF
Sample size	200	200	200	200	200	200
Token count	50,784	38,196	49,677	39,869	52,586	40,581
Token in total	271,693					

3.3 Syntactic data collection

The Syntactical Complexity Analyzer (Lu, 2010, 2011) was employed to annotate the syntactic features in the corpus, and linguistic data for 14 syntactic features for 1,200 samples collected to compile the dataset for further analysis. These 14 syntactic features were originally designed to assess L2 writing, and have been introduced into translation studies to investigate patterns of translated languages (Liu and Afzaal, 2021; Wang et al., 2023). This study adopts these indices, because they focus on robust and generalizable syntactic features, avoiding distraction by niche phenomena that may not reflect contact-induced changes. Even though they do not capture translation-specific phenomena like connective use or word order adjustments, these phenomena are inherently rooted in the underlying syntactic metrics the indices quantify. The list of syntactic features and their formulas are attached in Table 2.

3.4 Statistical methods

3.4.1 Descriptive statistics and pretest of data

This study first observed the descriptive statistics of 14 syntactic features across high/low degree of contact in non-fiction/fiction genre (Table 3). Features exhibiting extreme deviation from normality include MLS (skewness = 2.106, kurtosis = 13.059, standard deviation (SD) = 7.336), MLT (skewness = 3.370, kurtosis = 25.410, SD = 6.732), MLC (skewness = 5.474, kurtosis = 54.135, SD = 2.163) in non-fiction translations by students with lower contact degree; MLS (skewness = 10.512, kurtosis = 133.559, SD = 15.914), MLT (skewness = 11.674, kurtosis = 154.216, SD = 15.902), MLC (skewness = 13.787, kurtosis = 193.171, SD =15.746) in non-fiction translations by students with higher

TABLE 2 Calculation of syntactic features.

#	Abbreviations and calculation formulas
1	MLS (Mean Length of Sentence) = $\frac{\text{number of words}}{\text{number of sentences}}$
2	MLC (Mean Length of Clause) = $\frac{\text{number of words}}{\text{number of clauses}}$
3	MLT(Mean Length of T unit) = $\frac{\text{number of T units}}{\text{number of clauses}}$
4	C/S = $\frac{\text{number of clauses}}{\text{number of sentences}}$
5	C/T = $\frac{\text{number of clauses}}{\text{number of T units}}$
6	T/S = $\frac{\text{number of T units}}{\text{number of sentences}}$
7	DC/C = $\frac{\text{number of dependent clauses}}{\text{number of clauses}}$
8	DC/T = $\frac{\text{number of dependent clauses}}{\text{number of T units}}$
9	VP/T = $\frac{\text{number of verb phrases}}{\text{number of T units}}$
10	CT/T = $\frac{\text{number of complex T units}}{\text{number of T units}}$
11	CP/C = $\frac{\text{number of coordinate phrases}}{\text{number of clauses}}$
12	CP/T = $\frac{\text{number of coordinate phrases}}{\text{number of T units}}$
13	CN/C = $\frac{\text{number of complex nominals}}{\text{number of clauses}}$
14	CN/T = $\frac{\text{number of complex nominals}}{\text{number of T units}}$

contact degree; C/S (skewness = 1.999, kurtosis = 7.522, SD = 0.439), T/S (skewness = 2.624, kurtosis = 10.838, SD = 0.199), CP/T (skewness = 1.470, kurtosis = 5.141, SD =0.157) in fiction translations by students with higher contact degree. Mild deviation from normal distribution C/S (skewness = 1.228, kurtosis = 5.602, SD = 0.579) in non-fiction translations by students with lower contact degree; MLC (skewness = 0.954, kurtosis = 4.573, SD = 0.968), T/S (skewness = 1.005, kurtosis = 1.255, SD =0.157) in fiction translations by students with lower contact degree; C/S (skewness = 0.968, kurtosis = 2.070, SD = 0.599), T/S (skewness = 1.060, kurtosis = 0.970, SD = 0.337), CN/T (skewness = 0.873, kurtosis = 2.017, SD = 0.816) in non-fiction translations by students with higher contact degree; MLS (skewness = 1.018, kurtosis = 2.343, SD = 3.649), CP/C (skewness = 0.913, kurtosis = 2.689, SD = 0.088) in fiction translations by students with higher contact degree. The rest features have absolute skewness values $\leq$1.0 and absolute kurtosis values $\leq$2.0, indicating that they are slightly deviated to normal distribution. Given the data distribution characteristics, non-parametric tests, specifically the Mann-Whitney *U* test herein, generally have greater statistical power and have been frequently used in corpus-based translation studies, e.g., studies by Redelinghuys and Kruger (2015), Liu and Afzaal (2021) and Hu et al. (2019).

Next, Mann-Whitney *U* tests were employed to compare the means of 14 syntactic features between translations by students with lower and higher degrees of language contact in fiction and non-fiction genres. Benjamini-Hochberg (BH) correction is added to adjust P values of Mann-Whitney *U* tests address Type I error accumulation. The mean ranks in Mann-Whitney *U* tests indicate whether a linguistic feature is ranked higher or lower in translations by students with higher degrees of language contact when compared with those by students with lower degrees. Null hypotheses in the test are that their mean values will be equal for randomly selected values from n populations (Mann-Whitney *U* tests, $n = 2$). After BH correction, if the adjusted P value of a syntactic feature is < 0.05, the null hypothesis is rejected, indicating that the language change regarding the syntactic feature is statistically significant. In the Mann-Whitney *U* tests, mean ranks of all syntactic features indicate language changes between translations by students with lower and higher degrees of language contact. In the results (Table 4), none of the syntactic features exhibit statistical significance suggesting no significant difference can be observed in translations by students with lower and higher degrees of language contact. The results at individual feature level echo with the premise that studying individual linguistic feature to examine contact-induced language changes is insufficient. The pre-test results empirically validate the necessity of multivariate analysis, because they confirm that contact-induced changes manifest not as discrete, isolated differences in single features, but interrelations among multiple syntactic features, which multivariate methods enables the extraction of latent, underlying dimensions that drive collective linguistic variation.

3.4.2 Multivariate analyses to detect syntactic-level language changes

Multivariate analyses were performed to explore the dimensions that may better explain variations in translation language changes among different samples. The multivariate analysis method adopted in this study was Principal Component Analysis (PCA).

To check the suitability of the data for PCA, this study performed the Kaiser-Meier-Olkin (KMO) test and Bartlett's test of sphericity first. The KMO test measures the proportion of variance among variables for factoring, with values ranging from 0 to 1. A value $\geq$ 0.600 indicates adequate sampling and data suitability for factor analysis. The null hypothesis in Bartlett's test is that the syntactic features as variables are independent, indicating unsuitability of analyzing relations among the data via factor analysis. If the test shows a P value <0.050, factor analysis of the data is informative to explain variations among samples. The KMO value of the data was 0.677, and Bartlett's test had a *P* value of <0.001, suggesting that factor analysis can meaningfully be performed on the data.

Then, PCA was conducted to extract factors. Syntactic features with factor loadings $\geq|0.350|$ are salient, and only salient variables are considered in interpreting factors as dimensions for this research (Kline, 2014). Generally, the minimum number of salient variables to construct a reliable factor is three (Watkins, 2018), but two variables can be acceptable in certain situations if their factor loadings are high ($\geq|0.700|$) (Tabachnick and Fidell, 2018). Factor loading is an indicator suggesting the power of the relationships of each syntactic feature to all factors. As long as factor loadings exceed the threshold of |0.350|, whether positive or negative, the feature is important and should be further analyzed. The next step involves interpreting factors into dimensions. Based on the syntactical features supporting a factor, dimensions to explain language variations in translations are interpreted. Here, features with positive and negative factor loadings indicate that the features occur in a complementary relationship for the same factor. Then,

TABLE 3 Descriptive statistics of syntactic features.

Degree of contact	Genre	Features	Standard deviation (SD)	Skewness	Kurtosis
NF	L	MLS	7.336421	2.106	13.059
		MLT	6.732456	3.370	25.410
		MLC	2.163211	5.474	54.135
		CS	0.579300	1.228	5.602
		VPT	0.572786	0.745	0.665
		CT1_A	0.430178	0.787	0.435
		DCC	0.114212	−0.104	−0.199
		DCT	0.346433	0.754	0.549
		TS	0.303836	0.943	0.728
		CTT	0.181384	0.069	−0.065
		CPT	0.325709	0.627	0.757
		CPC	0.145972	0.257	0.015
		CNT	0.850542	0.567	−0.171
		CNC	0.279514	0.554	1.628
	H	MLS	15.914396	10.512	133.559
		MLT	15.901526	11.674	154.216
		MLC	15.746107	13.787	193.171
		CS	0.598852	0.968	2.070
		VPT	0.593345	0.860	0.990
		CT1_A	0.429126	0.988	1.035
		DCC	0.120951	−0.060	−0.310
		DCT	0.370190	0.878	0.513
		TS	0.337441	1.060	0.970
		CTT	0.191090	0.110	−0.137
		CPT	0.318902	0.406	0.581
		CPC	0.146134	0.013	0.973
		CNT	0.815661	0.873	2.017
		CNC	0.255019	0.294	−0.128
F	L	MLS	3.637122	0.590	0.083
		MLT	2.555595	0.516	0.217
		MLC	0.968227	0.954	4.573
		CS	0.379760	0.873	1.250
		VPT	0.317758	0.461	−0.026
		CT1_A	0.245001	0.611	0.339
		DCC	0.086450	0.021	−0.058
		DCT	0.209627	0.691	0.734
		TS	0.157028	1.005	1.255
		CTT	0.133915	0.160	0.173
		CPT	0.135758	0.334	0.455
		CPC	0.078732	0.259	0.848
		CNT	0.347543	0.574	1.010
		CNC	0.151686	0.115	0.104

(Continued)

TABLE 3 (Continued)

Degree of contact	Genre	Features	Standard deviation (SD)	Skewness	Kurtosis
	H	MLS	3.648636	1.018	2.343
		MLT	2.543193	0.708	0.696
		MLC	1.046934	0.965	1.660
		CS	0.439458	1.999	7.522
		VPT	0.307504	0.465	−0.154
		CT1_A	0.236196	0.736	0.704
		DCC	0.082943	0.250	−0.478
		DCT	0.197610	0.712	0.016
		TS	0.199057	2.624	10.838
		CTT	0.131337	0.342	0.122
		CPT	0.157349	1.470	5.141
		CPC	0.087593	0.913	2.689
		CNT	0.351075	0.769	1.062
		CNC	0.168103	0.499	0.771

factor scores of translation samples along each dimension are computed. Values of all features were transformed into z-scores. Such standardization of data prevents high numerical values from having an inordinate influence on the calculation of factor score. Z-scores of syntactic features (factor loadings $\geq|0.350|$) were summed as a factor score under each factor.

Observing the ranges of factor scores of translations along each dimension via boxplots revealed language changes based on the range overlapping among different translations. Factor scores of expert translations along each dimension were also calculated. The range of expert translations set a reference to observe whether translation changes by students with a higher degree of language contact are acceptable. Dimensions characterizing prolonged and community-acceptable language changes in translations involving intense interactions with different languages are identified as dimensions to quantify language contact in translations at the syntactic level. The normality of factor score data was tested via skewness and kurtosis statistics. Proper statistical tests were conducted to compare factor scores of student translations with higher and lower degrees of language contact along each dimension.

4 Results

4.1 Factors for language changes in translations

PCA generates four factors (Table 5), with factor 1's shared variance of 47.815%, factor 2 of 19.480%, factor 3 of 12.109% and factor 4 of 9.801%. These four factors account for a cumulative 89.206% of total translation variance. The syntactic features all have salient factor loadings $\geq |0.350|$ for at least one factor. In Table 6, factor 1 includes the most salient features (five features with factor loading $\geq |0.350|$), followed by factors 4 (four salient features), and factors 2 (three salient features). Even though factor 3 only has two salient features, both of their factor loadings exceed $|0.700|$, indicating their importance in the factorial structure.

4.2 Interpretation of factors as dimensions

The author interprets the four factors as four dimensions, considering features under each factor, relationships among the features, factor loadings and functions of the features. Factor 1 is interpreted as the dimension of structure diversity. Factor loadings of all five features for factor 1 exceed $> |0.500|$, suggesting they are powerful in the factorial structure. The features, including dependent clause per unit (DC/T, factor loading is 0.962), clause per unit (C/T, 0.950), verb phrase per unit (VP/T, 0.915), dependent clause per clause (DC/C, 0.870), and complex t-unit per unit (CT/T, 0.625), represent various syntactic structures translators adopt in their translations. As all five features have large positive factor loadings on factor 1, co-occurrence of these features indicates a preference for using comparatively diverse syntactic structures in translations. Thus, the dimension underlying factor 1 is interpreted as structure diversity.

Factor 2 is interpreted as the dimension of production length. Three features fall into factor 2, with mean length of t-unit (MLT, 0.968), mean length of clause (MLC, 0.960) and mean length of sentence (MLS, 0.906) being powerful in factorial structure. All three features are associated with translators' production of sentence length at syntactic levels of t-unit, clause, and sentence. Given the features with positive factor loadings under factor 2, this factor describes the language changes in translations regarding the inclination toward producing longer sentences in translations. Therefore, factor 2 is interpreted as the dimension of the production length dimension.

Factor 3 is interpreted as sentence complexity. Even though only two features are under factor 3, their factor loadings indicate importance that should not be ignored. T-unit per sentence (T/S

TABLE 4 Pre-test results.

Genre	Feature	Mean ranks by Mann-Whiteny *U* Test (L vs. H)	Adjusted *P* values by Benjamini-Hochberg correction
NF	MLS	187.190 vs. 213.810	0.196
	MLC	193.840 vs. 207.170	0.084
	MLT	183.390 vs. 217.610	0.634
	C/S	200.450 vs. 200.560	0.992
	C/T	203.280 vs. 197.720	0.689
	T/S	205.690 vs. 195.320	0.658
	DC/C	199.320 vs. 201.680	0.939
	DC/T	202.220 vs. 198.780	0.975
	VP/T	194.900 vs. 206.100	0.882
	CT/T	213.230 vs. 187.770	0.189
	CP/C	195.500 vs. 205.500	0.637
	CP/T	200.820 vs. 200.180	1.028
	CN/C	193.830 vs. 207.170	0.694
	CN/T	199.130 vs. 201.870	0.989
F	MLS	210.200 vs. 190.810	0.439
	MLC	207.360 vs. 193.640	0.731
	MLT	204.630 vs. 196.370	0.700
	C/S	211.650 vs. 189.350	0.302
	C/T	202.800 vs. 198.200	0.921
	T/S	215.260 vs. 185.740	0.154
	DC/C	209.820 vs. 191.180	0.428
	DC/T	208.570 vs. 192.430	0.571
	VP/T	205.550 vs. 195.450	0.669
	CT/T	205.000 vs. 196.000	0.678
	CP/C	199.220 vs. 201.790	0.961
	CP/T	200.800 vs. 200.200	0.995
	CN/C	206.390 vs. 194.610	0.663
	CN/T	207.110 vs. 193.890	0.590

TABLE 5 Top-ranking factors.

Factor	Total	% of variance	Cumulative %
1	6.694	47.815	47.815
2	2.727	19.480	67.296
3	1.695	12.109	79.405
4	1.372	9.801	89.206

0.956) and clause per sentence (C/S, 0.819) function together to mark variation in syntactic complexity. The co-occurrence of the two features measures frequencies of t-units and clauses translators use, signifying language changes of sentence complexity in the translations.

TABLE 6 Rotated factorial structure.

Feature	Factor 1	Factor 2	Factor 3	Factor 4
DC/T	0.962			
C/T	0.950			
VP/T	0.915			
DC/C	0.870			
CT/T	0.652			
MLT		0.968		
MLC		0.960		
MLS		0.906		
T/S			0.956	
C/S			0.819	
CP/C				−0.939
CP/T				−0.934
CN/C				−0.916
CN/T				−0.847

The features for factor 4 describe the frequencies of coordinate phrases and complex nominals in the translation. Both serve the core linguistic functions of systematizing discrete information and condensing distributed meaning. Coordinate phrases structure parallel information for systematic integration via coordinating conjunctions (Temperley, 2005). A complex nominal is a noun phrase with layered modifiers or embedded structures to compress complex semantic relationships into a compact nominal structure (Biber et al., 2000). Both organize fragmented information into a coherent, compact and logically connected whole. Furthermore, the values of their factor loading for factor 4 are negative, suggesting an inverse relationship between their frequencies and the factor. Therefore, it is interpreted as information fragmentation. That is, as the frequency of coordinate phrases and complex nominals increases, the degree of information fragmentation decreases.

Validity of four dimensions used to analyze language contact-induced changes in translations is further verified by comparing the ranges of the factor scores and the tendencies of language changes with the acceptance benchmark set by expert translations. Boxplots for factor scores of translations by experts, and students with higher and lower language contact degrees are displayed in Figure 1, with Figure 1-1 to 4 for non-fiction (NF) genre and Figure 1-5 to 8 for fiction (F) genre.

Boxplots in Figure 1 visualize the distributions of factor scores three types of translations get along four dimensions. Language changes from L samples to H samples are reflected via the differences of interquartile ranges, median lines, and whiskers between the two groups. In both non-fiction and fiction genres, interquartile ranges, median lines and whiskers of all translation samples changed from the students with lower language contact with English and Chinese (L) to those with higher language contact (H), regarding to structure diversity, production length, sentence complexity and cognitive ambiguity. The median lines of L and H samples in two genres differ along all four dimensions, suggesting

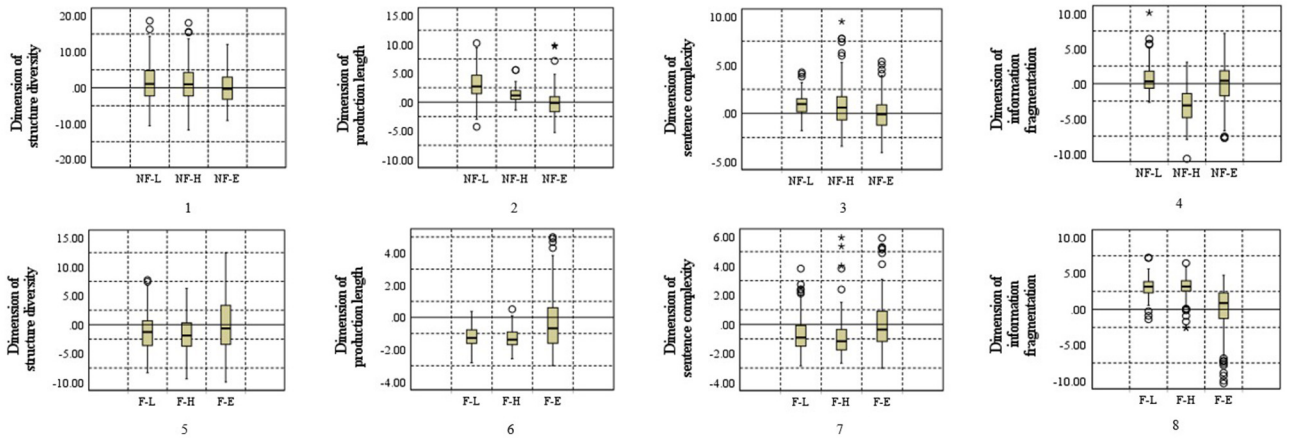

FIGURE 1
Boxplots for factor scores of L, H, E translations along four dimensions.

language changes regarding central tendencies. Language changes reflected in structure diversity focuses on a decrease in the interquartile range from NF-L to NF-H (Figure 1-1) and from F-L to F-H (Figure 1-5). The interquartile ranges indicate that middle 50% of the NF-H samples use less diverse sentence structure than NF-L samples do, and translations in fiction genre share a similar pattern. Compared with NF-L samples, NF-H samples exhibit lower variability in terms of production length, as the interquartile range and whisker length become shorter in Figure 1-2. The spreads of NF-H samples have greater variability regarding sentence complexity and cognitive ambiguity, as the width of the box for their factor scores along these two dimensions increases, together with their whisker lengths (Figure 1-3 and Figure 1-4). As for fiction translations in the dimensions of production length and sentence complexity (Figure 1-6 and Figure 1-7), the variabilities decrease in F-H, as the whisker lengths become shorter. The spread of factor scores for F-H along cognitive load is also less variable due to a narrower interquartile range in Figure 1-8.

As Figure 1 depicts, language changes from L to H samples fall into the ranges of E samples along corresponding dimensions, indicating that the language changes in translations along the four dimensions are acceptable to expert translations across fiction and non-fiction genres. Ranges of NF-H along the production length dimension, and of F-H along structure diversity, production length, and sentence complexity are fully covered by the ranges of the E samples in corresponding dimensions and genres (Figures 1-2,5 to 7). Though the spreads of NF-H along structure diversity, sentence complexity and cognitive ambiguity (Figures 1-1,3,4), and F-H along cognitive ambiguity (Figure 1-8) do not overlap their E sample counterparts, their interquartile ranges, and whisker lengths suggest that over 75% or even higher of their spreads are consistent with those of the E samples. Results indicate that after intensive interactions with Chinese and English, student translations witness language changes, which tend to be consistent with expert translations regarding structure diversity, production length, sentence complexity and cognitive ambiguity, and meet the acceptance benchmark set by the expert translations. This further confirms the effectiveness of using the identified four dimensions to measure contact-induced language changes in translations.

4.3 Quantifying the impact of language contact on translations

By calculating factor scores of each sample along four dimensions and comparing H and L samples in non-fiction and fiction genres, language contact's impact on translation syntax was quantified. The descriptive statistics of factor score data (Table 7) indicated that all of the data did not follow a normal distribution to varied extent, thus Mann-Whitney *U* tests were conducted for comparison. To eliminate the negative impact brought by Type I errors, HB corrections were adopted.

In the non-fiction genre (Table 8), the impact of language contact on language changes in translations is reflected along the four dimensions. Along the dimensions of structure diversity, the mean ranks of L translations (214.220) and H translations (186.780) indicate that translations with greater language contact tend to exhibit significantly less syntactic diversity compared to those with less language contact on average (adjusted $P = 0.024$). Production length, sentence complexity and information fragmentation share such patterns of decreasing from L to H translations, suggesting that translations produced by students with higher degrees of language contact tend to have lower structure diversity, production length, sentence complexity and information fragmentation, compared to translations by those with lower language contact. As all students involved in this study are advanced learners with stabilized language and translation proficiency, such syntactical simplicity does not refer to reduced proficiency but reflects a strategic choice to enhance conciseness and information integration. This is evidenced by the lower mean ranks of NF-H (153.880, 174.140 and 139.290) compared to NF-L (247.120, 226.860 and 261.710), with adjusted *P* values of 0.000. It indicates that more interactions with source and target languages

TABLE 7 Descriptive statistics of factor score.

Genre	Degree of contact	Dimension	Mean	Standard deviation (SD)	Skewness	Kurtosis
NF	L	Structure diversity	1.5016	4.98667	0.370	0.219
		Production length	3.0205	2.57949	0.154	0.090
		Sentence complexity	0.9722	1.40990	4.078	36.798
		Information fragmentation	0.6847	1.86797	1.173	2.835
	H	Structure diversity	1.1582	4.94663	0.540	0.742
		Production length	1.5222	4.84234	13.007	178.717
		Sentence complexity	0.8766	2.04232	1.256	2.326
		Information fragmentation	−3.1283	2.45699	−0.194	0.240
F	L	Structure diversity	−1.1702	3.25081	0.209	−0.009
		Production length	−1.2097	0.62268	0.331	−0.240
		Sentence complexity	−0.6900	1.12856	1.041	1.277
		Information fragmentation	3.0604	1.22656	−0.195	1.417
	H	Structure diversity	−1.4896	3.10170	0.349	−0.268
		Production length	−1.2847	0.60750	0.365	−0.323
		Sentence complexity	−0.8713	1.41074	2.512	10.334
		Information fragmentation	3.0884	1.36902	−1.140	2.971

TABLE 8 Mann-Whitney *U* test results along four dimensions.

Genre	Dimension	Mean ranks by Mann-Whiteny U Test (L vs. H)	Adjusted *P* values by Benjamini-Hochberg correction
NF	Structure diversity	214.220 vs. 186.780	0.024
	Production length	247.120 vs. 153.880	0.000
	Sentence complexity	226.860 vs. 174.140	0.000
	Information fragmentation	261.710 vs. 139.290	0.000
F	Structure diversity	204.960 vs. 196.040	0.440
	Production length	249.610 vs. 151.390	0.000
	Sentence complexity	213.560 vs. 187.440	0.027
	Information fragmentation	280.730 vs. 120.280	0.000

result in language changes of using fewer dependent clauses, clauses, verb phrases, dependent clauses, and complex t-units, decreased length and quantity of clauses and t-units, but more coordinate phrases and complex nominals in translations.

In the fiction genre (Table 8), the four dimensions quantify translation language changes. H group with greater language contact through interacting with source and target languages has significantly decreased factor scores in translations regarding production length (151.390), sentence complexity (187.440) and information fragmentation (120.280), than L group (249.610, 213.560 and 280.730) does (adjusted $P < 0.05$). It reveals shorter length of t-units, clauses, and sentences and more coordinate phrases and complex nominals in translations as language contact increases. In contrast, the dimension of structure diversity does not exhibit a statistically significant decrease with more interactions with source and target languages, as the adjusted *P* value is 0.440. This implies that language changes induced by language contact in translation involve lower frequencies of dependent clauses, clauses, verb phrases, dependent clauses, complex t-units and t-units, but such variations lack statistical validity.

5 Discussion

Patterns in range comparisons occur among L, H and E samples, validating the use of dimensions for distinguishing contact-induced language changes in translations. As reviewed, language changes resulting from contact with different languages over time and accepted in a community are identified as language contact-induced changes (Matras, 2020; Thomason, 2020). The differences between H and L in boxplots indicate that language changes along the four identified dimensions are the distinctive patterns of numerous translation samples with different periods of interaction with source and target languages. Meanwhile, the overlapping ranges of H and E in boxplots reveal that such language

changes along the four dimensions are accepted in a community represented by expert translations.

Distinctions are observed in mean rank comparisons among L and H samples, justifying the usefulness of the four dimensions in quantifying language contact in translations. In the univariate analysis of Mann-Whitney *U* test, statistically insignificant distinctions occur between L and H. However, such distinctions regarding individual features vary, making it difficult to adopt individual linguistic features to observe the impact of language contact. This pre-test result also supports the views of previous scholars who suggest multivariate analysis in language studies because language results from joint functions of multiple features (De Sutter and Lefer, 2020; Kajzer-Wietrzny and Ivaska, 2020; Kruger, 2019).

Dimensions extracted with multivariate analysis describe language changes based on the cooccurrences of multiple features. L and H factor scores along each dimension quantify translation language contact. The statistically significant reduction in structure diversity, production length, and sentence complexity aligns with the hypothesis of translation universals (Baker, 1993). Translation universals describe shared patterns of the translated language compared across genres and in various language pairs. The universal of simplification refers to an inclination to simplify the language of translation (Baker, 1996; Laviosa, 1998). In this study, translations exhibit greater syntactic simplicity as translators' interaction with the source and target languages accumulates. Furthermore, the findings of decreasing information fragmentation via higher frequencies of coordinate phrases and complex nominals are consistent with prior studies on the universal of explicitation. Explicitation is a tendency in translation to render explicit meanings or logical relationships (Toury, 2004; Englund Dimitrova, 2005). Such tendency was also identified in translations under the impact of language contact across genres. As inherent byproducts of the translation process, translation universals (e.g., simplification and explicitation) manifest whenever translation occurs. Collectively, the findings of this study imply that language contact may function as a factor that amplifies these inherent universal tendencies in translated language.

One exception lies in the dimension of structure diversity for fiction translations. Different from language changes above, fiction translations show non-significant change in structure diversity as impact of language contact accumulates. The results imply that genre and language change mechanism are important factors when considering language contact in translations. Genre-specific norm of structural expressiveness in fiction influences contact-induced changes in structure diversity. Non-fiction in this study is sourced from magazines, news reports, and popular science articles. Such informative texts are characterized by factual and objective descriptions, concise structures, organized structure, and high accessibility for readers to convey information, explain concepts, or provide facts about a topic (Valdeón, 2009). In contrast, fiction features narratives of stories, artistic use of language, literary devices of imagery, symbolism, and irony (Lodge, 2012). This genre-specific demand constrains the simplification tendency amplified by language contact, as excessive structural streamlining would compromise the artistic integrity of fictional works. Even with intensified bilingual interaction, translators of fiction are motivated to preserve structural diversity, leading to non-significant statistical changes.

Another factor is the language change mechanism. The results suggest that some language changes are highly noticeable, while some are not. Within the genre of fiction, compared with language changes along the dimensions of production length, sentence complexity, and information fragmentation, changes regarding structure diversity are not very salient. Such varied effects of language contact are consistent with previous studies on contact-induced language change mechanisms (Matras, 2020; Thomason and Kaufman, 2001; Winford, 2005, 2003). Specifically, layered and varied effects of language contact on language changes are found in studies by Armostis and Karyolemou (2022) and Rabanus (2024).

6 Conclusion

This study quantifies language contact in translation with corpus-based linguistic data and four identified dimensions. The data suggests structure diversity, production length, sentence complexity, and information fragmentation validly measure language contact-induced language changes in translation syntax. The results also imply that an amplifying effect of language contact on translation universals of simplification and explicitation. Meanwhile, genre and language change mechanism are important factors when considering language contact in translations. This study provides an alternative perspective to explain language changes in translations and expands the research scope of language contact.

Despite the encouraging findings of multidimensional measurement, there are limitations. First, the time length of language contact in this study is comparatively short. Even though in translation training, translators' interactions with the source and target languages were extremely intensive. Time difference between the two groups of samples was only one year. Time is essential to allow translation changes occurring via language contact. Future analyses across longer periods would complement the current findings. Additionally, translation samples under investigation were sourced from Chinese-English translations by students in China. This study was confined to translations with one language pair and in one translation direction. Third, while rigorous efforts have been made to control confounding variables (e.g., proficiency, institutional context, social contact patterns), the complete exclusion of unobserved variations (e.g., individual differences in learning motivation, extracurricular language exposure, prior translation experience) may introduce subtle biases into the analysis of contact-induced variation, which is worth more attention in future studies. The author hopes this study will stimulate more research on language contact in translations and more inspiring findings based on advances in language contact and translation studies.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

YL: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author declared that financial support was received for this work and its publication. This work was supported by the Fundamental Research Funds for the Central Universities under (Grant D5000250091).

Acknowledgments

The author would like to thank the reviewers for their insightful suggestions.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/flang.2025.1712405/full#supplementary-material

References

Adamou, E., and Shen, X. R. (2019). There are no language switching costs when codeswitching is frequent. *Int. J. Bilingual.* 23, 53–70. doi: 10.1177/1367006917709094

Armostis, S., and Karyolemou, M. (2022). Contact-induced change in an endangered language: the case of Cypriot Arabic. *Languages* 8:10. doi: 10.3390/languages8010010

Baker, M. (1993). "Corpus linguistics and translation studies—Implications and applications," in *Text and technology*, eds. M. Baker, G. Francis, and E. Tognini-Bonelli (Amsterdam: John Benjamins Publishing Company), 233–250. doi: 10.1075/z.64.15bak

Baker, M. (1996). "Corpus-based translation studies: The challenges that lie ahead," in *Terminology, LSP and translation: Studies in language engineering in honour of Juan C. Sager,* ed. H. Somers (Amsterdam: John Benjamins Publishing Company), 175–186. doi: 10.1075/btl.18.17bak

Becher, V., House, J., and Kranich, S. (2009). "Convergence and divergence of communicative norms through language contact in translation," in *Convergence and divergence in language contact situations* (Amsterdam: John Benjamins Publishing Company), 125–152. doi: 10.1075/hsm.8.06bec

Biber, D. (1991). *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., and Finegan, E. (2000). *Longman Grammar of Spoken and Written English* (Vol. 999). Harlow: Longman.

Bisiada, M. (2013). Changing conventions in German causal clause complexes: a diachronic corpus study of translated and non-translated business articles. *Lang. Contr.* 13, 1–27. doi: 10.1075/lic.13.1.01bis

Carter, P. M., Valdez, L. L., and Sims, N. (2020). New dialect formation through language contact: vocalic and prosodic developments in Miami English. *Am. Speech Linguistic* 95, 119–148. doi: 10.1215/00031283-7726313

Chen, J., Li, D., and Liu, K. (2024). Unraveling cognitive constraints in constrained languages: a comparative study of syntactic complexity in translated, EFL, and native varieties. *Lang. Sci.* 102:101612. doi: 10.1016/j.langsci.2024.101612

Chesterman, A. (2004). "Hypothese about translation universals," in *Claims, Changes and Challenges in Translation Studies,* eds. G. Hansen, K. Malmkjær, and D. Gile (Amsterdam; Philadelphia: John Benjamins Publishing Company), 1–14.

Comrie, B. (2000). "Language Contact, Lexical Borrowing, and Semantic Fields," in *Languages in Contact*, *(Brill)*, 73–86. doi: 10.1163/9789004488472_008

De Sutter, G., and Lefer, M.-A. (2020). On the need for a new research agenda for corpus-based translation studies: a multi-methodological, multifactorial and interdisciplinary approach. *Perspectives* 28, 1–23. doi: 10.1080/0907676X.2019.1611891

Durkin, P. (2020). Contact and Lexical Borrowing. *The Handbook of Language Contact* (Hoboken, NJ**:** John Wiley & Sons Ltd), 169–179. doi: 10.1002/9781119485094.ch8

Englund Dimitrova, B. (2005). *Expertise and Explicitation in the Translation Process*. Amsterdam: John Benjamins Publishing Company. doi: 10.1075/btl.64

Glynn, D. (2021). Outline of a theory of non-translation. *Across Lang. Cult.* 22, 1–13. doi: 10.1556/084.2021.00001

Göpferich, S. (2015). *Translation Competence. Interdisciplinarity in Translation and Interpreting Process Research*. Amsterdam: John Benjamins Publishing Company. 72:63. doi: 10.1075/bct.72.06goe

Han, Z. (2004). *Fossilization in Adult Second Language Acquisition* (Vol. 5). Bristol, UK: Multilingual Matters. doi: 10.21832/9781853596889

Hennecke, I., and Wiesinger, E. (2023). Language contact phenomena in multiword units: the code-switching–calquing continuum. *Int. J. Bilingual.* 29:13670069231190209. doi: 10.1177/13670069231190209

Hickey, R. (2010). *Language Contact: Reconsideration and Reassessment. The Handbook of Language Contact.* Hoboken, NJ: Wiley. (1–28). doi: 10.1002/9781444318159.ch

Houlind, M. (2001). Translation and adaptation. *Perspect. Stud. Translatol.* 9, 127–138. doi: 10.1080/0907676X.2001.9961411

Hu, G. (2003). Translation as adaptation and selection. *Perspect. Stud. Translat.* 11, 283–291. doi: 10.1080/0907676X.2003.9961481

Hu, X., Xiao, R., and Hardie, A. (2019). How do English translations differ from non-translated English writings? A multi-feature statistical model for linguistic variation analysis. *Corpus Linguist. Linguist. Theory* 15, 347–382. doi: 10.1515/cllt-2014-0047

Huber, M. (2020). "Diachronic studies of pidgins and creoles: exploring pidgins and creoles over time," in *The Routledge handbook of pidgin and creole languages* (New York: Routledge), 418–433. doi: 10.4324/9781003107224-27

Hymes, D. (2020). Pidginization and creolization of languages: their social contexts. *Int. J. Soc. Lang.* 2020, 99–109. doi: 10.1515/ijsl-2020-2088

Jakobson, R. (1959). *On Linguistic Aspects of Translation*. Cambridge, MA: Harvard UP.

Jin, Y., and Fan, J. (2011). Test for English majors (TEM) in China. *Language Testing*, 28, 589–596. doi: 10.1177/0265532211414852

Jon-And, A., and Aguilar, E. (2019). A model of contact-induced language change: testing the role of second language speakers in the evolution of Mozambican Portuguese. *Plos One* 14:e0212303. doi: 10.1371/journal.pone.0212303

Kajzer-Wietrzny, M., and Ivaska, I. (2020). A multivariate approach to lexical diversity in constrained language. *Across Languages and Cultures* 21, 169–194. doi: 10.1556/084.2020.00011

Kline, P. (2014). *An Easy Guide to Factor Analysis*. New York: Routledge. doi: 10.4324/9781315788135

Kruger, H. (2019). That again: a multivariate analysis of the factors conditioning syntactic explicitness in translated English. *Across Lang. Cult.* 20, 1–33. doi: 10.1556/084.001

Kruger, H., and Rooy, B. (2012). Register and the features of translated language. *Across Lang. Cult.* 13, 33–65. doi: 10.1556/Acr.13.2012.1.3

Laviosa, S. (1998). Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta* 43, 557–570. doi: 10.7202/003425ar

Liu, K., and Afzaal, M. (2021). Syntactic complexity in translated and non-translated texts: a corpus-based study of simplification. *Plos One* 16:e0253454. doi: 10.1371/journal.pone.0253454

Lodge, D. (2012). *The Art of Fiction*. New York: Random House.

Lörscher, W. (2005). The translation process: methods and problems of its investigation. *Meta* 50, 597–608. doi: 10.7202/011003ar

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *Int. J. Corpus Linguist.* 15, 474–496. doi: 10.1075/ijcl.15.4.02lu

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quart.* 45, 36–62. doi: 10.5054/tq.2011.240859

Malamatidou, S. (2016). Understanding translation as a site of language contact: the potential of the code-copying framework as a descriptive mechanism in translation studies. *Target Int. J. Transl. Stud.* 28, 399–423. doi: 10.1075/target.28.3.03mal

Matras, Y. (2020). *Language Contact*. Cambridge: Cambridge University Press. doi: 10.1017/9781108333955

Mauranen, A. (2004). "Corpora, universals and interference," in Translation universals: Do they exist?, eds. A. Mauranen and P. Kujamäki (Amsterdam; Philadelphia: John Benjamins Publishing Company), 65–82. doi: 10.1075/btl.48.07mau

Mellinger, C. D. (2018). Re-thinking translation quality: revision in the digital age. *Target. Int. J. Trans. Stud.* 30, 310–331. doi: 10.1075/target.16104.mel

Muysken, P. (2013). Language contact outcomes as the result of bilingual optimization strategies. *Biling. Lang. Cogn.* 16, 709–730. doi: 10.1017/S1366728912000727

O'Neil, D. (2019). The Middle English creolization hypothesis: persistence, implications, and language ideology. *Stud. Anglica Posnaniensia* 54, 113–132. doi: 10.2478/stap-2019-0006

Paloposki, O. (2011). "Domestication and foreignization," in *Handbook of Translation Studies: Volume 2*, eds. Y. Gambier and L. van Doorslaer (Amsterdam: John Benjamins Publishing Company), 40–42. doi: 10.1075/hts.2.dom1

Pang, S., and Wang, K. (2020). Language contact through translation: the influence of explicitness in English–Chinese translation on language change in vernacular Chinese. *Target* 32, 420–455. doi: 10.1075/target.19001.pan

Pang, S., and Wang, K. (2025). What transcends "translation universals" across time? A multi-dimensional diachronic analysis of the changes in translated and original Chinese register features (1919–2019). *Babel* 71, 449–477. doi: 10.1075/babel.24133.pan

Pešková, A. (2024). Intonational and Syntactic Innovations in a Language Contact Situation: An Explorative Study of Yes/No Questions in Paraguayan Guarani–Spanish Bilinguals. *Lang Speech* 67, 561–589. doi: 10.1177/00238309221082940

Poplack, S. (2020). *A variationist perspective on language contact. The Routledge handbook of language contact*. New York: Routledge, (46–62). doi: 10.4324/9781351109154-5

Pym, A. (2004). Propositions on cross-cultural communication and translation. *Target Int. J. Trans. Stud.* 16, 1–28. doi: 10.1075/target.16.1.02pym

Rabanus, S. (2024). *Two modes of contact-induced change in minority languages: phonology and syntax vs. inflectional morphology. Speakers and Structures in Language Contact: Pluralistic Approaches to Change and Variation*. Berlin: De Gruyter Brill, (31, 65). doi: 10.1515/9783111188348-003

Redelinghuys, K., and Kruger, H. (2015). Using the features of translated language to investigate translation expertise: a corpus-based study. *Int. J. Corpus Linguist.* 20, 293–325. doi: 10.1075/ijcl.20.3.02red

Salamah, D. (2021). Translation competence and translator training: a review. *International J. Linguist. Literat. Trans.* 4, 276–291. doi: 10.32996/ijllt.2021.4.3.29

Sankoff, G. (2004). "Linguistic Outcomes of Language Contact," in *The Handbook of Language Variation and Change*, (Hoboken, NJ: John Wiley and Sons), 638–668. doi: 10.1002/9780470756591.ch25

Shen, L., and Kotze, H. (2025). Linguistic indicators differentiating translated and untranslated diplomatic discourse: a diachronic analysis of the United Nations General Debate (1946–2022). *Lingua* 320:103933. doi: 10.1016/j.lingua.2025.103933

Tabachnick, B. G., and Fidell, L. S. (2018). *Using Multivariate Statistics*. London: Pearson.

Teich, E. (2003). *Cross-Linguistic Variation in System and Text: A Methodology for the Investigation of Translations and Comparable Texts*. Berlin and New York: De Gruyter Mouton. doi: 10.1515/9783110896541

Temperley, D. (2005). The dependency structure of coordinate phrases: a corpus approach. *J. Psycholing. Res.* 34, 577–601. doi: 10.1007/s10936-005-9165-2

Thomason, S. (2020). *Contact explanations in linguistics. The handbook of language contact*, (Hoboken, NJ: John Wiley and Sons), 31–49. doi: 10.1002/9781119485094.ch1

Thomason, S. G. (2022). *Language Contact: An Introduction*. Edinburgh: Edinburgh University Press.

Thomason, S. G., and Kaufman, T. (2001). *Language Contact. Edinburgh: Edinburgh University Press*. doi: 10.1016/B0-08-043076-7/03032-1

Tirkkonen-Condit, S. (2002). Translationese—a myth or an empirical fact?: A study into the linguistic identifiability of translated language. *Target. Int. J. Transl. Stud.* 14, 207–220. doi: 10.1075/target.14.2.02tir

Toury, G. (1995). *Descriptive Translation Studies—And Beyond*. Amsterdam: John Benjamins Publishing.

Toury, G. (2004). "Probabilistic explanations in translation studies: Welcome as they are, would they qualify as universals?," in *Translation universals: Do they exist?* eds. A. Mauranen and P. Kujamäki (Amsterdam: John Benjamins Publishing Company), 15–32. doi: 10.1075/btl.48.03tou

Valdeón, R. A. (2009). Translating informative and persuasive texts. *Perspectives* 17, 77–81. doi: 10.1080/09076760903122611

Van Poucke, P. (2012). "Measuring foreignization in literary translation: An attempt to operationalize the concept of foreignization," in *Domestication and foreignization in translation studies*, (Berlin: Frank and Timme), 139–157.

Verschik, A. (2021). Yiddish–Slavic language contact in multilingual songs: describing deliberate code-switching. *Int. J. Bilingual.* 25, 1696–1717. doi: 10.1177/13670069211036931

Wang, F. (2014). An approach to domestication and foreignization from the angle of cultural factors translation. *TPLS* 4, 2423–2427. doi: 10.4304/tpls.4.11.2423-2427

Wang, Z., Liu, K., and Moratto, R. (2023). A corpus-based study of syntactic complexity of translated and non-translated chairman's statements. *IJTIS* 15, 135–151. doi: 10.12807/ti.115201.2023.a07

Watkins, M. W. (2018). Exploratory factor analysis: a guide to best practice. *J. Black Psychol.* 44, 219–246. doi: 10.1177/0095798418771807

Weinreich, U. (1953). *Languages in Contact, Findings and Problems*. Linguistic Circle of New York.

Wen, Q., and Wang, J. (2009). *Zhongguó dàxuéshēng yinghàn hàn ying kou biyì yǔliàokù (*中国大学生英汉汉英口笔译语料库*) [Parallel Corpus of Chinese EFL Learners]*. Beijing: Wàiyǔ jiàoxué yǔ yánjiu chubǎnshè (外语教学与研究出版社) [Foreign Language Teaching and Research Press].

Winford, D. (2003). *An introduction to Contact Linguistics*. Oxford, UK: Blackwell Publishing.

Winford, D. (2005). Contact-induced changes: classification and processes. *Diachronica* 22, 373–427. doi: 10.1075/dia.22.2.05win

Winford, D. (2010). Contact and Borrowing. *The Handbook of Language Contact*. Hoboken, NJ: Wiley (170–187). doi: 10.1002/9781444338159.ch8

Xiao, R. (2015). *Source language interference in English-to-Chinese translation. Yearbook of Corpus Linguistics and Pragmatics 2015: Current Approaches to Discourse and Translation Studies*, 139–162. doi: 10.1007/978-3-319-17948-3_7