Published by
Centre for Languages and Translational Studies

Pakistan Journal of Languages and Translation Studies
ISSN (Print) 2410-1230
ISSN (Online) 2519-5042
Volume 11 Issue 1 2023
Pages 80-100

Open Access

Simplification Reviewed: A Corpus Based Study of Urdu Language Specialized Translated Genres

Publication Details

Paper Received:
March 15, 2023

Paper Accepted:
May 15, 2023

Paper Published:
June 30, 2023

Humaira Khurshid
PhD Scholar, Department of Applied Linguistics, Government College University Faisalabad
humaira.khurshied@yahoo.com

Prof. Dr. Muhammad Asim Mahmood
Government College University Faisalabad
masimrai@gmail.com

Dr. Rashid Mahmood
Government College University Faisalabad
ch.raashidmahmood@gmail.com

Abstract

Based on the research paradigm of translation universals and analyzed within the methodological framework of corpus linguistics this research has investigated the relevance of the translation universals hypothesis of simplification for the English, Urdu language pair genre specific parallel, comparable corpus (USTC). The corpus compiled for this study consisted of three genres i.e legal statutes (LSC), newspaper articles (NSC), and academic prose (ASC). The analytical framework to investigate the translation specific linguistic traits was adapted from Zanettin(2013). The analysis has

Published by Licensee CeLTS. Copyright: © the author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

revealed that the translated Urdu component (TUT) in USTC has shown a de simplifying tendency for all the linguistic measures. The findings can be attributed to the genre specific nature of the specialized translated texts and the typological features of the Urdu source target language. The study would add to the existing knowledge of translators, post-editors, trainers, contrastive genre analysts, terminology banks, and pedagogy developers in the domain of translation studies and machine translation.

Keywords: translation universals, simplification, specialized translation, genre specific, parallel and comparable corpus, target language

1. Introduction

The product of translation process has been termed as 'third code' since pre corpus and pre descriptive era Frawley (2000). The term 'third code' implies that the product owes linguistic features distinct from both source and target texts. The factors that contribute to make it a third code are; the level of translators' expertise, target audience or readers, interference of both source and target language as well as socio-cultural factors in which the original text was produced and in which the translated text would be received. The characteristic linguistic features that have given translation the status of a third code are called translation universals and these translation universals have been the subject of inquiry since the 1980s. The acceptance of translation as a third code, heralded a research convention to verify the translation universals hypothesis. This convention to study translation universals investigated the features of simplification, explicitation, normalization, interference, and leveling out in different languages and text types. This study attempted to explore the relevance of simplification hypothesis has been prompted by the following questions:

The analysis in this research was motivated by the following research questions:

1. Do the genre specific specialized translated texts constituting USTC provide evidence to support simplification hypothesis?

2. Have genre specific linguistic traits or general language specific features informed linguistic choices in the case of translated texts in USTC and its constituent sub-corpora i.e NSC, ASC, and LSC?

2. Literature Review

2.1. Translation universal of simplification

The theoretical basis for the simplification hypothesis was propounded on the contention that translators play the role of mediators in the process of communication, and they continuously try to bridge the gap between the source and target content readers. During this mediation, the translators reproduce the text in such a way as to maximize the communicative effect with the least effort. This tendency has been termed as simplification and it has been reckoned among the translation universals by Baker (1996). Toury (1991) considered it the "most persistent and unbending norm". Translators purposefully simplify the lexical, syntactic, and stylistic structure of language to come up to the expectation of target readers or what Toury (1991) has termed as "sociolinguistic norms" of expectancy. According to Baker (1996) simplification is inherent in the very process of translation and it is ensued in the variation in the two linguistic systems to be adapted during the process of translation. Chesterman (2004) has categorized simplification among "T universals by considering it result of the target language influence. Bernardini and Zanettin (2004) recognized it as a strategy adopted to deal with the differences between the target and source language differences and it can be traced during the analysis of the parallel corpus.

2.2. Empirical Studies on Simplification

As far as the empirical research on the simplification hypothesis is concerned it dates back to Blum-kulka and Levenston's (1983) study based on the lexical simplification of Hebrew English language pair specific content and the findings supported the hypothesis. Then followed a stream of studies dealing with the subject matter of simplification. Vandereuweraa (1985) dealt with syntactic simplification in terms of the conversion of paratactic clauses into hypotactic ones. Shlesinger (1989) analyzed simplification in court interpretation and found that the interpreted content was made free from all phatic expressions, inappropriate comments, and any reference made to the interpreter self. Braithwait (1996) confirmed simplification hypothesis in terms of lexical density and information load. Laviosa (2002) again confirmed the hypothesis by measuring lexical density, type-token ratio, and comparative proportion of high and low-frequency words. Hu (2007), Wen (2009), and Xiao and Dai (2010) studied different versions of translated Chinese and confirmed the simplification hypothesis. Hu (2007) studied translated contemporary Chinese fiction while Wen (2009) studied Chinese detective fiction and both studies verified simplification hypothesis and found that translated Chinese had lesser mean sentence length, lower type-token ratio, and lexical density as compared to non-translated Chinese language texts. However, simplification like the normalization hypothesis too has been contradicted on empirical grounds. Mauranen (2004), Jauntunen (2001, 2004), and Braithwait's (1996) research dealing with lexical and syntactic simplification has shown contradictory to hypothesis findings.

2.3. Research Work Produced in the Domain of Translation Universals in Pakistan and the Existing Gap

As far as Urdu language is concerned, most of the work deals with literary stylistic analysis or thematic issues raised in epoch making literary works and their translated counterparts. Asghar et al., (2020), Hussain (1996), Anwar et al., (2020), Shakur (2008), Hashmi (2014), Shahzad et al., (2020), Nazir (2019), and Qasim et al., (2020), Mahmood and Azhar (2020), Yaqub and Shakir (2019), Iqbal (2020) are a few examples of this dominance of literary texts in the research domain of translation studies. The proportion of research work dealing with linguistic issues related to the translation process is very low. Ali et al., (2019), Qasim et al., (2021), Khurshied and Mehmood (2022), khurshied and mehmood (2020) are a few instances of work produced in this regard.

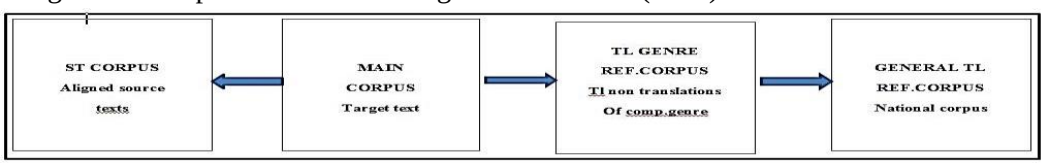
Hence a wide gap exists, no previous study has been conducted to analyze specialized or nonliterary texts translated from and into the Urdu language. Rather specialized translated texts had been an outcast in the research world dealing with Urdu language translational analysis. This research has endeavored to bridge this gap. The subject matter of analysis is specialized, parallel, comparable corpus comprising three genre specific texts i.e. legal statutes, academic prose, and newspaper articles. Not only the subject matter of research is new and unattended earlier, the corpus design itself comprising simultaneously of parallel, comparable, and general reference components is also so far an unexploited research paradigm in the body of Urdu language translation research.

3. 3. Materials and Methods

3.1. Corpus architecture

This study has investigated multi-faceted genre specific translated texts falling under the umbrella term of specialized or technical translation Rogers (2018), to put into practice claims put forward by Sutter & Lefer (2020) and Biel (2017) in favor of empirical translation studies, having a broader methodological perspective along with an in-depth theoretical underpinning. The genre specific corpus compiled for this study comprises of the following structure.

Figure 3.1:Corpus architecture design based on Biel (2017)



The data sources, size and the constituents of the three subcorpora are given below:

3.2.1. NSC and its Constituent Components

The parallel component for newspaper article genre was collected from 'The Financial Times' source English text SET(N) and 'The Jang' translated Urdu text (TUTN). Urdu newspaper 'The Jang'publishes articles translated from 'The Financial Times' on weekly basis. Again, web has been the source for the data and e papers of the above two newspapers were accessed having prior permission for data extraction.

Table3.1 : Description of NSC

Corpus	SET(N)	TUT(N)	CUT(N)
Tokens	89389	105260	102070
Types	10574	9476	9851

The comparable component for the newspaper articles (CUTN) has been again sought from 'The Jang' e-paper. Nontranslated originally produced articles in Urdu language approximating in length to parallel texts, published in 'The Jang' during last two years were extracted.

3.2.2. ASC and its Constituent Components

As far as the parallel corpus for academic prose is concerned two texts served the purpose of data; the first one was Iqbal's (2013) 'Reconstruction of Religious Thought in Islam' and the second one was the 'Education and Social Order' by Russell (2009). Only the first six chapters from the later text were included since they were sufficient to provide the required data limit. Again web has provided the source English text for academic prose SET(A) and target Urdu text TUT(A) data in the case of academic parallel sub-corpus. The comparable component for academic prose CUT(A) was accessed from the Urdu Web. The text titled 'Ab To Jag Jao' served the purpose of data being compatible in genre, size, time, and mode of production.

Table3.2.: Description of ASC

Corpus	SET(A)	TUT(A)	CUT(A)
Token	111506	104553	108,591
Types	13,244	8615	9415

3.2.3. LSC and its Constituent Components

The corpus for the legal genre was compiled from the official site of the National assembly of Pakistan. The site has both source English legal text SET(L) and the target Urdu text TUT(L) of the constitution of Pakistan. Moreover, the text of the Punjab local government act issued on July 25, 2016 (source English and target Urdu) was also included in the legal parallel corpus to meet the required size of the corpus component. As far as the comparable legal CUT(L) corpus component is concerned, it was formed by the Punjab Pure Food Act 2011, again accessed from the web from the official site of the Department of Parliamentary Affairs and Law. The comparable legal component was compatible with the parallel legal component in terms of genre, text size, time, and mode of production.

Table 3.3: Description of LSC

Corpus	SET(L)	TUT(L)	CUT(L)
Tokens	135477	136493	136388
Types	6233	6066	7212

3.2.4. REF (GEN) Corpus

Reference corpus in its conventional sense consists of a larger databank of any language representing all its varieties and genres. It is also alternatively termed as general or national corpus. No such corpus for the Urdu language has been collected yet. Hence the larger reservoir of Urdu language from the Sketch Engine was accessed.

The Urdu corpus available at Sketch Engine is representing its multiple varieties and genres and the site also provides lexico-grammatical analytical assistance required for the purpose of this study.

Table 3.4:Description of REF(GEN) Corpus Urdu Web 2018 (ur TenTen 18)

Corpus	REF(GEN)
Tokens	245656128
Types	274,682

3.3. Analytical Design

The analytical design proposed by Zanettin (2013) in its adapted form was employed to extract the characteristic lexico-grammatical features by means of which simplification hypothesis has been investigated.

Table 3.5: Descriptive Features and Linguistics Operators adapted from Zanettin 2013

Features	Formal Operator	
	Lexis	Syntax
Simplification	St. T/T ratio	Average sentence length.
	Lexical density	

3.4. Data Tagging

To find the distinct linguistic categories that can serve as operators for translation universals, the data was tagged to retrieve POS information. The source English language texts for all the three genres were tagged by the CLAWS 5 POS tagger designed by the University of Lancaster. The Urdu data for parallel and comparable categories was tagged by using Urdu POS tagger devised by the Centre For Language Engineering KICKS, UET, Lahore, Pakistan. The reference corpus was available in

tagged form on Sketch Engine by Urdu Shallow Parser (IIIT) and the tag set is termed as "UNIFIED PARTS OF SPEECH (POS) Standard in Indian languages." Since three different taggers have been used to tag the data, the refined compatible features for parts of speech were extracted by manual scrutiny from the broader categories in the cases where taggers had not tagged the equivalent categories for parts of speech.

3.5. Analytical Tools

To extract data for linguistic operators to determine the translation universals, the raw frequencies for the tagged data from the corpus were further analyzed by WordSmith devised by Oxford University Press. The Wordlist tool was used for statistical findings i.e type-token ratio and average sentence length.

4. Data Analysis

Keeping in view the convention set by the earlier studies on simplification, the following linguistic operators have been investigated: lexical density or information load of the corpus with a focus on the relative proportion of content words to total words, mean sentence length, and standard type-token ratio. Two important descriptive measures used to report findings in this research are normalized frequencies and their graphical presentation. The raw frequencies for different linguistic variables as indicators of the translation universals, were normalized per thousand words following Biber (1988) to develop for further analysis. . Below is given a detailed analysis of the three sub corpora:

4.1. Analysis for simplification hypothesis in NSC

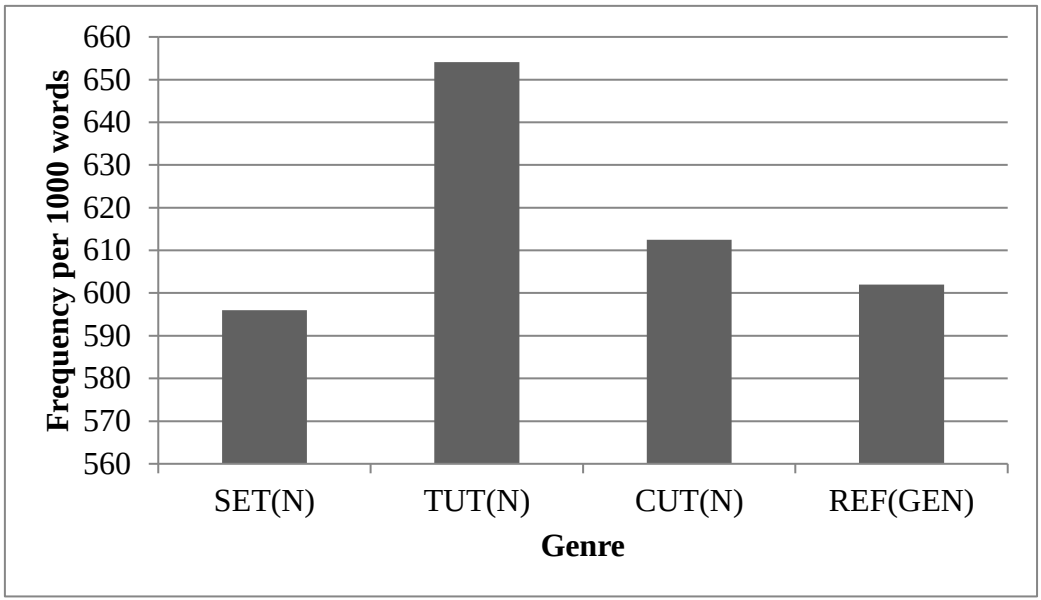
The detail of analysis and results for all analytical measures is given below:

4.1.1 Results for lexical density measure in NSC

Table 4.1: Normalized Frequency of Content Words in NSC

SET(N) Content Words	TUT(N) Content Words	CUT(N) Content Words	REF(GEN) Content Words
595.97	654.113	612.48	601.99

Figure 4.1: Normalized Frequency of Content Words in NSC



The frequency of content words has significantly increased in the TUT(N) sub-corpus component as compared to non-translated source SET(N) component. . The CUT(N) and REF(GEN) components have a significantly lower frequency of the content words in comparison to translated one.

Table 4.2: Standard Type Token Ratio in NSC

SET(N)	TUT(N)	CUT(N)	REF(GEN)
46.32	43.42	44.89	44.34

The results regarding the second measure for the assessment of the simplification hypothesis in the NSC have supported the research hypothesis since the STTR in the TUT(N) is considerably lower not only in comparison to the source text but also in comparison to the other two non-translated corpus constituents i.e CUT(N) and REF(GEN) corpus. However, it is noticeable that though STTR is higher in CUT(N)

and REF(GEN) components it is lower than the STTR of SET(N) and SET(N) has the highest STTR.

4.1.3. Results for Mean Sentence Length Measure in NSC

Table4.3: Mean Sentence Length in NSC

SET(N)	TUT(N)	CUT(N)	REF(GEN)
26.30	*31.68*	*25.77*	*30.31*

In the NSC mean sentence length in translated component TUT(N) has been higher as compared to the source text that indicated contrary to research hypothesis findings. While in the comparable component CUT(N) mean sentence length is lower than the translated one and the frequency is almost equal to that of the source text. In the REF(GEN) component, the mean sentence length is higher than that of comparable but still lower than translated one. It can be said alternatively that the translated component has the highest mean sentence length in the NSC.

4.2. Analysis for Simplification Hypothesis in ASC

Below is given a detail of the analysis for ASC.

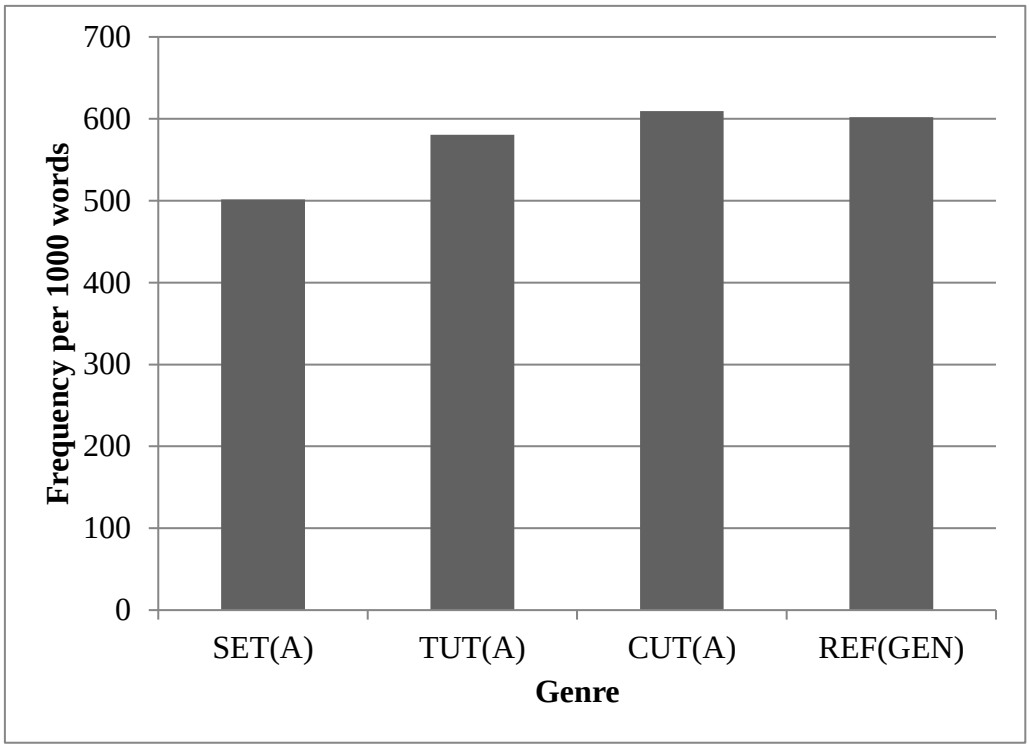
4.2.1. Results for Lexical Density Measure in ASC

Table 4.4: Normalized Frequency of Content Words in ASC

SET(A) Content Words	TUT(A) Content Words	CUT(A) Content Words	REF(GEN) Content Words
501.64	580.6	609.46	601.99

The analysis displays an overuse of lexical items in the TUT(A) as compared to the source text. The same tendency is displayed by the CUT(A) and REF(GEN).

Figure 4.2: Normalized Frequencies of Content Words in ASC



4.2.2. Results for Standard Type Token Ratio Measure in ASC

Table 4.5: Standard Type Token Ratio in ASC

SET(A)	TUT(A)	CUT(A)	REF(GEN)
27.58	31.61	27.96	44.34

The findings for the STTR measure in the ASC do not support the research hypothesis rather they have retained the null hypothesis. Since STTR in the TUT(A) has been higher than the SET(A). In ASC, source text had the lowest STTR. The other two non-translated Urdu language corpus components i.e CUT(A) and REF (GEN) have a higher STTR.

4.2.3. Results For Mean Sentence Length Measure in ASC

Table4.6 Mean Sentence Length in ASC

SET(A)	TUT(A)	CUT(A)	REF(GEN)
25.13	34.73	31.37`	30.31

The translated text has the highest mean sentence length in the ASC. Not only it is higher in comparison to the source text but also higher in comparison to the other two corpus components i.e comparable and reference corpus. Hence, the translated text has not simplified in terms of mean sentence length. Moreover, the same tendency of higher mean sentence length dominates in the other two Urdu language sub-components in ASC.

4.3. Analysis for Simplification Hypothesis in LSC

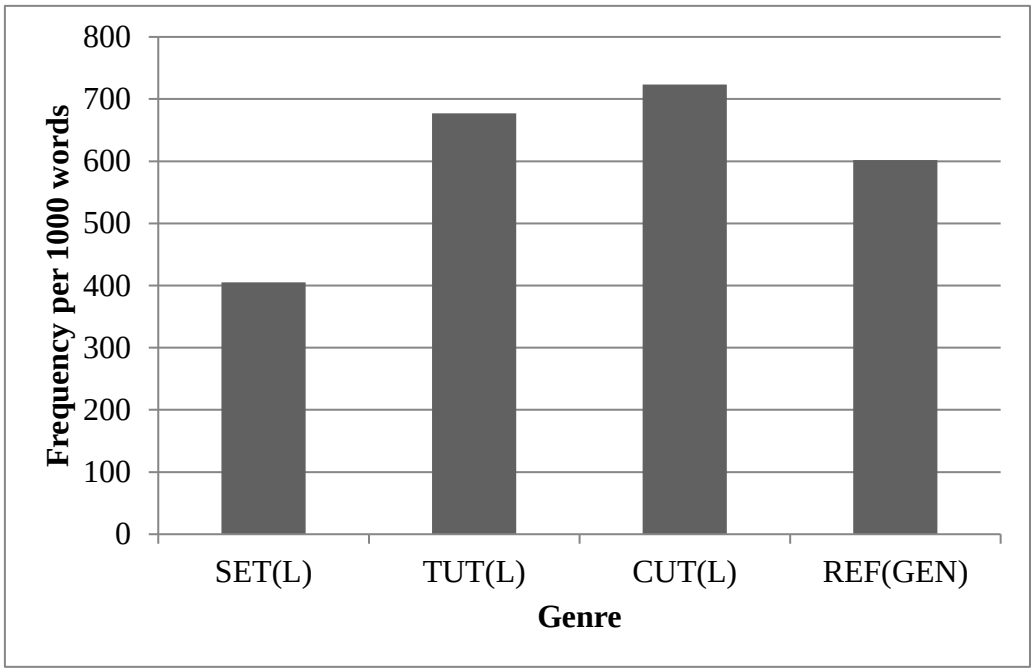
The detail of analysis and results for all analytical measures in LSC is given below:

4.3.1. Results for Lexical Density Measure in LSC

Table 4.7: Normalized Frequency of Content Words in LSC

SET(L) Content Words	TUT(L) Content Words	CUT(L) Content Words	REF(GEN) Content Words
405.22	677.24	723.56	601.99

Figure 4.3: Normalized Frequency of Content Words in LSC



The analysis demonstrates an overuse of lexical words in the TUT(L) as compared to its source text. A similar tendency of overuse of lexical words is also observable in CUT(L) and REF(GEN).

4.3.2. Results for Standard Type Token Ratio Measure in LSC

Table 4.8: Standard Type Token Ratio LSC

SET(L)	TUT(L)	CUT(L)	REF(GEN)
41.93	42.46	42.94	44.34

As far as the results for the third sub-corpus i.e LSC are concerned again source text contains the lowest STTR and the frequency is higher in the translated text TUT(L),

and the comparable text CUT(L) contains slightly lower STTR as compared to the translated text while the general reference corpus has the highest STTR. The findings are suggestive of contrary hypothesis results. Moreover, a tendency for higher Standard Type Token Ratio in Urdu language components has been recorded.

4.3.3. Results for Mean Sentence Length Measure in LSC

Table 4.9: Mean Sentence Length Measure in LSC

SET(L)	TUT(L)	CUT(L)	REF(GEN)
22.08	26.78	32.25	30.31

As far as the findings for the LSC are concerned the translated text TUT(L) has a higher mean sentence length as compared to the source text. Moreover, this trend in the increase of mean sentence length is also followed by the CUT(L) and REF(GEN) components. The findings have not provided evidence to support the simplification hypothesis. Moreover, the trend of higher mean sentence length is also followed by CUT(L) and REF(GEN) sub-scorpus components.

5. Findings and Discussion

5.1. Findings for Simplification in NSC

The TUT(N) has not simplified in terms of lexical density and mean sentence length measures. However, it has simplified as far as the STTR measure is concerned. Results of the first simplification evaluative measure i.e proportion of content words have not favored the hypothesis. The tendency to have a higher proportion of content words is also followed by genre-specific non-translated CUT(N) and the REF(GEN) component has a comparatively lower frequency of content words. As far as the results for the second measure i.e STTR are concerned, they have provided evidence in support of the simplification hypothesis. Moreover, the tendency to have a lower type token ratio has been favored by both the genre specific CUT(N) and REF(GEN) non-translated components, it is rather a tendency adhered by the Urdu language. The simplification hypothesis has not been supported by the third simplification measure i.e mean

sentence length, again similar trend has been reflected by the comparative and reference components.

5.2. Findings for Simplification in ASC

Regarding the findings for the analysis of simplification measure lexical density, the null hypothesis has been retained since the normalized frequency of content words is significantly higher in the translated text. No significant difference has been recorded in the genre specific non translated comparative component in the use of content words when compared to the translated text. A similar tendency of a significant increase in content words has also been reported in the CUT(A) and REF(GEN) components. The findings for STTR have also retained the null hypothesis with a similar trend of a higher STTR in the CUT(A) and REF(GEN) components. Mean sentence length measure has also retained null hypothesis. The TUT(A) component has not simplified in terms of lexical density, STTR, and mean sentence length. Moreover, the non-translated Urdu language components CUT(A) and REF(GEN) have shared the tendency of higher lexical density, STTR, and mean sentence length.

5.3. Findings for Simplification in LSC

As far as the findings for the LSC sub-corpus are concerned the lexical density reported contrary to hypothesis findings since the translated component has a higher frequency of content words and translated text is lexically dense as compared to the source text. A similar higher density level is followed by comparable and reference components. The TUT(L) component has not simplified in terms of STTR and the tendency of a higher STTR is visible in the CUT(L) and REF(GEN) components too. The results for the mean sentence length measure have not supported the research hypothesis and the mean sentence length in the TUT(L) is higher than the SET(L) component. The trend is also noticeable in the genre specific non-translated and reference components. Hence the TUT(L) component has not simplified in terms of any measure investigated in this study.

5.4 Discussion

For all the measures investigated in the present study i.e lexical density, mean sentence length, and STTR, the TUT component has not provided evidence to support the simplification hypothesis. The investigation has revealed that the translated Urdu corpus under investigation is lexically dense or has a higher load capacity, it is lexically more varied. Moreover, these tendencies of the semantic richness of the TUT corpus component under study are also a common feature of the CUT and REF(GEN) components in USTC.

The findings in the case of the simplification hypothesis have clear traces of the target language influence, particularly in the case of lexical density and mean sentence length measure, the bars in the figures representing the normalized frequency of content words in the TUT component are skewed towards the CUT and REF(GEN) components as compared to SET component in all the three sub-corpora. Williams(2005) in her study on translated English and the translated French language has found target language-oriented tendency in her analysis for lexical density, mean sentence length, and STTR. According to Williams (2005) mean sentence length for the English-translated text was higher while it was lower for the French translated texts. French translated texts were having higher lexical density and STTR while English translated texts were having lower lexical density and STTR. The findings in the case of the present study for the simplification hypothesis have reiterated Williams' (2005) contention that the target language does play a detrimental role in linguistic pattern formation rather than the translation process.

At this point, it should be kept in the mind that translation universals hypothesis was propounded having repeated research on the English language translated texts. The hypothesis has not taken into account the typological differences in different languages and the Urdu language is typologically different from its European cousins as established by Kachru (2018). These very typological differences are visible in the translated Urdu language texts as in the case of the French translated texts. Wang and Qin (2010) too observed the language-oriented tendency of the translated content. Wang and Qin (2010) found that the average sentence length in translated Chinese is higher than the native English. Moreover, translated Chinese sentence is longer than the non-translated Chinese sentence. Chan (2007) found a similar tendency of higher mean sentence length in translated Chinese language legal texts and Hu (2006) found it in Chinese translated novels.

Genre specific linguistic features and syntactic tendencies too play a significant role in determining the linguistic patterns of any translated text. This very fact is observable not only in this study but also in the earlier studies. Liu and Afzal (2021) in their study on COCE and FLOB found that translated news is homogenous to non-translated news in sentence structure. Translated academic prose was less complex than its non-translated counterparts in syntactic structure, while translated fiction was found more complex regarding sentence structure in comparison to non-translated fiction. In the case of the present study, the three sub-corpora too have divergence in terms of findings that can be attributed to the genre specific nature of the sub-corpora. The LSC and ASC translated text have been unbending in terms of all the measures for simplification hypothesis while the NSC has been prone to adopt the norms of translation in terms of STTR.

5.5. Limitations of the Study:

The findings of the study are language pair, text type and genre specific. The corpus comprises of three genres from the domain of specialized translation i.e newspaper articles, legal statutes, and academic prose. The findings in this study are confined to the linguistic features specified in the analytical design adapted from Zanittin (2013) outlined in the methodology section. study has not undertaken any investigation of the cognitive process informing the linguistic choices in the case of translated texts investigated in this study, nor has it attempted to investigate the process involved in the translation formation.

5.6. Practical Implications of the Research

The study would provide fruitful insight for the pedagogical development both in the domain of specialized translation studies and genre analysis, which is a dire need in the educational system of an underdeveloped country like Pakistan and under-resourced languages like Urdu. The integration of genre knowledge both in source and target language conventions with translator competence in specialized translation training would bring translation practice in the Urdu language on par with the international practice. Since the translation competence models like EMT and PACTE consider genre knowledge as an integral sub-competence required for accurate translational transformation. The findings of this study and further research conducted on the corpus compiled for this research would help in terminology extraction and development of terminology and data banks that would facilitate programming for machine translation.

References

Ali, G., Iqbal, M. J., & Hayat, M.(2019). Transparency in Translation: A Corpus-Based Study of Contemporary English-Translated Fiction from Urdu and Non-Translated English Fiction. Pakistan Journal Of Translation Studies,7(2)

Anwar, A., Ali, G., & Shaheen, F. (2017). Translation and Colonial Discourse: Analyzing the Urdu translation of English novel “First Love and Last Love” by ‘James Grant’. *Pakistan Journal of Languages and Translation Studies*, 87.

Asghar, J., Butt, M. I., & Ali, G. (2020). Postcolonial Polysystems, Anglo-American Canonical Hegemony and the Marginalization of Urdu in Translation. *Kashmir Journal of Language Research*, *23*(1).

Bernardini, S., & Zanettin, F. (2004). Methodologies for The Investigation of Translation Universals. *Translation Universals: Do They Exist?*, *48*, 51.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.

Biel, Ł. (2017). Researching Legal Translation: A multi-perspective and mixed-method framework for legal translation. *Revista de Llengua i Dret, 68,* 76-88.

Blum-Kulka, S., & Levenston, E. A. (1983). Universals of lexical simplification. Dlm. C. Faerch & G. Kasper,(pnyt), Strategies in interlanguage communication,(hlm. 119-139).

Chan, R. (2007). One Nation, Two Translations: China's Censorship of Hillary Clinton's Memoir. In *Translating and interpreting conflict* (pp. 119-131). Brill.

Chesterman, A. (2004). Beyond the particular. *Translation universals: Do they exist*, *33*, 49.

De Sutter, G., & 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), 1-23.

Financial times news epaper link: https://www.ft.com/todaysnewspaper/edition/asia#

Frawley, W. (2000). Prolegomenon to a theory of translation. *The translation studies reader*, 250-63.

Hashmi, A. M. (2014). A Comparative Study of Eastern and Western Approaches in the Translation of Urdu Language Oral Narrative Daastaan-e Amir Hamza. *LITERARY QUEST*. https://lexically.net/wordsmith/version6/

Hu, K., Wu, Y., & Qing, T. (2007). Yuliaoku yu yixue yanjiu: qushi yu wenti—2007 yuliaoku yu yixue yanjiu guoji xueshu yantaohui zongshu [Corpora and translation studies: Trend and problems–A critical review of the international symposium of corpora and translation studies]. Journal of Foreign Languages, 5, 64-69.

Hu, X. (2006). Dangdai fanyi xiaoshuo de guifan yuliaoku yanjiu [A corpus-based study on the translational norms of contemporary Chinese translated fiction.(Unpublished PhD dissertation)]. *East China Normal University, Shanghai*.

Iqbal, M. J., Sibtain, M., & Shahzadi, R. (2020). Translational Style: A Corpus-Based Comparative Analysis of English Translations of Urdu Fiction by Memon and Naqvi. *sjesr*, *3*(3), 283-292.

Jantunen, J. (2001). Synonymity and lexical simplification in translations: A corpus-based approach. *Across Languages and Cultures*, *2*(1), 97-112.

Kachru, Y., & Bhatia, T. K. (2018). Hindi-Urdu. In *The world's major languages* (pp. 409-426). Routledge.

Khurshied, H., & Mahmood, M. A. (2020). A Corpus Based Study of Simplification in the Legal Translated Documents of Pakistan. *Al Qalam*, *25*(1), 334-344.

Khurshied,H.,Mahmood,M.A.,Mahmood,R.,(2022). Contrastive Analysis of Global Discourse Features In Pakistani Translated Legal And

Narrative Texts. *Harf-O-Sukhan*, *6*(1), 170-180.

Laviosa, S. (2002). *Corpus-based translation studies: theory, findings, applications* (Vol. 17). Rodopi.

Laviosa-Braithwaite, S. (1996). *The English Comparable Corpus (ECC): A resource and a methodology for the empirical study of translation* (Doctoral dissertation, University of Manchester).

Liu, K., & Afzaal, M. (2021). Syntactic complexity in translated and non-translated texts: A corpus-based study of simplification. *Plos one*, *16*(6), e0253454.

Mahmood, M. A., & Azher, M. A Corpus Based Study Of Explicitation In Pakistani Literary Text.

Mauranen, A., & Kujamäki, P. (Eds.). (2004). *Translation universals: Do they exist?* (Vol. 48). John Benjamins Publishing.

Nazir, F. (2019). The Task of the Translator: Cultural Translation or Cultural Transformation?

Qasim, H. M., Sibtain, M., & Sufi, M.(2020). Studying Translation Strategies in Subtitling of English Movies for The Pakistani Audience: An Interlingual Perspective. Pakistan Journal of Languages and Translation Studies Issue, No. 1(CeLTS)

Rogers, M. (2018). Specialised translation today: a view from the JoSTrans bridge. *JoSTrans: The Journal of Specialised Translation*, 3-22.

Roxas, R. E., Hussain, S., & Choi, K. S. (2011, November). Proceedings of the 9th workshop on asian language resources. In *Proceedings of the 9th Workshop on Asian Language Resources*.

Shahzad, A. K., Sibtain, M., & Qasim, H. M. (2020). Studying Strategies Employed to Minimize Cultural Differences in Translating Umera Ahmed's "Pir-e-Kamil" into English. *Al-Qalam*, *25*(1.), 345-361.

Shakur, N. (2008). Constructing Pathways to Translation: A Study in Translation Pedagogy and Process [from English into Urdu and Vice Versa] Part I.

Shlesinger, M. (1989). Extending the Theory of Translation to Interpretation: Norms as a Case in Point. *Target: International Journal of Translation Studies*, *1*(2), 111-15.

Sketchengine urdu parts of speech tagset link: https://www.sketchengine.eu/urdu-part-of-speech-tagset/

Sketchengine urtenten urdu corpus link: https://www.sketchengine.eu/urtenten-urdu-corpus/

Toury, G. (1991). *" Everything Has Its Price": An Alternative to Normative Conditioning in Translator Training*. ERIC Clearinghouse.

University of Lancaster, UK. "Claws Part-of-Speech Tagger for English." March 242012. http://ucrel.lancs.ac.uk/claws/.

Urdu POS tagger. (n.d.). Retrieved December 26, 2017, from http://www.cle.org.pk/software/langproc/POS_tagger.htm

Vanderauwera, R. (1985). *Dutch novels translated into English: The transformation of a minority literature*. Brill Rodopi.

Wen, T. H. (2009). *Simplification as a recurrent translation feature: a corpus-based study of modern Chinese translated mystery fiction in Taiwan* (Doctoral dissertation, University of Manchester).

Whitfield, A. (1992). Barbara Folkart. Le Conflit des énonciations. Traduction et discours rapporté. Montréal, Les Éditions Balzac, 1991, 481 p. *TTR: traduction, terminologie, rédaction*, *5*(2), 257-260.

Williams, D. A. (2005). *Recurrent features of translation in Canada: A corpus-based study* (Doctoral dissertation, University of Ottawa (Canada)).

Xiao, R., & Hu, X. (2015). *Corpus-based studies of translational Chinese in English-Chinese translation*. Springer Berlin Heidelberg.

Xiao, Z., & Dai, G. (2010). Xunqiu Disanyuma: Jiyu Hanyu Yiwenyuliaoku de Fanyi Gongxing Yanjiu (In pursuit of the "third code"). *Waiyu Jiaoxue yu Yanjiu (Foreign Language Teaching and Research)*, *42*(1), 52-58.

Yaqub, H., & Shakir, A. (2019). A Corpus-Based Study of Hypotactic and Paratactic Thematic Relations in English and Urdu Clause Complexes. *International Journal of English Linguistics*, *9*(5).