

Normalization in Specialized Translation; A Corpus Based Study of Genre Specific Urdu Language Texts

Dr. Humaira Khurshied¹, Dr.M. Asim Mahmood²(corresponding Author),

Dr. Samina Ali Asghar³, Dr. Rashid Mahmood⁴

Abstract

This exploratory study has investigated the relevance of the translation universals hypothesis of normalization to 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 study has investigated if genre specific specialized translated texts normalize in an Indo Aryan language Urdu. The analytical framework to investigate the translation specific linguistic traits is based on Zanettin(2013) WordSmith 7 has been used to find the distribution of list heads for the purpose of analysis . The analysis has revealed that the translated Urdu component (TUT) in USTC has shown a denormalizing tendency in two sub corpus. Moreover, comparable genre specific non translated language (CUT) and general Urdu language corpus components REF(GEN) also favor the same tendency in all the supcorpora. An adherence to generic conventions, the context of the production and consumption of specialized translated texts, and the content-oriented informative role of the texts have provided sufficient justification for contrary to the hypothesis findings. 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: Normalization, Genre, parallel and comparable Corpus, Specialized translation, Contextual factors.

1.Normalization

The concept of normalization in the context of translational practice stands for “conservatism” Baker (1996), “conventionalization” Mauranen (2007), “confirmation” of the typical patterns of the target language, rather “sanitization” (Kenny 1998) of the translated content to the extent that it may appear more “palatable” to the target language readers. Williams (2005) relates it to the tendency of translators to conform to the norms of written language in the target language and a restraint to use language innovatively and creatively. Hence, translated texts would be categorized as more norm governed and less creative in comparison to non translated ones. In other words, it can be generalized that translators produce texts that ought to be “acceptable” in terms of linguistic choices and text patterns in the target culture.

Normalization tendency endows translated material with the overuse of typical grammatical patterns and generic features, lexical bundles, and an over adaptation of punctuation patterns familiar to the written language conventions in the target language culture. It results in the domestication of culturally loaded terms, formalizing the colloquial style, simplifying the syntactic structures, and enhancing the coherence of the target text by rearranging the sentences,

1 Dr. Humaira Khurshied, Government Graduate College Sahiwal Pakistan. Email: humaira.khurshied@yahoo.com

2 Prof. Dr. Muhammad. Asim Mahmood, Government College University Faisalabad Pakistan Email: masimrai@gmail.com

3 Dr. Samina Asghar University of Education, Faisalabad Campus

4 Prof. Dr. Rashid Mahmood, Riphah University, Faisalabad Campus

paragraphs, and chapters. Even some of the features considered representative of simplification universal have also been associated with normalization i.e lower degree of lexical and syntactic diversity, a higher proportion of list heads and a lower proportion of low-frequency words (Laviosa 2002), fewer contracted forms (Olohan 2003) and less diversified collocations (Kenny 2001).

1.1. Empirical Studies on Normalization

Since Vanderauwera's (1985) earlier contentions about normalization, a series of empirical studies have been conducted to verify the hypothesis. Shlesinger's (1991) study investigating translated English texts of Hebrew origin found a normalizing tendency in translated texts by adding complete utterances and sentences and preferring grammatical accuracy that was left unattended in the source texts. May's (1997) study dealing with the comparative analysis of translated literary texts in Russian and French languages from English origin found that the original texts expressed characters' mental phases in short, long, complex, and compound sentences, in the translated versions the compound sentences were replaced by complex sentences and complete sentences had been used instead of incomplete ones. Munday's (1998) and Overas (1998) investigated literary translations and their investigation supported the normalization hypothesis. Malmkjaer (1998), Hansen, and Teich (2001) too supported evidence of normalization. Mauranen (2000) had mixed observations in her analysis of translated academic prose and popular nonfiction. She found unusual lexical patterns as well as a higher repetition rate of the words that formed those patterns.

Kenny's studies (1998, 1999, 2001, 2006) illustrated the point of how untypical collocation patterns and compounds in the German language had been rendered as normalized ones in the English language during the process of translation. Her findings suggest that the degree of normalization varies from translator to translator and most of the times translators normalize the lexical patterns of the text. Translators prefer recurrent lexical bundles as compared to the varied lexical bundles that are specific to the original authors. Empirical studies not only favored the normalization hypothesis but also challenged it. Blum-kulka and Levenston's (1983) analysis of Hebrew and English translations revealed that translators adopted the strategy of word coinage to fill "semantic voids" instead of relying on existing vocabulary. Kenny's (2001) study also unfolded that some translators may opt for creative solutions. The study conducted by Tirkknen-Condit (2002) on translated Finnish established the deviation from the normalization hypothesis.

Diachronic changes in the normalization trends in translated Chinese literary texts of English origin were investigated by Xia (2010). The findings suggest that the normalization tendency represented by using conventional language features is relative in the sense that it varies with the passage of time depending upon the prevalent translational norms and different socio-cultural constraints. Moreover, the normalization tendency is most of the time reflected in the surface structure of the translated texts hence a corpus based analysis can be opted to grasp this tendency. The translated texts that form the sub corpora for this study belong to specialized translated category.

1.2. Specialized Translation

There has been a longstanding convention to focus on the translation of literary works (including Philosophical, rhetorical, and religious works) as the sole subject matter of translation studies. According to Holmes (1972), the efforts to develop theories for the translation of scientific texts were relatively new. As far as the defining boundaries of Specialized translation are concerned Baker and Saldanha's (2009) domain specific categorization includes several entries: commercial translation; dealing with business context, institutional translation; 'translating in and for specific organizations', scientific and technical translation; dealing with the domain of science and technology. To make the matter short Postolea (2016) defines specialized translation as "translation carried out in specialized context i.e the context which involves specialist source text producers, specialist topics, or a specialist activity". The product of specialized translation approximates 80% of the translation market and the remaining 20% goes for both literary and Biblical content (Wilss 1999). On the other hand, the proportion of scholarly endeavors in terms of research in the domain of specialized translation is in sharp contrast to the huge bulk of its product. The volume of research publications in the domain had been almost 1.4% till the 1950s and thereafter it increased to 10.2% till the 1990s Aixela (2004). Even the statistics of recent years still points toward this imbalance in research and theoretical work Olohan (2013), Rogers (2015). Rogers (2018) further probes the issue and states,

"JoSTrans is still, to my knowledge, the only international refereed journal dedicated to specialized translation".

1.3. An overview of Urdu language

The status of English as the dominant lingua franca and storehouse of knowledge is well established. Urdu is the national language and lingua franca of Pakistan and is spoken by approximately by 170 million people (including those who speak it as their second language). It is the 11th most widely used language (Ethnologue 2018) around the globe and has an established literary inheritance of three centuries. It is an Indo Arian language having origin in Sanskrit and heavily influenced by Persian, and Arabic from where it has borrowed most of its literary and technical vocabulary. This morphologically rich and highly inflectional language is written in Nastalique style from right to left and lacks case discrimination. Moreover, the script is context bound i.e form of an alphabet is determined by the alphabets surrounding it. The isogloss representing Urdu is not only confined to the geographical boundaries of Pakistan, it is also spoken in parts of India, Afghanistan, and Bangladesh. Besides being the lingua franca of a wider business community within Pakistan, across South Asia and U.K, it is one of the widely spoken languages in the U.K.

1.4. Status of Specialized Translation in Pakistan and existent gap

The international scenario presented above is also indicative of the status of research work being conducted in the domain of specialized translation in a developing country like Pakistan. The problems like postcolonial inheritance and resultant prestigious status of English, lack of immense translational activity, the education policy of promoting the English language as custodian of knowledge instead of transforming the knowledge in the native or national

languages, and nonavailability of institutional, financial, and technical assistance to undertake the specialized translation projects and research work has further aggravated the situation.

The earlier endeavors in corpus based translation studies dealing with the features of the translated language were restricted to translation from and into European languages. Latter, the subject matter of inquiry shifted to other languages like Chinese, Arabic, Persian and Russian. The present research is a further extension of this convention to investigate English Urdu language pair, genre specific parallel comparable corpus from the perspective of translational universals.

1.5. Research Questions

This study is designed to find the translation universal of normalization in the Urdu language specialized translated texts. The analysis in this research was motivated by the following research questions:

1. Do the translated texts in USTC constituent genre specific specialized subcorpora provide evidence to support:
 - a. Normalization as universals of translated language?
 - b. Have genre specific linguistic traits or general language specific features informed linguistic choices in the case of translated texts in USTC constituent subcorpora i.e NSC, ASC, and LSC?

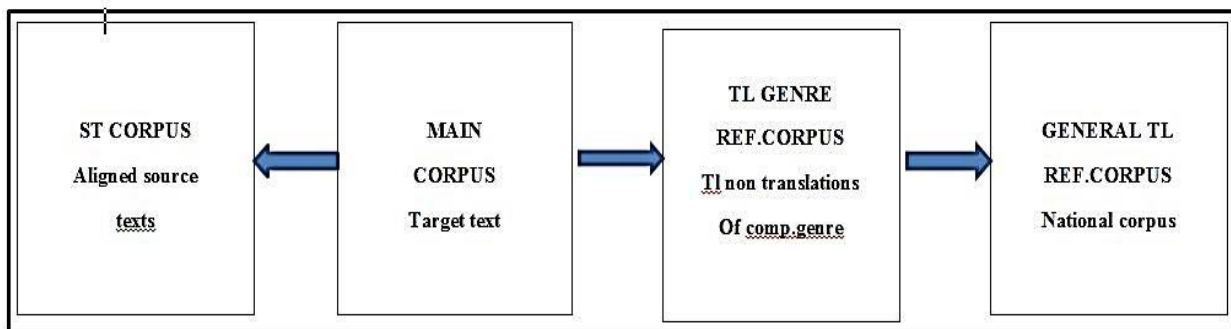
2. Methodology

The corpus designed for the purpose of this study comprises three sub subcorpora i.e legal statutes, newspaper articles, and academic texts. The analytical tool to measure the normalization hypothesis has been adapted from Zanettin 2013. Wordlist tool in WordSmith 7 has been used to find the distribution of list heads.

2.1. Corpus design

The corpus compiled for the purpose of analysis in the present study can be termed as a composite bilingual parallel one, following Laviosa (2006) and Bernardini (2011). The corpus contains a Parallel bilingual corpus consisting of source English text in all the genres and their translated Urdu texts as well as comparable Urdu texts in all the genres and a general reference corpus. The following corpus structure has been adopted for corpus construction.

Figure 01: Corpus architecture design based on Biel (2017)



2. Analysis

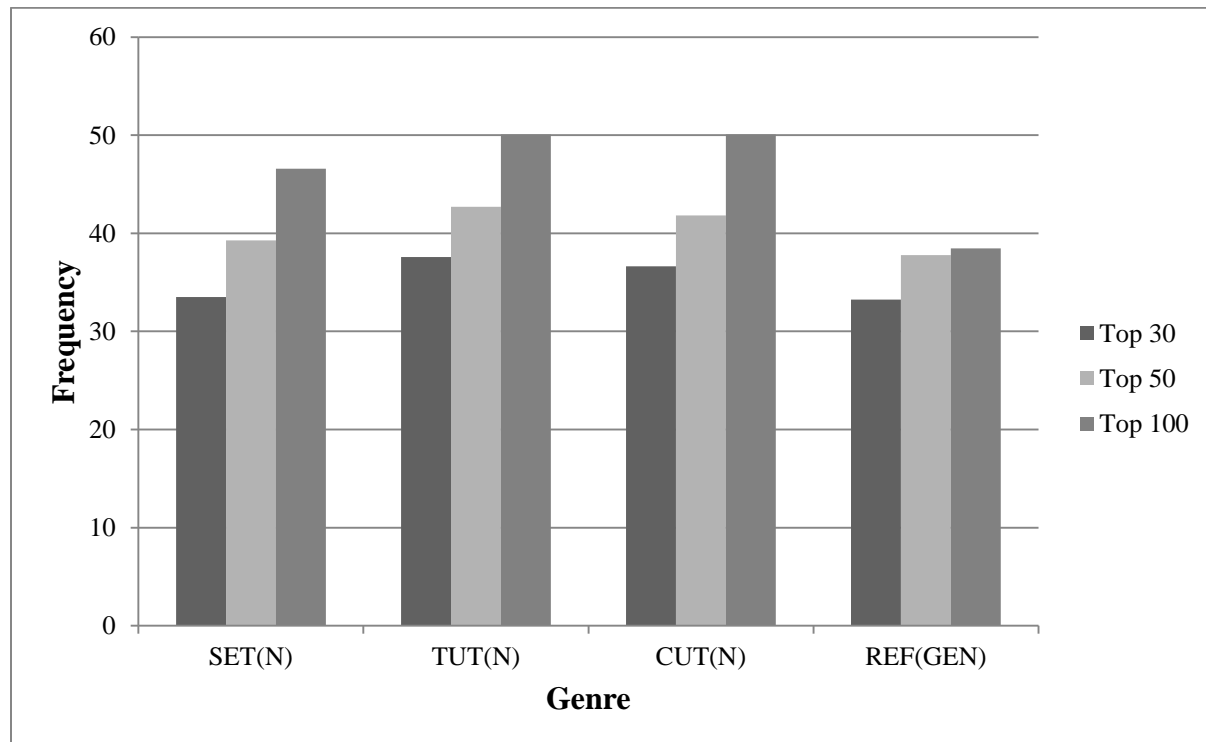
The analysis based on the distribution of list heads in all the subcorpora revealed the following tendencies;

3.1. Distribution of List Heads in NSC

Table 01: Distribution of list heads in Sub Corpus NSC

| NSC | SET(N) %age | | | TUT(N) %age | | | CUT(N) %age | | | REF(GEN) %age | | |
|-----|-------------|--------|---------|-------------|--------|---------|-------------|--------|---------|---------------|--------|--------------|
| | TOP 30 | TOP 50 | TOP 100 | TOP 30 | TOP 50 | TOP 100 | TOP 30 | TOP 50 | TOP 100 | TOP 30 | TOP 50 | TOP 100 |
| | 33.51 | 39.27 | 46.59 | 37.57 | 42.69 | 50.07 | 36.63 | 41.81 | 50.06 | 33.25 | 37.79 | 38.47 |

Figure 02: Distribution of list heads in NSC



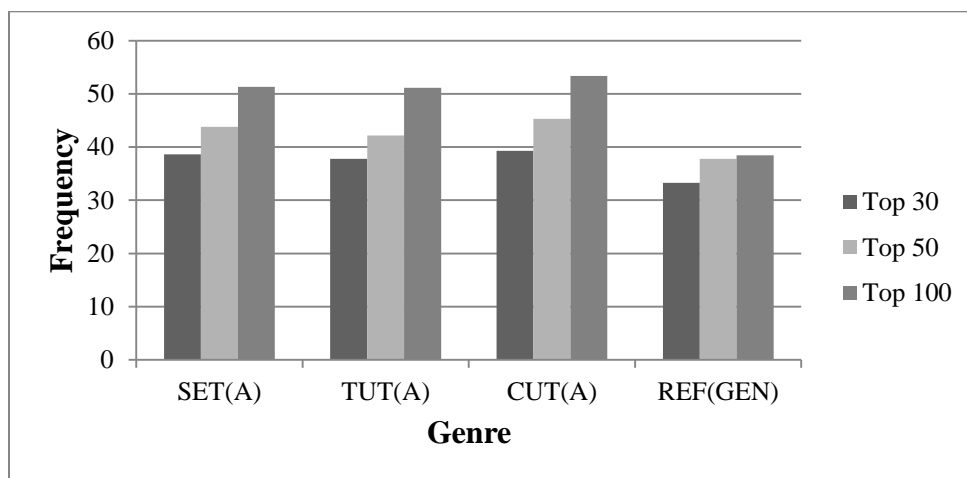
The data elicited from the Sub corpus NSC indicates that the top 30 words in the SET constituted 33% of the total source text size, the top 50 words constituted 39 % of the source text and the top 100 words constituted almost 47 % of the SET(N) corpus component. A comparative view of the TUT(N) indicates that the relative percentage of all three list heads is higher as 38% for top30 words, 42 % for top 50 words, and 50% for top 100 words. This tendency of the TUT(N) component has provided evidence in support of the research hypothesis. When the frequencies are compared with the third subcorpus component i.e comparable one we find that the percentage is quite closer to TUT(N). Since in CUT(N) top30 words form 37 % of the total data, the top 50 words form 41 % of the data and the top 100 words. As far as the REF(GEN) corpus is concerned, the percentage of list heads is quite low from the other two Urdu corpus components i.e 33%, 37%, and 38 % for the top 30, top 50, and, top 100 words respectively. Here the frequencies seem to be slightly tilted to SET(N).

3.2. Distribution of List Heads in ASC

Table 02: Distribution of list heads in ASC

| | SET(A) %age | | | TUT(A) %age | | | CUT(A) %age | | | REF(GEN) %age | | |
|-----|-------------|--------|---------|-------------|--------|---------|-------------|--------|---------|---------------|--------|---------|
| | TOP 30 | TOP 50 | TOP 100 | TOP 30 | TOP 50 | TOP 100 | TOP 30 | TOP 50 | TOP 100 | TOP 30 | TOP 50 | TOP 100 |
| ASC | 38.64 | 43.78 | 51.29 | 37.79 | 42.18 | 51.14 | 39.31 | 45.32 | 53.37 | 33.25 | 37.79 | 38.47 |

Figure 03: Distribution of list heads in Corpus ASC



The percentage of list heads in the second subcorpus ASC is indicating contrary to the research hypothesis trend since the TUT (A) list heads percentages are lower than SET(L) i.e 37%, 42%, and 51% respectively for the top 30, top50, and top100 word categories. However, the trend in

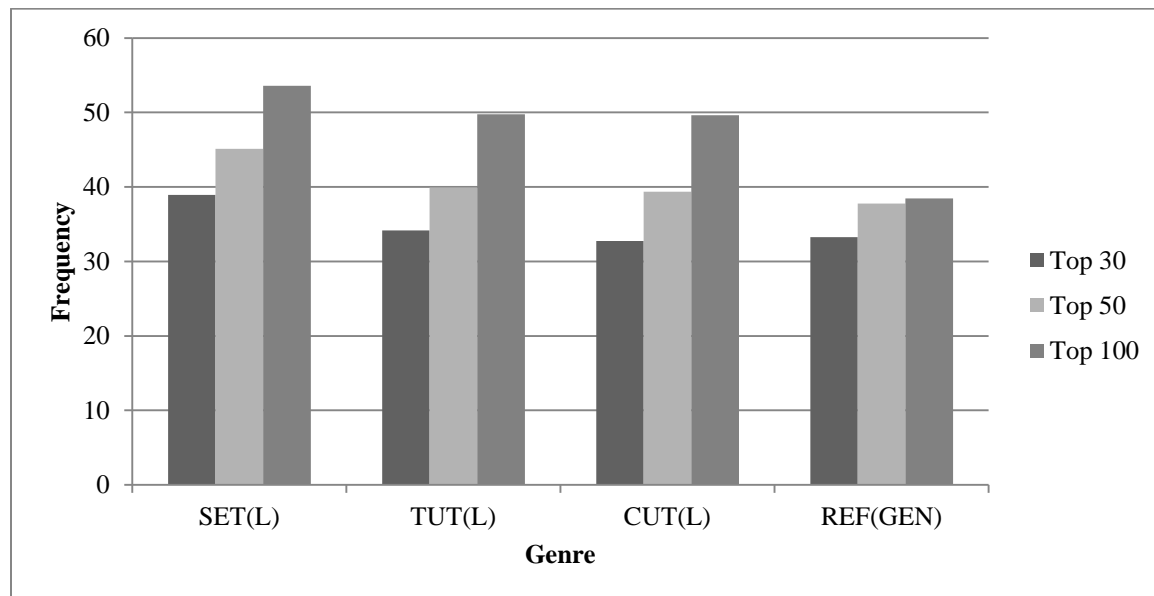
the lower percentage of list heads has not been followed by CUT(A) since this component contains 39%, 45%, and 53% in top30, top50, and top100 word categories respectively. However, the REF(GEN) corpus component seems to be in a unique position since here the percentage for all the three list heads is the lowest and resultantly this text seems to be the least normalized or highly creative among the ASC subcorpus components in terms of list head frequencies.

3.3. Distribution of List Heads in LSC.

Table03: Distribution of list heads in LSC.

| | SET(L) %age | | | TUT(L) %age | | | CUT(L) %age | | | REF(GEN) %age | | |
|-----|-------------|--------|---------|-------------|--------|---------|-------------|--------|---------|---------------|--------|---------|
| | TOP 30 | TOP 50 | TOP 100 | TOP 30 | TOP 50 | TOP 100 | TOP 30 | TOP 50 | TOP 100 | TOP 30 | TOP 50 | TOP 100 |
| LSC | 38.92 | 45.14 | 53.60 | 34.16 | 39.93 | 49.78 | 32.75 | 39.35 | 49.62 | 33.25 | 37.79 | 38.47 |

Figure 04: Distribution of list heads in LSC



The analysis of the list heads in the LSC subcorpus indicates that the frequency of all the three categories of list heads is lower 34.16%, 39.93% and 49.78% for top30, top50, and top100 words

respectively in the TUT(L) component as compared to the SET(L) component. This provides contrary to research hypothesis evidence and reveals that translated component has used varied vocabulary in terms of list heads analyzed in the text. This trend of variance in the use of vocabulary in terms of list heads has also been supported by the other two Urdu components i.e CUT(L) and REF(GEN). Because the frequencies of list heads in these two corpus components are too lower not only than the SET(L) but also lower than TUT(L) subcorpus component.

4. Conclusion

The Urdu language translated components for the two specialized genres have not provided evidence in support of normalization hypothesis. Only NSC translated component shows less lexical diversity in terms of the percentage of list heads; and provides evidence in support of research hypothesis. However, the translated components in ASC and LSC subcorpora have shown higher lexical diversity. The trend of higher linguistic creativity is also visible in the CUT and REF(GEN) components. This trend is a manifestation of strong adherence to Urdu language norms by the translated texts under study. It can be inferred from the findings ;

The generic norms have influenced linguistic patterns; The contrary to hypothesis findings in the translated Urdu language can be attributed to the very nature of texts that constituted the corpus. The excessive use of nominalized verbs is a typical syntactic feature of legal texts (Bhatia, 1994; Gibová, 2009, Pavlíčková, 2012). The translated draft of the legal text understudy also followed this generic convention and it has a higher frequency of nominalized verbs. For Halliday (1994) nominalization is associated with impersonal or abstract voicing of ideas. For the texts investigated in this study i.e legal statutes, newspaper articles, and academic prose, it is natural to use nominalizations in order to give an air of formality, aloofness, and objectivity. This very feature has been rendered characteristic of academic prose and newspaper articles by Biber and Conrad (2019). This text type specific feature accounts for higher lexical creativity in terms of lexical and collocational diversity and creativity in the translated components of the three genres in this study.

The findings for the normalization hypothesis in this study has established that the translated Urdu language corpus has rather denormalized. Instead of retaining the status of ‘third code’ clear traces of target language system adherence are visible. Since, the translated components in genre specific subcorpora have closer affinity the comparable and reference Urdu language components in lexical creativity. This trend of the translated components, to have closer numerical affinity with the Urdu language non translated comparative components is visible in all the three sub corpora.

4.1. Contribution of the Study

- This study has provided a re appraisal of translation universals hypothesis from the perspective of Urdu language translated genre specific texts.
- This study has incorporated three corpus components simultaneously i.e. parallel, comparable, and reference corpus in the corpora for analysis.
- This research has compiled genre specific specialized corpus (data consisting of three functional or specialized genres i.e. statutes, news articles, academic prose) while earlier

endeavors had been restricted to the compilation of either translated literary texts or other variants of nonliterary texts.

- This study focuses on the improvement of translational transformation of specialized texts and it would provide food for thought to bilingual lexicographers, translators, lawyers, comparative law specialists, academicians, newspaper article writers, and national and international stakeholders involved in the act of TSP communication.

4.2. Limitations of the Study

The findings of the study are language pair, text type and genre specific. Keeping in view the limitations associated with the compilation of parallel corpus and practical hindrances of converting Urdu word and Pdf documents into text files, resultant time-intensive proofreading issues due to the nonavailability of any standard Urdu language optical recognizer, the corpus construction convention of extracting data from multiple sources has been set aside. And data set comprises full texts rather than extracted samples from different sources. The findings are restricted to the linguistic operator of distribution of list heads as specified in the analytical framework Zanittin (2013) for the measurement of normalization universal of translation.

4.3. Insight for Further Research

Since earlier studies in the domain of translation studies and genre analysis dealt with the European languages this research based on the analysis of Indo Aryan language i.e Urdu, not only would pave way for further research within the Indo Aryan body of languages, it would also broaden the horizon of translation studies to frame extended theoretical contentions to encompass all languages.

The analytical framework of this study can be applied to other genres within the registers of journalistic texts, academic prose, legal texts, and other registers to add information to the existing sum of knowledge in the domain of specialized translation in the Urdu language.

This very corpus can be used for the contrastive analysis of collocation patterns and phraseology analysis and the findings would inform about the recurrent word combinations and would facilitate specialized translators.

References

- Baker, M. (1996). Corpus-based translation studies: The challenges that lie ahead. *Benjamins Translation Library*, 18, 175-186.
- Baker, M., & Saldanha, G. (2009). *Routledge Encyclopedia of Translation Studies*. 2nd edn. London & New York.
- Bernardini, S. (2007, July). Collocations in translated language: Combining parallel, comparable and reference corpora. In *Proceedings of the Corpus Linguistics Conference* (pp. 1-16).
- Bhatia, V. K. (1994). Generic integrity in professional discourse. *Text and talk in professional contexts*, (6), 61-76.
- Biber, D., & Conrad, S. (2019). *Register, genre, and style*. Cambridge University Press.
- Biel, L. (2017). Researching Legal Translation: A multi-perspective and mixed-method framework for legal translation. *Revista de Llengua i Dret*, 68, 76-88.
- Franco Aixelá, J. (2004). The study of technical and scientific translation: an examination of its historical development. *The journal of specialised translation*, 1, 29-49.



- Gibová, K. (2009). EU translation as the language of a reunited Europe reconsidered. *Language, Literature and Culture in a Changing Transatlantic World*, 192202.
- Halliday, M. (1994). *Introduction to Functional Grammar* 2nd Ed (London: Edward Arnold).
- Hansen, S., & Teich, E. (2001, July). Multi-layer analysis of translation corpora: methodological issues and practical implications. In *Proceedings of EUROLAN 2001 workshop on multi-layer corpus-based analysis* (Vol. 45). Iasi, Romania.
- Holmes, J. S. (1975). *The name and nature of translation studies* (pp. 172-185). Amsterdam: Translation Studies Section, Department of General Literary Studies, University of Amsterdam.
- Kenny, D. (1998). Creatures of habit? What translators usually do with words. *Meta: journal des traducteurs/Meta: Translators' Journal*, 43(4), 515-523.
- Kenny, D. (1999). The German-English parallel corpus of literary texts (GEPCOLT): A resource for translation scholars. *Teanga*, 1, 25-42.
- Kenny, D. (2001). *Lexis and creativity in translation. A corpus-based study*. Manchester: St. Jerome Publishing.
- Malmkjær, K. (1998). Cooperation and literary translation. *The pragmatics of translation*, 25-40.
- Mauranen, A. (2000). Strange strings in translated language: A study on corpora. *Intercultural faultlines. Research models in translation studies I. Textual and cognitive aspects*, 119-41.
- May, R. (1997). Sensible elocution: How translation works in & upon punctuation. *The translator*, 3(1), 1-20.
- Munday, J. (1998). A computer-assisted approach to the analysis of translation shifts. *Meta: journal des traducteurs/Meta: Translators' Journal*, 43(4), 542-556.
- Olohan, M. (2003). How frequent are the contractions?: A study of contracted forms in the Translational English Corpus. *Target. International Journal of Translation Studies*, 15(1), 59-89
- Olohan, M. (2013). Scientific and technical translation. In *The Routledge Handbook of Translation Studies* (pp. 443-455). Routledge.
- Pavličková, E. (2012). The role of nominalisation in English legal texts. *English Matters*, (3), 4-10.
- Postolea, S. (2016). Translating in a Specialised Context: Challenges and Risks. *Bulletin of the Polytechnic Institute of Iasi. Section: Social Sciences*, 51-66.
- Rogers, M. (2015). Terminology and Specialised Translation: A Historical Perspective. In *Specialised Translation* (pp. 81-109). Palgrave Macmillan, London
- Rogers, M. (2018). From binaries to borders: Literary and non-literary translation. In *Moving Boundaries in Translation Studies* (pp. 151-167). Routledge.
- Schäffner, C., & Adab, B. (2001). The idea of the hybrid text in translation: Contact as conflict. *Across languages and cultures*, 2(2), 167-180.
- Shlesinger, M. (1991). Interpreter latitude vs. due process. Simultaneous and consecutive interpretation in multilingual trials. *Empirical research in translation and intercultural studies*, 147-155.
- Tirkkonen-Condit, S. (2002). Translationese—a myth or an empirical fact?: A study into the linguistic identifiability of translated language. *Target. International Journal of Translation Studies*, 14(2), 207-220.
- Vanderauwera, R. (1985). *Dutch novels translated into English: The transformation of a minority literature*. Brill Rodopi.
- Williams, D. A. (2005). *Recurrent features of translation in Canada: A corpus-based study* (Doctoral dissertation, University of Ottawa (Canada)).
- Zanettin, F. (2013). Corpus methods for descriptive translation studies. *Procedia-Social and Behavioral Sciences*, 95, 20-32.