



## THEORETICAL CORPUS-BASED TRANSLATION STUDIES

Bozorova Vasila Ilhom qizi

Teacher at Uzbekistan state World Languages University

**Abstract:** *Theoretical corpus-based translation studies (CBTS) are a methodological approach that uses corpora as primary data sources to study translation phenomena. It focuses on corpus linguistics and translation theory, analyzing large collections of text to understand language patterns and usage. The research objectives include empirical investigation and generalizability across different languages and contexts. Key concepts include translation universals, quality assessment, and translator training. CBTS has applications in translator training programs, machine translation systems, and interdisciplinary perspectives, such as cognitive linguistics and sociolinguistics and pragmatics. It also addresses ethical concerns related to the use of corpora and may integrate with technology to enhance translation tools and methodologies. By bridging theory with practical application, CBTS continues to advance translation studies and inform scholarly research and professional practice.*

**Key words:** *methodological approach, corpus, corpus-based translation, language patterns, machine translation, pragmatics, sociolinguistics*

### INTRODUCTION

Baker remained at the vanguard of intense theoretical discussions. She authored three articles between 1996 and 1999 that significantly contributed to the theoretical development of the connections between CTS and target-oriented methodologies. She offered suggestions on how to improve theories and procedures in order to make theoretical ideas like the idea of universals operational and verifiable. Furthermore, according to Baker, determining the patterns that distinguish translation as translation should be done in tandem with determining the relative standing of source and target languages—that is, prominent versus less prestigious languages. Furthermore, descriptive academics must concentrate on the interaction of three factors: readership expectations, theoretical declarations, and professional practice, since translated texts are unique communication events that are formed by their own aims, pressures, and production contexts. In order to better understand them, extralinguistic data sources like historical records, book reviews, author and translator interviews, publishing businesses' output, funding organizations' decisions, and interviews can be combined with textual studies.

What is corpus linguistics?

One significant distinction between corpus linguistics and DTS/CTS is the incorporation of information to be collected outside of the corpus. In the latter case, particularly when taking Sinclair's stance into account, the corpus is the sole acceptable subject of

"It's all in the language," asserts Sinclair (Sinclair, personal communication 1994). Alternatively, he may invoke Hatim's (1999) ideas and state that the text itself will show what is in and out of the text. Nevertheless, in order to supplement and clarify the results of



linguistic descriptions, descriptive translation studies must also look elsewhere. Examples of this include the significance of psycholinguistic research models like Think Aloud Protocols or the examination of the interim tactics used by the translator as made clear by manuscripts, typescripts, or corrected proofs. DTS and CTS both consider various approaches, provided that they are rigorous and originated from inside the discipline (Toury, 1995).

However, from a theoretical perspective, other novel syntheses were being proposed: Halverson (1998) adopted prototypical categories for characterizing the subject of study in corpus-based investigations and breaking the impasse caused by the contradicting statements that professional translations enjoy a higher status, primarily because of evidence from psycholinguistic studies, and that any translation that is "presented or regarded as such within the target culture, on whatever grounds" (Toury 1985, 20) constitutes legitimate data for empirical and theoretical research. She proposed that the target parent population of translated works be seen as a prototype category, the center of which is occupied by professional translations, but only for the cultures of industrialized Western countries, while there are professional translations occupying the periphery for collections of various sorts of translation, such as those done in the target language or by trainee translators.

This means that a corpus that aims to reflect the population of translated writings will comprise a variety of subcorpora that each offer varying degrees of relevance and are all considered valid subjects of study. Prototypes and the corporations created to represent a particular parent demographic are both culturally specific. This brings up the difficult question of whether the study's purpose is comparable and whether particular research findings should be generalized as a result.

It is a persistent issue in CTS that has repeatedly surfaced during the construction of monolingual and multilingual corpora. Using the same criteria used to create the English Comparable Corpus (ECC), it would be difficult to create a representative and balanced monolingual comparable corpus of narrative texts in a country like Brazil, where translated literature makes up about 90% of all published literature (Magalhaes, 2001). This challenge in less widely spoken languages extends beyond literary genres to everyday language use as well. Kenny (1998) notes that a large number of nonliterary text types in Irish Gaelic, for example, are translations, primarily from English. The Corpus of Literature's design has been impacted by the impact of translation policies. Finnish translations collected at the Savonlinna School of Translation Studies, with Anna Mauranen serving as the director.

Since this text category is not translated into Finnish, academic texts in the natural sciences have been removed. Bi-directional parallel corpora are also of interest in this situation. The selection of texts for the English-Norwegian Parallel Corpus was restricted due to the fact that a large number of text categories have been translated into Norwegian but a comparatively small number into English (Johansson & Hofland 1994; Johansson 1998).

Applied corpus-based translation studies



The University of Bologna in Forlì's CEXI, a bi-directional parallel English-Italian corpus assembled under Guy Aston's supervision, has likewise been impacted by this kind of imbalance. The issue, which is thoroughly described by Zanettin (2002), stems from the significantly different demographics of the parent populations of translated narrative and non-fiction works in English and Italy. This has an impact on the representativeness of the original subcorpora and the degree of comparability that can be attained for the translational and original components of the corpus as a whole. Thus, the kinds of comparative analysis that can be performed are restricted. To put it another way, there is a trade-off between representativeness and balance/comparability. The challenge will be overcome.

by striking a compromise: representativeness will be attained by encircling the core corpus with unidirectional parallel subcorpora that more closely resemble the makeup of the translation parent populations, while balance will be maintained inside the core corpus. Therefore, with some maneuvering between what is "given" and what is "taken" as the object of study in a specific socio-cultural milieu, internal balance, representativeness, and comparability can be achieved in corpus design to a tolerable level. While cross-cultural comparability is undoubtedly a much more complex issue, it is possible to reach agreement on the design principles and make them clear so that various research communities can communicate with one another about the meaning of each other's findings and the justification for specific decisions.

The core of theoretical considerations in CTS has always been corpus design. Apart from Halverson (1998), the approach was critically remarked upon and proposed by Shlesinger and Malmkjaer. A new kind of monolingual comparable corpus, comprising interpreted speeches from a range of source languages, original spoken texts produced in comparable settings, and written translations of oral source texts produced in comparable settings, was proposed by Shlesinger (1998) as a way to reveal the unique characteristics of interpreting in comparison to written translation and original spoken discourse. In order to provide information helpful to academics interested in the study of equivalencies and those who concentrate on, Malmkjaer (1998) suggested a new kind of parallel corpus, one that includes as many distinct translations of the same source text as feasible on the act of translating itself. This is due to the possibility that the conventional unidirectional parallel corpus may obscure a crucial facet of the translation process, namely the distinct decisions and tactics used by many translators.

#### REFERENCES:

1. Blum-Kulka, S. and E. A. Levenston. 1983. Universals of lexical simplification. In *Strategies in interlanguage communication*, ed. C. Faerch and G. Kasper, 119-139, London: Longman.
2. Bowker, L. 1998. Using specialized monolingual native-language corpora as a translation resource: A pilot study. *Meta* 43 (4):631-651. (Special issue: The corpus-based approach, ed. S. Laviosa.)



3. Bowker, L. 2000. A corpus-based approach to evaluating student translations. *The Translator* 6 (2):183-210. (Special issue: Evaluation and Translation, ed. C. Maier.)
4. Brisset, A. 1990. *Sociocritique de la Traduction: Théâtre et Altérité au Québec, 1968-1988*. Longueuil, Canada: Le Préambule.
5. R. Gill and R. Gannon. Toronto: University of Toronto Press. Burnett, S. 1999. A corpus-based study of translational English. Unpublished MSc dissertation. Centre for Translation and Intercultural Studies, UMIST, Manchester.
6. Camargo, D. C. de 2001. Corpus-based translation research on legal, technical and corporate texts. *Across Languages and Cultures* 2 (1):113-125.
7. Carey, F. H. 1844. *The vision of hell, purgatory and paradise of Dante Alighieri*. London: George Bell & Sons.
8. Chafe, W. L., J. W. DuBois and S. A. Thompson. 1991. Towards a new corpus of Spoken American English. In *English corpus linguistics in honor of Jan Svartvik*, ed. K. Aijmer and B. Altenberg, 64-82, London: Longman.
9. Chesterman, A. 1998. Causes, translations, effects. *Target* 12 (1):200-230. Chesterman, A. 2000. A causal model for translation studies. In *Intercultural faultlines. Research models in translation studies: Textual and cognitive aspects*, ed. M. Olohan, 15-27, Manchester: St. Jerome.
10. Chesterman, A. and R. Arrojo. 2000. Shared grounds in translation studies. *Target* 12 (1):151-160. Frawley, W. 1984. Prolegomenon to a theory of translation. In *Translation: literary, linguistic and philosophical perspectives*, ed. W. Frawley. London: Associated University Presses.