A syntactic complexity analysis of translational English across genres

Jiajin Xu and Jialei Li*

National Research Centre for Foreign Language Education/Artificial Intelligence and Human Languages Lab, Beijing Foreign Studies University, Beijing, China

ORIGINAL RESEARCH PAPER

Received: January 18, 2020 • Accepted: February 4, 2021

© 2021 The Author(s)

ABSTRACT

This study compares the syntactic complexity between translational and non-translational English across four genres (i.e. fiction, news, general prose, and academic prose) and examines the connections between various forms and degrees of syntactic complexity measures and explicitation. Fourteen syntactic complexity indices were examined based on a one-million-word translational English corpus (COTE) and a one-million-word non-translational English corpus (i.e. FLOB), respectively. This study shows that syntactic explicitation in translations varies with the formality of discourse. The most significant complexity difference between translational vis-à-vis non-translational English is found in fiction, which is regarded as the major contributor to translational English syntactic complexity. No significant difference in syntactic complexity was observed between the two types of academic English texts. Translational English news and general prose stand between fiction and academic texts. Translational fiction and news are characterised by more phrasal complexity features such as coordinate and complex nominal phrases, and a key indicator of translational English general prose complexity is subordination. The findings of this study will help students of translation to make informed decisions on the arrangement of sentence structures when given texts of different genres.

KEYWORDS

syntactic complexity, syntactic explicitation, translational English, genre, corpus-based translation studies

* Corresponding author. E-mail: lijialei@bfsu.edu.cn

1. INTRODUCTION

The term “syntactic complexity” defines the variety and sophistication of syntactic structures (Ortega, 2003:492). Variety refers to the categories of syntactic structures involved and the complexity of these syntactic structures. Syntactic complexity often occurs with the term syntactic explication in translation studies during which complex inter-clausal relations in source texts are rendered explicit in translations (Al-Jabr, 2006:203). Without deciphering the inter-clausal relations accurately, it would be difficult for the translator to demarcate sentence boundaries and produce an accurate translation.

Regarding syntactic shifts in translation, a single word may be rendered as a phrase, a phrase as a clause, and a clause as a clause complex in the target texts. The condensed passage is unpacked and redistributed into larger units. Syntactic complexity in translations is influenced by three factors (Al-Jabr, 2006). The first is the inherent lexico-grammatical features of the source and target languages. Another factor is the text genre, as different text genres favour different forms of syntactic units to fulfil their textual functions. For instance, Elness (1984:520) investigated the conditioning of genres on *that/zero* variation and found that *that* is preferred in learned and scientific writing but tends to be omitted in press and Western fiction. The third factor underlying syntactic complexity is the idiosyncratic preferences of authors and translators.

Previous research investigating syntactic complexity in translations has been limited in several ways. First, no previous studies have used a method for providing a comprehensive sketch of the indices of syntactic complexity in translations. Existing literature is restricted to single features such as average sentence length in source texts and translations (Serbina, 2014). Second, most studies on syntactic complexity are restricted to lexical searches for specific words signalling the use of complex sentences (Frankenberg-Garcia, 2019). These studies do not allow us to precisely quantify the syntactic shifts beyond the selected words in translations. Finally, very few studies have systematically examined syntactic complexity and genre variation. This study aimed to fill these research gaps by providing a detailed analysis of syntactic complexity in translational and non-translational English across genres².

In Section 2, we summarise previous studies on syntactic complexity and translation. Section 3 describes the corpora and design of the study. Section 4 explores syntactic complexity across genres in translational and non-translational English. Section 5 discusses the factors that are potentially relevant to syntactic complexity and genre variation. Section 6 is devoted to concluding remarks.

2. PREVIOUS LITERATURE ON SYNTACTIC COMPLEXITY AND EXPLICITATION

Previous literature suggests that syntactic complexity in translations interacts with “explicitation,” one of the universals of translation first proposed by Vinay and Darbelnet (1995:342), which renders implicit elements explicit in the target texts. Although the explicitation hypothesis began to be investigated (Vanderauwera, 1985) before the advent of corpus linguistics, corpus-based translation studies enabled testing of the explicitation hypothesis with large-scale corpus data (Baker, 1993). Syntactic shifts involved dividing one single sentence in the source text into

two or more sentences in the target language or the extension and elevation of phrases in the source text into clauses in the target text (Klaudy & Károly, 2005:17). The structural shift within groups, clauses, and clause complexes may produce more explicit target texts than source texts (Al-Jabr, 2006). Syntactic complexity in translations also interacts with “informational density” (Fabricius-Hansen, 1998), the connection of information conveyed in the texts with the number of sentences, clauses, and words entailed. Fabricius-Hansen (1998) discovered that condensed syntactic structures are expanded into clauses in translations but fail to clarify whether explicitation is concerned with the translation process or the translators’ idiosyncratic styles (Konšalová, 2007). The following paragraphs summarise how syntactic complexity correlates with explicitation and the weaknesses of previous studies, which prompted our research questions.

Much of the literature on syntactic explicitation in translations seem to have been based on specific words that can serve as complex syntax indices. There is extensive literature on connectives signalling syntactic complexity in translations. To investigate the explicitation of clausal relations in translational and non-translational Finnish children’s literature, Puurtinen (2004) compared the relative frequencies of conjunctions, specific adverbs, and relative pronouns. The result of her analysis was inconclusive because it did not reveal more frequent usage of connectives either in translational or non-translational texts and thus failed to provide support for the explicitation hypothesis. In the same vein, Ramm (2004) queried adverbs as markers of relative clauses and found that many relative clauses were upgraded into independent clauses in German–Norwegian translation. Another study on connectives was conducted by Bisiada (2013), who retrieved concessive and clausal conjunctions to examine how the translators coped with complex syntax and discovered that sentence splitting is stronger than sentence combining.

Research on syntactic explicitation in translations also showed that translators prefer more explicit encodings than implicit ways of expression (e.g. the optional complementiser in translations). Olohan and Baker (2000) conducted an empirical study on the retention and omission of reporting verbs *say* and *tell*. The results revealed that the complementiser *that* is more frequent in translational English. However, the complementiser zero is more frequent in non-translational English, strengthening the trend toward syntactic explicitation in translational English. Kruger (2019) took a step further to disentangle three hypotheses – cognitive complexity, pragmatic risk aversion, and source language transfer – behind increased syntactic explicitation in translational English reflected through the *that* complementiser. While these studies enable us to examine the syntactic shifts in translations, they cannot examine the changes beyond the selected linguistic patterns. Additionally, these studies only provide a partial view of syntactic complexity in translations.

Apart from the partial representation, previous studies on syntactic explicitation are also problematic because measures of syntactic complexity lack uniformity. For instance, Frankenberg-Garcia (2009) proposed two methods for measuring text length: word count and morpheme count (i.e. morphemes with equivalent meanings). However, there are certain drawbacks associated with both measures. First, text length is a pluralistic concept incorporating the mean length of clauses, sentences, and T-units (Hunt, 1970:4). Neither word nor morpheme count is valid enough to measure text length. Second, previous standards for the automatic demarcation of sentence boundaries based on full-stops, ignore that full-stops can also appear in abbreviations such as *Mr.* and decimal numbers.

Our paper seeks to remedy these problems by adopting a comprehensive catalogue of syntactic complexity indices first automated by Lu (2010). Our paper covers a broad range of

syntactic complexity indices: length measures, sentence complexity, subordination, coordination, and key constituent units to avoid partial interpretation of syntactic explication based on selected lexis. In terms of sentence demarcation, we used Stanford parser, which automatically recognises sentence-final punctuation marks following a capitalisation letter (marked as ‘ROOT’ in parsed texts). In addition, we examined the syntactic complexity of translational English across genres in detail.

Two research questions are formulated for this study:

1. How does the syntactic complexity of translated English texts compare with non-translational English ones in quantitative terms?
2. What are the connections between various forms and degrees of syntactic complexity measures and explication across genres?

3. DATA AND METHODS

3.1. Corpus profile

Two balanced corpora of present-day English, the Corpus of Translational English (COTE) and the Freiburg-LOB Corpus of British English (FLOB) were analysed for the current study. Both corpora were constructed in accordance with the sampling frame of the Brown Corpus (Francis & Kučera 1964), which consists of text samples of approximately 2,000 words from 15 text types (see Table 1).

The total tokens of the Brown Corpus and those of other member corpora of the Brown family are one million words. COTE and FLOB, in our case, are two comparable Brown family English corpora. COTE¹ is a translational English corpus compiled by Richard Xiao, which contains texts published in the 1990s. The FLOB corpus, developed at Albert-Ludwigs-Universität Freiburg under Christian Mair’s leadership, comprises non-translational British English text samples published in 1991.

The two corpora represent comparable translational and non-translational English texts from the 1990s, respectively. They match perfectly in terms of corpus size, corpus composition, and date of publication. It is worth noting that corpora of one million words can hardly be regarded as large data sources in the age of big data, since mega corpora of ten billion are no longer rare. The one-million-word COTE and FLOB suffice for our purpose of syntactic analysis, in that grammatical variation and change are far less common than those of lexis.

This study will focus on translational language. Therefore, the source language of the translational English texts is critically important in generalising source-target language shifts regarding syntactic complexity. The COTE corpus texts were translated from over 60 source languages¹, such as Afghan, Catalan, Egyptian, Italian, Vietnamese, and Yiddish, to name a few. The top source languages of individual texts are French (17.8%), German (14.8%), Russian (8.7%), Chinese (6.2%), Japanese (5.1%), and Spanish (5.1%). The wide and balanced coverage suggests that the COTE corpus is a representative translational English corpus from the perspective of source languages.

Also, of essential relevance to this study is that major genres, such as press, general prose, academic discourse, and fiction, can be extracted from both corpora for cross-genre comparison of syntactic complexity.

Table 1. Sampling frame of the Brown Corpus

Genre	Sub-genre	Code	# of texts
Press	Reportage	A	44
	Editorial	B	27
	Reviews	C	17
General prose	Religion	D	17
	Skills and hobbies	E	36
	Popular lore	F	48
	Belles lettres, biography, etc.	G	75
	Miscellaneous	H	30
Learned	Learned and scientific writings	J	80
Fiction	General fiction	K	29
	Mystery and detective fiction	L	24
	Science fiction	M	6
	Adventure and western fiction	N	29
	Romance and love story	P	29
	Humour	R	9
Total			500

A comparable corpora based contrastive study is the overarching methodology for this research. Mona Baker used this method in the early 1990s in her TEC (Translational English Corpus) project (Luz & Baker, 2000). Section 3.2 explains the tools and procedures for syntactic parsing and complexity analysis and analyses syntactic complexity and genre variation with independent samples *t*-tests.

3.2. Syntactic parsing and complexity analysis

We analysed syntactic complexity in translations in two steps: syntactic parsing and complexity analysis. Syntactic parsing is a process used to identify the syntactic structures and production units of sentences, and we used the Stanford parser³ (Klein & Manning, 2003) to automate this process. We chose the Stanford parser because this software has built-in functions for segmentation, tokenisation, and POS tagging. Raw texts can be uploaded without pre-processing. Assuming that java is on the path, double-click “lexparser-gui.bat” to access the graphic interface. Then, load the corpus texts for parsing and select englishPCFG.ser.gz from modals jar to parse the sentences. The output consisted of a sequence of parse trees, each representing the syntactic structures of the sentences. Using the parse tree in Fig. 1 as an example, the labels indicate the POS, phrasal, and clausal relations, which form the basis for syntactic complexity analysis in step 2.

Fig. 1. Phrase structure tree for the sentence: *He looked forward to closer working relations between China and Uganda*

After the corpus texts in COTE and FLOB were syntactically parsed, we fed those parsed texts into the BFSU Syntactic Complexity Analyser (Xu & Jia, 2011). The analyser extracted and counted nine production units and syntactic structures. These structures included words, sentences (S), clauses (C), dependent clauses (DC), T-units (T), complex T-units (CT), coordinate phrases (CP), complex nominals (CN), and verb phrases (VP), as shown in Table 2

Table 2. Nine production units and their definitions

Categories	Codes	Definitions
Words	W	Total number of words excluding punctuation marks
Sentences	S	Delimited by punctuation marks that signal sentence endings, annotated with “ROOT” in the parse tree
Clauses	C	A subject and a finite verb
Dependent clauses	DC	Finite adjective, adverbial, or nominal clauses
T-units	T	Main clauses with subordinate or non-clausal structures (Hunt, 1970:4)
Complex T-units	CT	T-units with a dependent clause
Coordinate phrases	CP	adjective, adverb, noun, and verb phrases
Complex nominals	CN	(i) nouns plus adjective, possessive, prepositional phrases, relative clauses, participle, or appositive, (ii) nominal clauses, and (iii) gerunds and infinitives in subject position
Verb phrases	VP	finite and non-finite verb phrases

Table 3. Fourteen syntactic complexity measures and their definitions

Measures	Codes	Definitions
Length measures		
Mean length of clause	MLC	W/C
Mean length of sentence	MLS	W/S
Mean length of T-unit	MLT	W/T
Sentence complexity		
Sentence complexity ratio	C/S	C/S
Subordination		
T-unit complexity ratio	C/T	C/T
Complex T-unit ratio	CT/T	CT/T
Dependent clause ratio	DC/C	DC/C
Dependent clauses per T-unit	DC/T	DC/T
Coordination		
Coordinate phrases per clause	CP/C	CP/C
Coordinate phrases per T-unit	CP/T	CP/T
Sentence coordination ratio	T/S	T/S
Key constituent units		
Complex nominals per clause	CN/C	CN/C
Complex nominals per T-unit	CN/T	CN/T
Verb phrases per T-unit	VP/T	VP/T

Note: The measures are slightly modified from Lu (2010:479).

(cf. Lu, 2010). We then calculated fourteen different syntactic complexity measures, categorised as length measures, sentence complexity, subordination, coordination, and key constituent units (see Table 3), in our corpus texts using these counts. For a detailed analysis of the production units, Tregex⁴ (Levy & Andrew, 2006) was used to query the parse trees with a set of Tregex patterns found in Lu (2010).

3.3. Syntactic complexity and genre variation

Fourteen measures of syntactic complexity in COTE and FLOB were retrieved. Then, we used SPSS to perform independent samples *t*-tests on COTE and FLOB in general and across four genres (i.e., fiction, news, general prose, and academic prose) to determine which of the measures were statistically significant across different genres. We also paid attention to syntactic complexity indices with *P*-values lower than 0.05.

Table 4. The overall syntactic complexity comparison between COTE and FLOB

Measure type	Code	T	Sig.
Length measures	MLC	3.82	0.00
	MLS	5.82	0.00
	MLT	5.87	0.00
Sentence complexity	C/S	3.80	0.00
Subordination	C/T	4.36	0.00
	CT/T	3.00	0.00
	DC/C	4.30	0.00
	DC/T	4.70	0.00
Coordination	CP/C	4.01	0.00
	CP/T	5.80	0.00
	T/S	0.30	0.77
Key constituent units	CN/C	2.99	0.00
	CN/T	4.36	0.00
	VP/T	4.06	0.00

4. RESULTS

4.1. The overall syntactic complexity comparison between COTE and FLOB

An independent samples *t*-test was first performed to compare the fourteen indices of grammatical complexity between COTE and FLOB in general.

Table 4 displays the summary statistics for the overall syntactic complexity across four different genres: fiction, general prose, learned, and press combined. Aside from T/S ($t = 0.30$, $P = 0.77$), thirteen syntactic complexity indices generate *P*-values lower than 0.05. Thus, translational English is statistically significantly more complex than non-translational English. This finding confirms our cross-linguistic intuition that translated language is more verbose. However, does the complexity hold for all genre types? Are longer and more complex texts necessarily more explicit than texts with fewer words? In the pages that follow, independent samples *t*-tests on the features of syntactic complexity mentioned above will be repeated across learned, general prose, press, and fiction to determine whether syntactic complexity and explicitation in translational English is genre-sensitive.

4.2. The syntactic complexity comparison between COTE-Learned and FLOB-Learned

As can be seen from Table 5, only three out of fourteen measures (i.e. MLS, $P = 0.02$; C/S, $P = 0.02$; T/S, $P = 0.02$) have *P*-values below 0.05, and along most of the complexity dimensions, translational and non-translational English are remarkably similar.

Table 5. The syntactic complexity comparison between COTE-Learned and FLOB-Learned

Measure Type	Code	<i>T</i>	Sig.
Length measures	MLC	−0.182	0.86
	MLS	2.44	0.02
	MLT	1.65	0.10
Sentence complexity	C/S	2.32	0.02
Subordination	C/T	1.62	0.11
	CT/T	1.11	0.27
	DC/C	1.37	0.17
	DC/T	1.47	0.14
Coordination	CP/C	0.46	0.65
	CP/T	1.47	0.14
	T/S	2.29	0.02
Key constituent units	CN/C	−0.12	0.91
	CN/T	1.33	0.19
	VP/T	0.91	0.37

A similar trend in syntactic features is intricately linked to the nature of academic writing. The principal goals of academic writing and translation are to achieve “clarity, concision, and correctness” (Herman, 1993). Academic writing is often characterised as an elaborated, compressed informational form of written communication shared in academia across languages and cultures (Biber and Gray, 2016:4). Translations of academic genres are expected to accurately and concisely recreate the reasoning and technical terms in the target language, ensuring that the texts are easier to understand for the readers who cannot read the original document. What is distinctive about the correctness of academic translations is that translators are not expected to discover or refute anything, even if the research papers contain mistakes. It is very unlikely that the translators would make corrections or add paraphrastic comments in their translations. Another explanation may be the prominent role of English as an academic lingua franca. As noted by Ammon (2000:112), “all around the world, English has come to serve extensively for research, or research-related communication.” Since English dominates international academic publishing, many non-native English-speaking researchers who are proficient in English can translate their work themselves. Besides, they may have their drafts proofread by a native speaker before submission to avoid unidiomatic expressions. Their research articles are thus similar to those produced by native researchers.

4.3. The syntactic complexity comparison between COTE-General prose & FLOB-General prose

Likewise, only one measure of syntactic complexity (i.e. DC/T, $P = 0.02$) has a P -value lower than 0.05 (see Table 6). Thus, almost no difference was found between translational and

Table 6. The syntactic complexity comparison between COTE-General Prose & FLOB-General Prose

Measure type	Code	<i>t</i>	Sig.
Length measures	MLC	1.03	0.30
	MLS	0.41	0.68
	MLT	1.59	0.11
Sentence complexity	C/S	1.02	0.31
Subordination	C/T	1.78	0.08
	CT/T	0.34	0.74
	DC/C	2.01	0.05
	DC/T	2.33	0.02
Coordination	CP/C	1.56	0.12
	CP/T	1.78	0.08
	T/S	–1.53	0.13
Key constituent units	CN/C	–0.63	0.53
	CN/T	–0.29	0.77
	VP/T	1.23	0.22

non-translational English general prose in terms of syntactic complexity. Texts under this category encompassed a wide range of subgenres, including Religion, Skills and hobbies, Popular lore, Belles letters, Biography, and Miscellaneous. One of the subcategories that deserve special attention is Religion. Its sacred nature necessitates the most faithful rendition of the original texts and the preservation of the original philosophies and culture as much as possible. Other sub-genres also fall into the overarching category of non-fiction and allow less leeway for translational shifts.

4.4. The syntactic complexity comparison between COTE-Press and FLOB-Press

Table 7 compares the syntactic complexity of COTE-Press with that of FLOB-Press. The results of the thirteen measure types are significantly different across the translational and non-translational texts (all complexity indices, except for C/S, have *P*-values lower than 0.05). The interpretation of these values is that texts in the translational corpus, the COTE-Press, are generally more long-winded, with more coordinate phrases and complex nominals.

The press genre covering reportage, editorials, and reviews, is a typical informative text. Under huge time pressure and responsibilities of the media they represent, the translators, especially those who are novices, may choose to stay close to the surface structures of the original texts or the guidelines provided by the news agencies. This practice avoids the twisting of meanings or producing unintended rendering, thereby contributing to phrasal and clausal complexity in the translations.

Table 7. The syntactic complexity comparison between COTE-Press and FLOB-Press

Measure type	Code	<i>t</i>	Sig.
Length measures	MLC	10.10	0.00
	MLS	8.83	0.00
	MLT	11.20	0.00
Sentence complexity	C/S	−0.16	0.88
Subordination	C/T	3.13	0.00
	CT/T	1.75	0.08
	DC/C	2.15	0.03
	DC/T	2.81	0.01
Coordination	CP/C	8.65	0.00
	CP/T	9.56	0.00
	T/S	−4.79	0.00
Key constituent units	CN/C	10.37	0.00
	CN/T	10.97	0.00
	VP/T	3.32	0.00

In a nutshell, compared with non-translational English newswire texts, translational English ones tend to be structurally ‘elongated,’ or ‘explicit’ in [Blum-Kulka’s \(1986\)](#) term, but hierarchically ‘flattened,’ or ‘simplified’ in [Baker’s \(1993\)](#) term. In addition, the syntactic complexity of translational English in different media sources can be influenced by various reporting styles. For instance, the broadsheets represent the mainstream stance and mentality of the country and enjoy a greater number of audiences compared with tabloids. The translators affiliated with quality newspapers tend to stay close to the original texts in their translation, while those working for tabloids have a greater degree of freedom. Different writing guidelines may cause the translators to deliberately alter the syntactic complexity of their translations, making the target texts more syntactically explicit. The findings of syntactic complexity features of translational English newswire texts should stand the test of more genres and media sources to make fairer claims about the syntactic behaviour.

4.5. The syntactic complexity comparison between COTE-Fiction and FLOB-Fiction

Who has produced more complex translational English texts, since translations of scientific, general, and journalistic texts do not significantly increase the complexity coefficient over non-translational English texts? Our quantitative analysis of the literary texts (see [Table 8](#)) seems to provide some reliable clues to this question. All fourteen syntactic complexity indices of translational English fiction texts are significantly more complex than non-translational English texts. Hence, translational English fiction can be regarded as the main contributor to the syntactic complexity of translational English.

Table 8. The syntactic complexity comparison between COTE-Fiction and FLOB-Fiction

Measure type	Code	t	Sig.
Length measures	MLC	4.89	0.00
	MLS	5.69	0.00
	MLT	5.43	0.00
Sentence complexity	C/S	4.52	0.00
Subordination	C/T	3.91	0.00
	CT/T	3.93	0.00
	DC/C	3.94	0.00
	DC/T	4.22	0.00
Coordination	CP/C	3.25	0.00
	CP/T	4.47	0.00
	T/S	4.24	0.00
Key constituent units	CN/C	4.54	0.00
	CN/T	5.25	0.00
	VP/T	4.66	0.00

The translation of fiction has long been essential to translation enterprises and research. With the mission of transferring both aesthetic and cultural significance from the source language to the target language, it varies widely from the translation of academic and commercial texts. Therefore, the translators must produce texts that carry literary merits of its own, a work that is ‘designed to be read as literature’ (France, 2000:xxi). In the face of linguistic and cultural phenomena that do not have equivalence in the target language, the translators tend to add paraphrastic elements to achieve the effects of the effect in the original text, producing texts with longer and more complex sentences. The sentences in fiction translations become longer, but they have a greater level of explicitness and generate a low cognitive processing load.

4.6. A scrutiny of clausal and phrasal complexity features of translational English

The formality of discourse correlates consistently with syntactic complexity differences between translational and non-translational English, if we regard academic prose, general prose, news texts, and fiction as a cline of formality. In this section, we will retrieve and quantify the salient clausal and phrasal complexity features in the four genres of translational English using Tregex (tree regular expressions), a tool for matching patterns in syntactically annotated texts in a bracketed tree format. Examples in the following section were taken from Press, General Prose, Learned, and Fiction of COTE. To search and account for sentences that contain two or more subordinate clauses, coordinate phrases, and complex nominals, the following regular expressions were used (cf. Lu, 2010):

Sentences with two or more subordinate clauses: ‘SBAR. SBAR’
CP: ADJP|ADVP|NP|VP < CC
CN: NP !> NP [<< JJ|POS|PP|S|VBG |<< (NP \$++ NP !\$+ CC)]

Independent samples *t*-tests in the preceding paragraphs revealed that clausal complexity stands out in General Prose and Press genres. To determine the exact number of subordinate clauses in four genres, an SBAR⁵ search was performed, and the results are summarised in Table 9. These results show that more complex subordination instances were found in translational English than in non-translational English. What stands out in this table is the significant differences between translational and non-translational English in general prose. The differences were represented by clausal subordination, which can be calculated by dependent clauses per clause (DC/C) or dependent clauses per T-unit (DC/T).

Truth is mostly unpalatable, /and when told by him, Kossuth, /it caught a tinge of bitterness from . . . the exile’s soul, /because exile was bitter homelessness and cheerless. (COTE: G38, biography, source language: Hungarian)

This example consists of three T-units that are divided from each other by slashes. Juxtaposed sentence elements can be easily observed and serve as measurements of syntactic complexity in the general prose. The number of dependent clauses in COTE is nearly twice as many as that in FLOB. Clausal complexity, therefore, is probably the most striking contributor in general prose.

If the definition appeals to the deduction in an arbitrary calculus with arbitrary rules of inference, one must keep in mind that the notion of deduction in a logical calculus in its most general form can only be specified within the framework of the notion of algorithm being defined. (COTE: J57, Scientific discourse, source language: Russian)

As shown in the example above, academic texts are more rigid and upright. Researchers make claims and hypotheses, then testify and elaborate. When those claims are cited and reused by others, they again become “facts”. Academic texts, the most informative genre, are characteristically more explicit, precise, and elaborate, which does not need an extended explanation. In regard to abstruse ideas or terminologies, scholars tend to define themselves as passive.

Table 9. Trees and subordinate clause complex in different genres

Genre	Sentences/Trees	Subordinate clause complex ³
COTE: J	988	2,974
FLOB: J	810	1,897
COTE: D-H	2,495	7,583
FLOB: D-H	1,846	3,829
COTE: A-C	612	1,221
FLOB: A-C	707	1,169
COTE: K-R	1,383	3,880
FLOB: K-R	1,011	2,070

The commission said in a statement that its work is continuing and that it will inform the general public of the results. (COTE: A-C50, news, source language: Chinese)

Both phrasal and clausal complexity stand out in the press. The most salient contributor is CN (complex nominals), followed by coordinate phrases and dependent clauses. However, the data in this genre should be interpreted with caution as the syntactic features of news translations are strongly influenced by the nature of different news media.

Her name was Sali (at least, that was what they called her; her full name was Sali-fu-Hamr) when Sali heard that the choice had fallen on her, she was afraid. (COTE: K01, fiction, source language: Italian)

As the main contributor to the syntactic differences between translational and non-translational texts, fiction is more complex in the length of language units, overall sentence complexity, amounts of subordination, amounts of coordination, and phrasal sophistication. Clauses and other units have an equal status in coordination, while in subordination syntactic units serve as the background of other syntactic elements. Literary texts are known for their creativity and imaginativeness, cultural richness, and aesthetic values. Regardless of the source language, defining and/or paraphrasing is necessary to lay bare culture-specific ideas.

In general, academic texts are the most informative among all these genres. The interpretation of specific concepts is embedded in the texts themselves rather than through additional notes in translation. Since literary and journalistic texts are narrative, they rely more on phrasal complexity measures for description and modification. General prose is informative but still has room for translators' creativity.

5. DISCUSSION

This study compares the syntactic complexity between translational and non-translational English across four genres (i.e., academic, general prose, news, and literary texts). Consistent with the previous literature, this study found that in thirteen out of fourteen syntactic complexity indices, overall translational English is significantly more complex than non-translational English. What is surprising is that syntactic complexity is genre-dependent, irrespective of language pairs. Translational English fiction can be regarded as the main contributor to the syntactic complexity of translations and is characterised by complex phrases, clauses, and coordinating conjunctions. Conciseness, clarity, and preciseness are the main features of academic texts. Complex nominals followed by coordinate phrases and dependent clauses are indicators of syntactic complexity in press. Clausal subordination has made the translations of the general prose complex. These phrases, clauses, and conjunctions serve different purposes in different genres.

The similarities in syntactic complexity between non-translational and translational English in academic settings can be partly explained by English as a lingua franca in academia. The high language proficiency of researchers who choose to rewrite or translate the papers themselves and their increased awareness of their papers proofread by a non-translational speaker before submission.

With regard to the press text type, translators, especially those working for national broadsheets, tend to be safe and cling to the established practices and guidelines provided by the

news media they represent, thereby consciously or unconsciously altering the syntactic complexity of their translations. Explication in news translation is closely connected with a risk-management framework (Pym, 2005). The translators are afraid of not getting paid or even losing their jobs. The more complex the source texts, the more likely the translators are to reduce ambiguities and make their translations explicit. They may also choose to translate from different perspectives by rendering some things explicit and others implicit. This practice might be the underlying reason for the increase in phrasal and clausal complexity in news translations. Besides the responsibility of reflecting and shaping public opinion, translators are afraid of distorting the meaning of the original texts and damaging the institution's image or even the country. One of the things that complicate this issue is that in the age of globalisation, news translations are highly likely to be picked up by journalists and reused in the reportage of news in other countries, making translators increasingly anxious. Burdened with huge time pressure, translators cannot sit in front of the computer, weighing their wording and reframing the syntactic structures for hours.

The translation of fiction serves as one of the major contributors to syntactic complexity. Unlike news translators who face great time pressure, fiction translators may spend hours or even days pondering on a single sentence, adding relevant notes, and making adjustments whenever necessary for cultural or aesthetic reasons. As the translation of fiction is less risky and time-sensitive, the translators dare more to 'rewrite' the components of sentences in fiction, leading to greater syntactic complexity.

General prose stays in the middle regarding the informative nature and tolerance level among the four genres. We may note that translators generally prefer complex sentences if the writer aims to provide readers with ideas used in complex syntax and reading experiences (Leech & Short, 2007).

The results of this study suggest that previous studies on syntactic structures might be biased, and a balanced corpus-based study and cross-genre comparison can present a fuller picture. By integrating various indices of syntactic complexity, it can be assumed that longer sentences are not necessarily more complex. In translational English, cutting up embedded sentences into several parts surely makes the texts longer but not necessarily more difficult to understand. With a clearer logic and order, the focus of English translations may become accessible to the audience. There are occasions when the original texts' logic is unclear, so the translators have no choice but to stay close to the structure of the original texts. Generally, coordinate structures and subordination may help the audience tease out the chain of meaning and logic between different components of the original texts.

With a close examination of the indices in the table, we found two major types of structures that complicate sentence structure: coordinate phrases and complex nominals. In translational English, more complex nominals are used instead of verbs. For instance, conjunctions such as "due to the fact that" are found to be more common than "because".

A note of caution is that the results of this study might have been influenced by the varied complexity of subgenres under the four broad categories, the news media the translators work for, and the language proficiency and translating experiences of different translators. For instance, syntactic complexity in the translation of religious texts may remain close to the original texts because of the information they hope to convey and their sacred nature. However, the syntactic complexity of text types such as skills and hobbies as well as biographies may vary because the translators have a greater degree of freedom in translation.

In addition, for the translation of academic texts, the syntactic complexity of hard science articles varies from that of humanities. The translators' language proficiency and experiences can also influence the syntactic structures of translational English even under the same subgenre.

6. CONCLUDING REMARKS

Previous literature claimed that translational English is more complex than non-translational English without considering genre variance. This study evaluated whether the syntactic complexity of translational and non-translational English varies in different genres irrespective of the language pairs involved. The findings suggest that in thirteen out of fourteen syntactic complexity measures, translational English is more complex than its non-translational counterparts. Among the four genres investigated, translational English fiction is regarded as the major contributor to syntactic complexity. A key measurement for general prose is subordination. Fiction and news are characterised by phrasal complexity features such as coordinate phrases and complex nominal phrases. Translators deliberately or unwittingly alter the syntactic complexity of different texts by trying to make some things explicit while others implicit. Texts in academic settings are regarded as the most informative genre, and the translation is strongly influenced by English as a lingua franca. However, general prose is distinguished by clausal complexity.

The findings of this study shed new light on translation training. Awareness of syntactic complexity among different genres may help students prioritise sentence structures when time is limited and alert them to stylistic requirements.

We recommend that further research be undertaken in the following areas: the syntactic complexity of research articles in different disciplines could be explored to determine if the results for academic texts are skewed. Another possible area of future research would be comparing syntactic complexity in translations made by novice translators that found in the work of professional translators.

NOTES

1. The text categories A–C of the COTE corpus were composed of merged short journalistic articles, which means that each of the 88 two-thousand-word texts might contain multiple source languages.
2. Varieties of language are termed differently as registers, genres, text types, domains, and styles, which are delineated in fine detail by some discourse analysts (Leech, 2001; Biber and Conrad, 2009). However, in our paper, we see 'genre' as a general name for the terminological mix, ignoring their nuanced differences.
3. Stanford parser is freely available at <https://nlp.stanford.edu/software/lex-parser.shtml>.
4. Tregex is freely available at <https://nlp.stanford.edu/software/tregex.shtml>.
5. We use the search string 'SBAR. SBAR' to collect two or more subordinate clauses in a sentence in a Tregex. The code SBAR marks subordinate clauses in the Stanford Parser annotated texts.

DISCLOSURE STATEMENT

The authors reported no potential conflict.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the funding by the Beijing Municipal Social Science Foundation project (20YYB013) and the support of the National Research Centre for Foreign Language Education at Beijing Foreign Studies University. The authors are extremely grateful to the editors and the anonymous reviewers for their very helpful comments and suggestions. The authors would also like to thank Professor Eniko Csomay for reading an early version of the article.

REFERENCES

- Al-Jabr, A. (2006). Effect of syntactic complexity on translating from/into English/Arabic. *Babel*, 52(3), 203–221.
- Ammon, U. (2000). Towards more fairness in international English: Linguistic rights of non-native speakers. In R. Phillipson (Ed.), *Rights to language: Equity, power, and education* (pp. 111–116). NJ: Lawrence Erlbaum.
- Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 233–250). Amsterdam: John Benjamins.
- Biber, D., & Conrad, S. (2009). *Real grammar: A corpus-based approach to English*. London: Pearson Longman.
- Biber, D., & Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge: Cambridge University Press.
- Bisiada, M. (2013). *From hypotaxis to parataxis: An investigation of English-German syntactic convergence in translation*. PhD thesis. University of Manchester.
- Blum-Kulka, S. (1986). Shifts of cohesion and coherence in translation. In J. House, & S. Blum-Kulka (Eds.), *Interlingual and intercultural communication: Discourse and cognition in translation and second language acquisition studies* (pp. 17–35). Tübingen: Narr.
- Elsness, J. (1984). That or zero? A look at the choice of object clause connective in a corpus of American English. *English Studies*, 65(6), 519–533.
- Fabricius-Hansen, C. (1998). Informational density and translation, with special reference to German-Norwegian-English. In S. Johansson, & S. Oksefjell (Eds.), *Corpora and crosslinguistic research. Theory, method, and case studies* (pp. 197–234). Amsterdam/Atlanta: Rodopi.
- France, P. (Ed.). (2000). *The oxford guide to literature in English translation*. Oxford: Oxford University Press.
- Francis, W. N., & Kučera, H. (1964). *Manual of information to accompany a standard corpus of present-day edited American English, for use with digital computers*. Providence, Rhode Island: Department of Linguistics, Brown University.

- Frankenberg-Garcia, A. (2009). Are translations longer than source texts? A corpus-based study of explicitation. In A. Beeby, P. R. Inés, & P. Sánchez-Gijón (Eds.), *Corpus use and translating* (pp. 47–58). Amsterdam: John Benjamins.
- Frankenberg-Garcia, A. (2019). A corpus study of splitting and joining sentences in translation. *Corpora*, 14(1), 1–30.
- Herman, M. (1993). Technical translation style: Clarity, concision, correctness. In S. E. Wright, & L. D. Wright (Eds.), *Scientific and technical translation* (pp. 11–19). Amsterdam-Philadelphia: John Benjamins.
- Hunt, K. W. (1970). Do Sentences in the second language grow like those in the first?. *TESOL Quarterly*, 4(3), 195–202.
- Klaudy, K., & Károly, K. (2005). Implication in translation: Empirical evidence for operational asymmetry in translation. *Across Languages and Cultures*, 6(1), 13–28.
- Klein, D., & Manning, C. D. (2003). Fast exact inference with a factored model for natural language parsing. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in neural information processing systems* 15 (pp. 3–10). Cambridge, MA: MIT Press.
- Konšalová, P. (2007). Explicitation as a universal in syntactic De/Condensation. *Across Languages and Cultures*, 8(1), 17–32.
- Kruger, H. (2019). That again: A multivariate analysis of the factors conditioning syntactic explicitness in translated English. *Across Languages and Cultures*, 20(1), 1–33.
- Leech, G., & Short, M. (2007). *Style in fiction: A linguistic introduction to English fictional prose*. London: Pearson.
- Levy, R., & Andrew, G. (2006). Tregex and tsurgeon: Tools for querying and manipulating tree data structures. 5th International Conference on Language Resources and Evaluation (LREC 2006).
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4), 474–496.
- Luz, S., & Baker, M. (2000). TEC: A toolkit and API for distributed corpus processing. In S. Bird, & G. Simmons (Eds.), *Proceedings of exploration-2000: Workshop on web-based language documentation and description* (pp. 108–112).
- Olohan, M., & Baker, M. (2000). Reporting that in translated English. Evidence for subconscious processes of explicitation? *Across Languages and Cultures*, 1(2), 141–158.
- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24(4), 492–518.
- Puurtinen, T. (2004). Explicitation of clausal relations: A corpus-based analysis of clause connectives in translated and non-translated Finnish children's literature. In A. Mauranen, & P. Kuusimäki (Eds.), *Translation universals. Do they exist?* (pp. 165–176). Amsterdam/Philadelphia: John Benjamins.
- Pym, A. (2005). Explaining explicitation. In K. Károly, & Á. Főris (Eds.), *New trends in translation studies: in honour of Kinga Klaudy* (pp. 29–43). Budapest: Akadémiai Kiadó.
- Ramm, W. (2004). Sentence-boundary adjustment in Norwegian-German and German-Norwegian translations: First results of a corpus-based study. In K. Aijmer, & H. Hasselgard (Eds.), *Translation and corpora* (pp. 129–147). Gothenburg: Acta Universitatis Gothoburgensis.
- Serbina, T. (2014). *Sentence splitting in the translation pair English-German*. Paper presented to the 4th Using Corpora in Contrastive and Translation Studies Conference. Lancaster University. 24–26 July 2014.

- Vanderauwera, R. (1985). *Dutch novels translated into English: The transformation of a 'Minority' literature*. Amsterdam: Rodopi.
- Vinay, J.-P., & Darbelnet, J. (1995). *Comparative stylistics of French and English: A methodology for translation*. Philadelphia: John Benjamins. (Translated by Sager, J. C. & Hamel, M.-J.).
- Xu, J., & Jia, Y. (2011). *BFSU syntactic complexity analyzer 1.0*. Beijing: National Research Centre for Foreign Language Education, Beijing Foreign Studies University.

Open Access. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium for non-commercial purposes, provided the original author and source are credited, a link to the CC License is provided, and changes – if any – are indicated.