

CORPUS-DRIVEN APPROACHES TO TERMINOLOGICAL UNIFICATION IN UNITED NATIONS INSTITUTIONAL DISCOURSE

Gafarova Zumrad Zokhirjonovna

Associate Professor (PhD) Head of Foreign Language and
Social Sciences Department, Asia International University

gafarovazumrad@oxu.uz

Orcid: 0009-0003-6034-6150

Phone: +998 94 778 86 68

Abstract

Terminological unification is one of the central linguistic problems of multilingual institutional communication. This issue is particularly significant in the discourse of the United Nations, where the same political, legal, social, economic and administrative concepts are repeatedly expressed in different documents and reproduced across several languages. In such a communicative environment, terminology performs more than a nominative function. A term identifies a specific institutional concept, preserves its semantic boundaries and ensures that the same concept is interpreted consistently in different documents. For this reason, variation in institutional terminology may influence not only linguistic form but also the conceptual accuracy of multilingual communication.

Keywords: corpus linguistics, corpus-driven approach, terminological unification, terminological variation, United Nations discourse, institutional discourse, parallel corpus, multilingual terminology, terminological harmonization, semantic equivalence, translation equivalence, institutional terminology, terminology management, specialized corpus.

Introduction

United Nations documentation is characterized by a relatively stable institutional style, but this stability does not necessarily mean that terminology is completely uniform. The same concept may occur in several lexical or structural forms. Some of these variants emerge because institutional language develops over time, while others result from

translation practice, differences between document genres or changes in conceptual interpretation. Consequently, the main problem is not the mere existence of terminological variants, but determining whether such variants represent real semantic distinctions or simply different linguistic realizations of the same concept. [9:10]

A corpus-driven approach provides an effective means of resolving this problem because it allows terminology to be examined through actual usage rather than through isolated definitions. When a specialized corpus of United Nations documents is analyzed, terminological units can be observed repeatedly in comparable contexts. Their frequency, lexical environment, syntactic position, collocations and translation equivalents reveal whether a particular form has become institutionally stable or remains one of several competing variants. This makes corpus evidence especially important for identifying the relationship between terminological variation and terminological unification. [10:311]

One of the most significant features of corpus analysis is that it reveals the difference between formal variation and conceptual variation. Two terms may appear different in form while referring to the same institutional concept. At the same time, two expressions that seem synonymous at the general lexical level may acquire different meanings in institutional discourse. Therefore, formal similarity or dictionary synonymy cannot be regarded as sufficient evidence for terminological equivalence.

This can be observed in institutional expressions such as capacity-building and capacity development, technical assistance and technical cooperation, or peacekeeping operation and peacekeeping mission. These expressions may appear semantically close outside institutional discourse. However, their actual use in United Nations documents may reveal differences in scope, institutional function or contextual distribution. For example, one term may be preferred in development-oriented documents, while another may be associated with a particular programme, institutional body or period. Corpus evidence therefore helps determine whether two terms should be unified or preserved as separate concepts. [1:230]

This distinction is crucial because terminological unification should not mean replacing every lexical variant with a single expression. Such an approach could eliminate conceptually necessary distinctions. The purpose

of unification is rather to reduce variation where several forms express the same institutional meaning without a functional reason for their coexistence. Where the corpus demonstrates that two terms consistently occur in different semantic or pragmatic contexts, they should not be treated as interchangeable. Where they occur in essentially identical contexts and perform the same institutional function, their coexistence may indicate unnecessary terminological duplication.

Materials and methods

The parallel structure of multilingual United Nations documents makes corpus-based analysis particularly valuable. A term used in one language can be compared directly with its equivalent in another language. If the same source-language term is consistently translated by one equivalent across a large number of documents, this indicates a relatively high level of terminological stability. Conversely, if a single source term is translated by several target-language forms in comparable contexts, the corpus reveals a zone of terminological instability. [2:47-50]

Such variation is especially important because multilingual institutional communication depends on conceptual equivalence rather than literal lexical correspondence. A translation may be linguistically acceptable but terminologically inconsistent with established institutional practice. For example, two target-language expressions may both be possible dictionary equivalents of an English term, while only one of them is systematically used in official United Nations discourse. Corpus analysis makes this distinction visible by showing which equivalent is dominant, in what contexts it occurs and whether competing forms correspond to genuine conceptual differences.

Another important dimension of terminological unification is contextual stability. Institutional terms rarely function independently of their surrounding lexical environment. They tend to form stable combinations with particular verbs, adjectives and other specialized expressions. As a result, the meaning of a term is partly reflected in its collocational behavior. If two competing variants regularly appear with the same lexical partners and within the same syntactic structures, this strengthens the assumption that they may represent the same concept. If their collocational profiles differ

systematically, such differences may indicate distinct institutional meanings.

The analysis of collocations is particularly useful in distinguishing near-synonymous terms. General dictionaries may provide similar definitions for two expressions, while corpus data reveal that they participate in different institutional structures. A term that typically occurs with verbs such as implement, ensure, strengthen or monitor may belong to a different functional category from another term occurring with provide, receive, request or deliver. Thus, terminological identity should be evaluated not only through definitions but also through recurring discourse patterns. [7:40]

Frequency also contributes to the identification of preferred terminology. A term used consistently across a large number of resolutions, reports, conventions and policy documents is likely to possess a higher degree of institutional stability than a rare competing form. Nevertheless, frequency should not be treated as an absolute criterion. A frequent term may reflect earlier institutional usage, while a less frequent variant may represent a newer terminological preference. For this reason, corpus-based terminological analysis must take chronological development into account.

Analyses

Diachronic analysis can reveal how institutional terminology changes over time. In some cases, two variants coexist during a transitional period before one gradually becomes dominant. Such processes demonstrate that terminological unification is often evolutionary rather than immediate. An older term may continue to appear in documents because of legal continuity or established textual formulas, whereas newer documents increasingly adopt another expression. Consequently, the preferred form cannot always be identified by total frequency alone; recent institutional practice and the distribution of forms across historical periods must also be considered. [11:49]

The same issue can be observed in socially sensitive terminology. Institutional language changes together with broader developments in international law, human rights discourse and social policy. Expressions that were formerly accepted may gradually be replaced by forms considered

more precise or more consistent with contemporary institutional principles. Corpus analysis allows such shifts to be documented empirically and shows how new terminology becomes conventionalized across different document types.

Multiword terminological units are particularly important in United Nations discourse because institutional concepts are frequently expressed through complex nominal structures rather than individual words. Terms such as international human rights law, sustainable development goals, technical assistance programme and institutional capacity-building illustrate this tendency. The stability of such expressions depends not only on their individual lexical components but also on their internal structure. [9:60]

Changes in one component may create a different institutional meaning. Therefore, corpus-driven unification should focus on terminological combinations as complete units. An isolated analysis of individual words may overlook conceptual distinctions encoded in the phrase as a whole. This is one reason why specialized corpora are more useful than general-language resources for institutional terminology: they demonstrate which combinations are conventionalized within the institution and which occur only sporadically. [15:39]

Terminological unification is also closely related to the distribution of terms across document genres. United Nations resolutions, conventions, reports and administrative texts perform different communicative functions. A term that is highly stable in legally binding documents may exhibit greater variation in analytical reports or policy discussions. Such variation should not automatically be interpreted as inconsistency. Instead, it may reflect differences in communicative purpose and degree of institutional formality. For this reason, a reliable analysis must compare terms within equivalent or comparable genres. If two variants occur in the same type of document, within the same conceptual field and in equivalent syntactic contexts, their functional similarity becomes more evident. If one variant predominates in legal instruments and another in explanatory or technical documentation, this distribution may indicate a genre-specific distinction rather than a terminological problem. [4:89]

Discussion

Corpus-driven analysis therefore demonstrates that terminological unification must be based on several interrelated criteria. Semantic correspondence is necessary because the units under comparison must refer to the same concept. Contextual correspondence is equally important because terms must perform similar functions in actual discourse. Institutional distribution shows whether a form is widely accepted across different bodies and documents. Translation correspondence indicates whether the same conceptual relation is preserved between languages. Chronological analysis reveals whether a particular term represents current or outdated institutional usage.

The interaction of these factors makes it possible to distinguish stable terminology from competing or inconsistent variants. A term that occurs frequently, appears across different document types, maintains a consistent semantic profile and corresponds regularly to the same equivalents in other languages can be regarded as highly institutionalized. A competing term that occurs irregularly or lacks stable translation correspondence may require terminological harmonization. [13:78]

At the same time, corpus-driven research shows that unification and harmonization should not be reduced to simple lexical replacement. Institutional terminology is connected with conceptual systems. Before one variant is selected as preferred, the researcher must establish that the variants actually represent the same concept. Otherwise, apparent standardization may lead to semantic loss. This is especially relevant in legal and diplomatic discourse, where even minor terminological distinctions may have significant interpretative consequences.

A specialized parallel corpus thus becomes more than a database of examples. It provides evidence of how institutional concepts are actually represented, repeated and translated. Through corpus observation, it becomes possible to identify dominant forms, detect unstable equivalents, trace diachronic changes and determine how terminology functions in different institutional contexts.

The principal value of the corpus-driven approach lies in its empirical character. Traditional terminological resources often present preferred terms as already established forms, whereas a corpus demonstrates the linguistic

processes through which such forms become preferred. It shows not only which terms are used but also how consistently they are used, where alternatives occur and whether variation is semantically motivated. [14:95] This perspective is particularly relevant to rapidly developing areas of United Nations activity. Fields such as climate policy, sustainable development, migration, digital governance, cybersecurity and artificial intelligence continuously introduce new concepts into institutional discourse. During the early stages of conceptual development, several competing expressions may coexist. Corpus monitoring can reveal which forms gradually acquire institutional stability and which remain marginal or context-specific.

Conclusion

Therefore, terminological unification in United Nations institutional discourse can be understood as the establishment of stable linguistic representation for institutionally recognized concepts while preserving differences that carry genuine semantic or functional value. The role of the corpus is to provide sufficient textual evidence for making this distinction. The analysis ultimately indicates that effective terminological unification requires a balance between consistency and conceptual precision. Excessive variation may reduce clarity and create problems in multilingual equivalence, while excessive standardization may eliminate meaningful distinctions. Corpus-driven analysis offers a way to maintain this balance because terminological decisions are based on observable patterns of authentic institutional use.

In this sense, specialized parallel corpora can serve as an important mechanism for the development of coherent United Nations terminology. They make it possible to identify recurrent institutional patterns, compare competing forms, examine translation correspondences and determine the degree of semantic and functional equivalence between terms. Terminological unification consequently becomes an evidence-based process aimed at preserving conceptual stability across multilingual institutional discourse rather than merely imposing formal uniformity.

References

1. McEnery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.
2. Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam: John Benjamins.
3. Faber, P. (Ed.). (2012). *A Cognitive Linguistics View of Terminology and Specialized Language*. Berlin/Boston: De Gruyter Mouton.
4. Zanettin, F. (2012). *Translation-Driven Corpora: Corpus Resources for Descriptive and Applied Translation Studies*. Manchester: St. Jerome Publishing.
5. Biel, Ł. (2014). *Lost in the Eurofog: The Textual Fit of Translated Law*. Frankfurt am Main: Peter Lang.
6. Biel, Ł. (2017). Enhancing the communicative dimension of legal translation: Comparable corpora in the research-informed classroom. *The Interpreter and Translator Trainer*, 11(4), 316–336.
7. Šarčević, S. (1997). *New Approach to Legal Translation*. The Hague: Kluwer Law International.
8. Cao, D. (2007). *Translating Law*. Clevedon: Multilingual Matters.
9. Prieto Ramos, F. (2011). Developing legal translation competence: An integrative process-oriented approach. *Comparative Legilinguistics*, 5, 7–21.
10. Prieto Ramos, F. (2014). International and supranational law in translation: From multilingual lawmaking to adjudication. *The Translator*, 20(3), 313–331.
11. ISO. (2009). *ISO 704: Terminology Work — Principles and Methods*. Geneva: International Organization for Standardization.
12. ISO. (2019). *ISO 1087: Terminology Work and Terminology Science — Vocabulary*. Geneva: International Organization for Standardization.
13. United Nations. *United Nations Editorial Manual Online*. New York: United Nations.
14. United Nations. *UNTERM: United Nations Multilingual Terminology Database*. New York: United Nations.
15. United Nations. *United Nations Parallel Corpus*. New York: United Nations.