

CORPUS-BASED MODELLING OF TERMINOLOGICAL HARMONIZATION IN
MULTILINGUAL UNITED NATIONS INSTITUTIONAL DISCOURSE

Gafarova Zumrad Zokhirjonovna

Head of Foreign Language and Social Sciences Department, Asia International University

Associate Professor (PhD)

gafarovazumrad@oxu.uz

Orcid: 0009-0003-6034-6150

Annotation. In this article you will know the increasing complexity of multilingual institutional communication has intensified the need for systematic approaches to terminological harmonization, particularly within international organizations such as the United Nations (UN). The production, translation and dissemination of institutional documents across multiple languages require not only lexical equivalence but also conceptual accuracy, terminological consistency and pragmatic correspondence. This study proposes a corpus-based model for investigating and systematizing terminological harmonization in multilingual United Nations institutional discourse. The research integrates the methodological principles of corpus linguistics, terminology studies, institutional translation and contrastive linguistics. Particular attention is paid to terminological equivalence, variation, conceptual correspondence, frequency, collocational behaviour and contextual stability across parallel documents.

Keywords: corpus linguistics, parallel corpus, terminological harmonization, institutional discourse, United Nations documents, terminology management, translation equivalence, multilingualism, terminological variation, specialized corpus.

Introduction. Global institutional communication is characterized by the simultaneous operation of several languages, legal traditions, administrative systems and conceptual frameworks. Within international organizations, terminology functions not merely as a collection of specialized lexical units but as an instrument for transmitting institutionally recognized concepts with maximum semantic precision. Consequently, the translation of institutional terminology requires a higher degree of consistency than ordinary lexical translation.

The United Nations represents one of the most complex environments of multilingual institutional communication. The Organization operates with six official languages – Arabic, Chinese, English, French, Russian and Spanish and maintains the UNTERM terminology portal to support consistent use of official terminology across the UN system. UN multilingualism guidelines explicitly identify UNTERM as a standard reference resource for United Nations terminology. The institutional status of terminology makes terminological consistency particularly significant. A terminological unit used in resolutions, conventions, reports, agreements or administrative instruments is frequently linked to a specific institutional concept. Its inaccurate or inconsistent translation may therefore produce not only linguistic divergence but also conceptual ambiguity. Corpus linguistics provides effective methodological instruments for addressing this problem. Parallel corpora make it possible to compare corresponding linguistic units in source and target texts systematically, identify recurrent translation equivalents and investigate variation in real communicative contexts. Since Baker's seminal discussion of the relationship between corpus linguistics and translation studies, corpus methods have developed into an important empirical approach to studying translation patterns. [1:235]

Research Methodology. The present research adopts an integrated methodology combining:

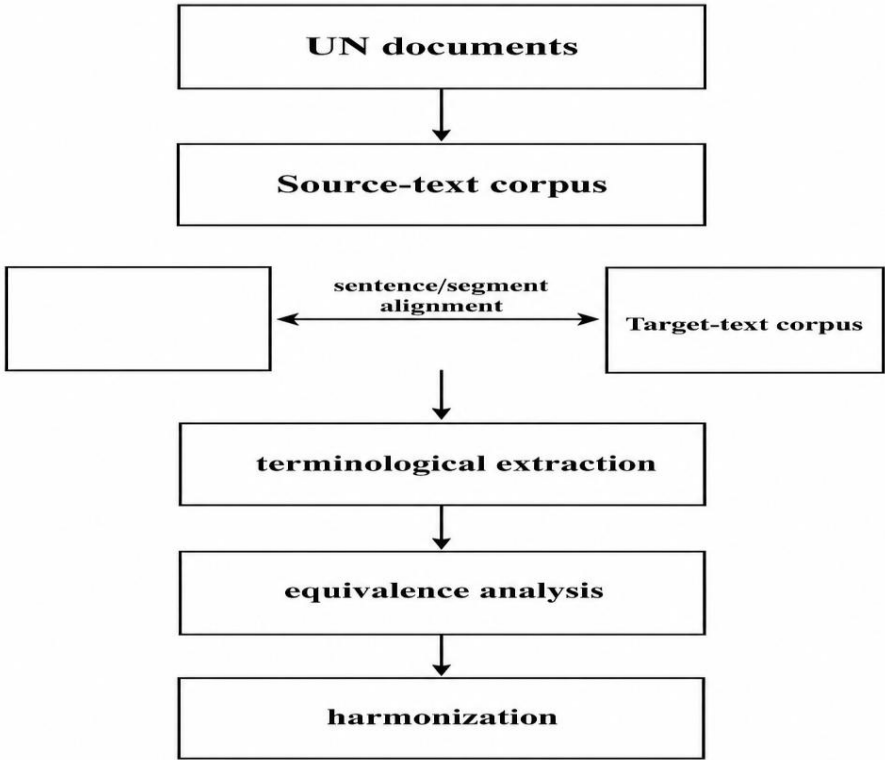
- corpus linguistics;
- contrastive linguistics;
- translation studies;

- terminology studies;
- semantic analysis;
- contextual analysis;
- frequency analysis;
- collocational analysis. [2: 256]

The methodological object is a specialized multilingual parallel corpus of United Nations institutional documents. The corpus may include the following genres: *UN resolutions* → *conventions* → *declarations* → *reports* → *agreements* → *policy documents* → *institutional guidelines*

For English-Uzbek research, English may function as the principal reference language, while Uzbek translations constitute the target-language subcorpus. Where necessary, Russian may be used as an auxiliary comparative language because of its historical influence on Uzbek administrative and legal terminology. [3:90]

A representative specialized corpus should contain metadata for each document: document title; UN document symbol; publication date; institutional body; thematic field; source language; target language; document genre; translation status; number of tokens; alignment information.



Picture 1. Terminological procedure of corpus

Document-level and sentence-level alignment are particularly important. Without alignment, researchers may identify term frequency but cannot reliably determine translation correspondence.

Terminological candidates may be extracted through a combination of automatic and manual procedures.

Automatic identification may involve frequency lists; keyword analysis; n-gram extraction; termhood calculations; collocation analysis; part-of-speech patterns.

The extraction of both single-word and multi-word terminology is essential because institutional terminology frequently occurs as complex nominal structures, for example:

international peace and security;
sustainable development goals;
human rights treaty bodies;
peacekeeping operations;
international legal obligations. [4: 230]

A significant theoretical implication of the proposed model is that terminological equivalence should not be interpreted as a single-dimensional relationship.

Analyses. The analysis suggests at least five dimensions.

1. Lexical equivalence

The target term corresponds to the lexical meaning of the source term.

2. Conceptual equivalence: Both forms represent the same specialized institutional concept.

3. Functional equivalence: The terms perform equivalent communicative functions within institutional discourse.

4. Collocational equivalence: Target-language combinations correspond naturally to the phraseological behaviour of the term.

5. Institutional equivalence: The target designation conforms to terminology officially accepted or conventionally established within the relevant institution.

Results. The corpus-based analysis of terminological harmonization in multilingual United Nations institutional discourse demonstrates that terminological correspondence cannot be reduced to direct lexical equivalence between source and target languages. The study reveals that the harmonization of institutional terminology is a multistage process in which semantic, contextual, collocational, functional, and institutional factors interact.

The first major result concerns the **systematic relationship between source-language terms and their translation equivalents**. Analysis of parallel institutional texts shows that a single source-language term may correspond to several target-language variants. These variants may arise because of lexical synonymy, morphological differences, syntactic restructuring, translation tradition, or the influence of previously established administrative and legal terminology. Consequently, the existence of more than one equivalent does not necessarily indicate translation inconsistency. Instead, each variant requires contextual and conceptual evaluation.

The second result is the identification of five principal dimensions of terminological equivalence: *lexical equivalence* → *conceptual equivalence* → *functional equivalence* → *collocational equivalence* → *institutional equivalence*.

Lexical equivalence reflects correspondence at the level of dictionary meaning, while conceptual equivalence determines whether the source and target terms refer to the same specialized concept. Functional equivalence concerns the role performed by the term within institutional communication. Collocational equivalence reflects the compatibility of the term with recurrent lexical combinations, whereas institutional equivalence indicates whether the selected form corresponds to established usage within official or professional discourse.

The analysis therefore suggests that the most reliable terminological equivalent is not necessarily the formally closest translation. A preferred equivalent should demonstrate a high degree of correspondence across all five dimensions.

A further result of the study is the classification of terminological variation into the following major types:

1. **orthographic variation** – differences in the graphic representation of terms;
2. **morphological variation** – differences resulting from word formation or grammatical structure;
3. **lexical variation** – the use of different lexical units for the same or closely related concept;
4. **syntactic variation** – differences in the structural organization of multi-word terms;

5. **translational variation** – the presence of several target-language equivalents for one source term;

6. **conceptual variation** – cases in which apparently synonymous terms express different conceptual scopes.

This classification makes it possible to distinguish acceptable linguistic variation from actual terminological inconsistency.

Another important finding is that frequency alone cannot serve as the principal criterion for terminological standardization. A frequently occurring translation may be historically established but conceptually imprecise, while a less frequent variant may correspond more accurately to the institutional concept. Therefore, terminological harmonization should be based on an integrated assessment of frequency, distribution, contextual compatibility, conceptual accuracy, collocational stability, and institutional acceptance.

Discussion. The corpus-based model developed in this study can be represented through the following sequence: *term extraction* → *parallel alignment* → *equivalent identification* → *semantic verification* → *variation analysis* → *corpus validation* → *harmonization* → *terminological codification*.

The application of this model demonstrates that parallel corpora can perform not only a descriptive but also a normative function. At the descriptive level, corpus analysis identifies the forms actually used in multilingual institutional documents. At the normative level, the collected evidence allows the researcher to determine which equivalent should be treated as the preferred term, which forms may be retained as acceptable variants, and which forms should be regarded as non-preferred.

The findings also reveal the importance of institutional phraseology in the harmonization process. Multilingual UN discourse contains numerous recurrent constructions such as *sustainable development*, *international peace and security*, *human rights mechanisms*, *in accordance with international law*, and *within the framework of*. These structures demonstrate that terminology should be examined not only as isolated lexical units but also as components of stable collocational and phraseological patterns. [5:145]

Accordingly, the corpus analysis establishes the following hierarchy of units requiring harmonization: *term* → *multi-word term* → *institutional collocation* → *institutional phraseological unit* → *recurrent discourse formula*.

The results further indicate that a specialized parallel corpus can reveal the degree of terminological stability across different document genres. Terms repeatedly occurring in resolutions, conventions, declarations, reports, and policy documents with the same target-language equivalent demonstrate a higher level of terminological stability. Conversely, terms represented by several competing equivalents across similar institutional contexts require further harmonization. [7:134]

For English–Uzbek institutional terminology, the proposed framework is particularly significant. The analysis suggests that some Uzbek equivalents of international institutional terms may have developed through both direct English–Uzbek translation and indirect transfer through Russian. Therefore, corpus comparison makes it possible to distinguish between:

English → Uzbek direct terminological correspondence and English → Russian → Uzbek mediated terminological correspondence. This distinction is important for determining whether an Uzbek term represents an internally normalized linguistic form, an established international equivalent, or a historically mediated translation.

On the basis of the obtained analytical results, terminological units may ultimately be divided into three functional groups: [10. 107]

Category	Description
Preferred term	Conceptually accurate, corpus-supported and recommended for standardized use
Admitted variant	Contextually acceptable alternative that does not create conceptual ambiguity
Non-preferred term	Inconsistent, semantically inaccurate or institutionally unstable variant

Thus, the central result of the research is the establishment of a corpus-based mechanism for moving from terminological variability toward systematic harmonization: *multiple translation variants* → *contextual and conceptual comparison* → *corpus verification* → *preferred equivalent* → *harmonized terminology*.

The findings confirm that specialized parallel corpora can provide an empirical foundation for multilingual terminology management. In the context of United Nations institutional discourse, such corpora make it possible to identify recurrent equivalents, detect inconsistencies, evaluate competing forms, and develop evidence-based recommendations for terminological standardization. [6:92]

Overall, the results support the proposition that terminological harmonization should be understood as a multidimensional corpus-driven process, rather than as the mechanical replacement of several variants with one standardized term. The proposed approach integrates linguistic description, translation equivalence, conceptual analysis, and institutional normalization and can therefore serve as a methodological basis for constructing an English–Uzbek database of harmonized United Nations terminology.

Translation practice. Translators can consult corpus-attested equivalents rather than relying exclusively on general dictionaries.

Terminography. Corpus data can support the compilation of multilingual specialized dictionaries and terminology databases.

Translation quality assurance. Inconsistent equivalents can be identified automatically or semi-automatically across large collections of documents.

Translator training. Parallel concordances can help students observe institutional terminology in authentic contexts.

Machine translation. Harmonized terminological databases can function as controlled terminology resources for computer-assisted and machine translation systems.

National terminology development. Languages with comparatively limited institutional terminology resources can establish standardized equivalents by analysing authoritative multilingual corpora. [8:120]

Conclusion. Terminological harmonization in multilingual institutional discourse is not merely a lexical standardization procedure but a complex linguistic, conceptual and translational process. The corpus-based framework developed in this study interprets harmonization as a systematic sequence involving terminological extraction, contextual alignment, identification of equivalents, semantic verification, variation analysis, corpus validation, harmonization and terminological codification.

The principal theoretical contribution of the study lies in conceptualizing terminological equivalence as a multidimensional relationship encompassing lexical, conceptual, functional, collocational and institutional correspondence.

The proposed model also distinguishes terminological variation from terminological inconsistency. While some variation represents a natural feature of multilingual communication, other forms may create conceptual ambiguity and therefore require harmonization.

A specialized parallel corpus of United Nations documents provides an appropriate empirical environment for implementing this methodology because it allows terminology to be studied in authentic, aligned and institutionally constrained contexts. In particular, the development of an English–Uzbek parallel corpus of UN documents may establish an empirical foundation for identifying standardized Uzbek equivalents of international institutional terminology.

References

1. Baker, M. 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* Amsterdam/Philadelphia: John Benjamins. 1993. – P. 233-250.
2. Biel, Ł. *Lost in the Eurofog: The Textual Fit of Translated Law*. Frankfurt am Main: Peter Lang. 2014. – P. 56-59.
3. Bowker, L., Pearson, J. *Working with Specialized Language: A Practical Guide to Using Corpora*. London/New York: Routledge. 2002. – P. 67.
4. Cabré, M. T. (1999). *Terminology: Theory, Methods and Applications*. Amsterdam/Philadelphia: John Benjamins.
5. Faber, P. (Ed.). *A Cognitive Linguistics View of Terminology and Specialized Language*. Berlin/Boston: De Gruyter Mouton. 2012. – P. 145.
6. Laviosa, S. (2002). *Corpus-Based Translation Studies: Theory, Findings, Applications*. Amsterdam/New York: Rodopi. 2002. – P. 59-60.
7. McEnery, T., & Hardie, A. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press. 2012. – P. 234.
8. McEnery, T., Xiao, R., Tono, Y. *Corpus-Based Language Studies: An Advanced Resource Book*. London/New York: Routledge. 2006. – P. 120.
9. Pearson, J. (1998). *Terms in Context*. Amsterdam/Philadelphia: John Benjamins.
10. Sager, J. C. (1990). *A Practical Course in Terminology Processing*. Amsterdam/Philadelphia: John Benjamins. 1990. – P. 37.
11. Sinclair, J. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press. 1991. – P. 102.
12. Stefaniak, K. Terminology work in the European Commission: Ensuring high-quality translation in a multilingual environment. In T. Svoboda, Ł. Biel & K. Łoboda (Eds.), *Quality Aspects in Institutional Translation*. Berlin: Language Science Press. 2017. – P. 35.
13. Temmerman, R. (2000). *Towards New Ways of Terminology Description: The Sociocognitive Approach*. Amsterdam/Philadelphia: John Benjamins.
14. United Nations. *UNTERM: The United Nations Terminology Database*. Department for General Assembly and Conference Management.
15. United Nations, Department for General Assembly and Conference Management. *Translation and Terminology Resources*. United Nations.
16. Wright, S. E., & Budin, G. (Eds.). *Handbook of Terminology Management. Vol. 1: Basic Aspects of Terminology Management*. Amsterdam/Philadelphia: John Benjamins. 1997. – P. 46.