Qualitative and quantitative methods of official document research

Yuliya Demyanchuk, Lviv State University of Life Safety, Lviv, Ukraine

This paper analyzes official documents for the purpose of determination of specific features of the linguistic sampling and formulation of some provisions and linguistic conclusions on the basis of the obtained results. In an effort to provide a more complete picture of official documents, this study uses both a qualitative approach to document research (Bowen 2009) and a quantitative approach specifically adopted for the analysis of official documents (Weber 1990). The study also attempts to determine the frequency of lexical components related to the official NATO, UN, and WTO documents. Further, the official documents were assessed and classified into three groups and an experimental text corpus was formed. Within the context of needs analysis, the data metric were created that allows tracking the groups of documents by using the distance between words. In consequence, advantages and disadvantages of the qualitative and quantitative methods of analysis of official documents were outlined.

Keywords: *official documents (OD), term combination, common phrase, qualitative method of document research, quantitative method of document research.*

1 Introduction

A distinctive feature of official documents is that it is special type of texts (regulations, decrees, programs, resolutions, orders, protocols and acts of international organizations) with a special lexical and semantic structure, which are dedicated to current problems in this domain. In the present study, we have used an analytical structure consisting of the functional classification developed by Kuchyk. The purpose of the functional classification is to identify the way of organization of official texts of international organizations in the discourse and text corpus (TC). According to Kuchyk's classification (2016), there are three positions that contribute to smoothness of the discourse. In this context, the researcher identifies the following categories of OD: 1) laws, regulations and all issues that affect the daily life of citizens; 2) resolutions, statutes, conventions, declarations, memoranda, agreements – constituent documents with a long history of historical development and consolidation of their text content (Kuchyk 2016: 503–507).

Based on a number of works examining the discourse of the political or academic community (Page et al. 1987; Graber 2010; Morris & Goldsworthy 2008; Raganato & Tiedemann 2018; Sanchez & Martınez 2018) and meeting the key requirements of the studied TC, we analyzed those components of OD that were the most frequent in samples of legal texts representing three different organizations: NATO (North Atlantic Treaty Organization), UN (United Nations Organization), and WTO (World Trade Organization). Actually, the possibility of combining the qualitative and quantitative approaches in the process of study of official documents was proposed by Weber, who stated that the emergence of political and security risks was caused by the improper use of the communication process.

According to numerous studies (Castells 2009; Goldman 2008; Sterling 2008; Weber 1990 and others), the terms of official documents are easily subjected to the linguistic analysis in view of their narrower specialization as compared to functional words. In addition, the

choice of terms for a particular international organization is convenient due to its easy comparison with the terms of other international organizations.

We agree with Goldman (2008) that within the framework of the qualitative and quantitative approaches, the methodology and objectification of data fit into a block of linguistic patterns, often derived from the political thinking or political behavior. Assuming that units determine the frequency of occurrence of features of an international organization in a text and the volume of text that contains a unit (Weber 1990), this study may shed more light on the issue of vector representations of terms in input data and language models and thus help to define an integrated hierarchical approach to the classification of official documents.

2 Official documents

2.1 Definition

The achievements of the past years gave a powerful impetus for the official documents going beyond the narrow framework of the strict discipline (especially international one), and later on official documents became one of the key concepts of the majority of human sciences. Agreeing to the postulate on performativity of formulas of the official business text, Arivazhagan believes "the situational motivation directly related to the worldview of the people" to be their characteristic feature (Arivazhagan et al. 2019: 5). In this case, the system of units of OD is connected with the material, social and spiritual culture of the language community and, therefore, is a reflection of its cultural and national experience at the international level.

The thematic layer of OD plays an important role in reflection of current trends in the international relations, studying the problems of globalization and integration processes, research of information war technologies and activities of the analytical institutes (Bau et al. 2019). There are many linguistic studies that consider the analysis of official documents at the international level (see Bowen 2009: 27–40; Kugler 2007: 8; Sterling 2008: 47 and others).

It should be noted that the official documents create the official terminology, which is most noticeable in international organizations, such as NATO, UN, WTO, where the representation of economic, political, military, and other processes affects the construction of the professional vocabulary. By resorting to grammatical and lexico-semantic experiments (Roelcke 1991; Temmerman 2000), researchers of official documents widely use the appropriate markers that help to outline the content of the official terminology, the semantics of which does not fit into any stereotypical matrix, however it is quite common in an official business environment.

2.2 Relevant research studies

The first researcher who laid the scientific foundations for distinguishing methods of the scientific research of official documents was the American scientist and philosopher Speed. In his article "Do Newspapers Now Give the News?" (1893), based on an array of empirical data obtained from Sunday issues of New York newspapers for 1881–1883, he described the distinctive features of the document (political, literature and religious) and operative information and materials on public issues (such as advertising, entertainment) and factors that affect them. Speed measured the volume of materials on each topic in inches and compared the results. It turned out that newspapers began to pay more attention to the events from the life of

celebrities (gossip and scandals), while less attention was paid to the documentary material. The next publication that claimed the collection, grouping and analysis of a large amount of information in official documents and caused a wide public resonance was the work of Thomas and Znaniecki "The Polish Peasant in Europe and America", written in 1918–1920. The authors of this work that became a classical one, decided to investigate the problems faced by poor Polish emigrants in the United States. They studied 8,000 issues of the newspaper – a special periodical of Polish emigrants in the United States – for the period of over 20 years, the card index and official documents. Of course, this method had both supporters and opponents. For example, the American researcher Blumer (1937), criticizing the work of Thomas and Znaniecki, believed that the material collected by them was purely illustrative and could not be used as a basis for conclusions. Contesting the thesis of the mutual dichotomy of the quantitative and qualitative research strategies, modern researchers (Flyvbjerg 2004; West 2001) note that using only one of them significantly limits the empirical resources of the study – the description and even more the explanation – of the phenomena interesting for the researcher (Borysova 2012: 44–48). The theory of the official terminology proposed by Cabre Castellvi deserves a special attention. It gained popularity at the beginning of the XXI century. In particular, the researcher proposed to analyze an official document based on the quantitative and formal-grammatical content of its components (Cabre Castellvi 2003: 170). Similarly, in the present study we will try to systematize documents, to outline the most common terminology used in official documents of international organizations such as NATO, UN, WTO, to propose a hierarchical approach to the classification of documents, as well as to identify the advantages and disadvantages of the qualitative and quantitative methods for research of official documents.

3 Material and methods

At the first stage of our study, we used a *qualitative method* of analysis of official documents. In this article, we present the results collected from 40 documents of NATO, UN, and WTO. As a result, term combinations and common phrases were separated from collocations. At this stage, we also managed to determine the frequency of occurrence of a term combination and compare it with the frequency of occurrence of a common phrase in official documents.

Using a total of 100 units of texts that represent three international organizations (NATO, UN, WTO), at the next stage of our study we created a text corpus comprising a total of 186.154 words, where the average text size equaled to 4.758 words per text. Using an open source software Orange, written in Python language and used for data analysis and machine learning, developed at the University of Ljubljana in collaboration with the Jožef Stefan Institute, Slovenia, we identified the most common terms with the help of *the quantitative* analysis of official documents and proposed an integrated hierarchical approach to classification of documents with the help of word spacing.
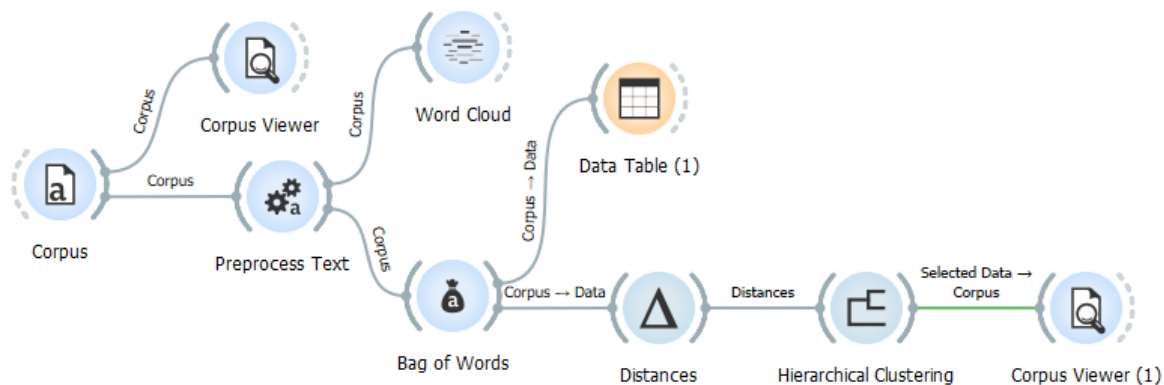
Figure1: Division of texts into groups (clustering)

3.1 *Qualitative Research Method according to Bowen (2009)*

As noted above, official documents of NATO, UN, and WTO were selected for the analysis in view of contextual differences of language units at the linguistic level and a lesser variability degree as compared to general official documents (orders, decrees, laws, declarations). In addition to the selected terminological phrases (a total of 58 units), we also managed to confirm the ambivalence of commonly used phrases in OD that do not affect the textual embodiment of concepts and phenomena of NATO, UN, and WTO.

Based on the assumption that the predominance of certain common tokens and phrases in documents does not affect the textual embodiment of concepts and phenomena that contain signs of the international peace and security, friendly relations between nations based on the principle of respect for equality and self-determination of nations, and international cooperation (Bowen 2009: 27–40), served as the initial hypothesis as mentioned before. The first phase of the linguistic research envisaged a systematic review of documents and enabled the selection of all terminological combinations and common phrases from the official documents contained in the processed collection of documents. At the next stage, we sorted them by the degree of their termhood. The results of the second stage of the linguistic analysis were determination of the number of texts that contain a term combination and a common phrase and development of a vector model for NATO, UN, and WTO term combination and common phrase.

3.2 *Quantitative Research Method according to Weber (1990)*

The purpose of the above-mentioned clustering was to clarify the content units that can be unambiguously recorded and transferred into quantitative indicators, represented by researches carried out by Weber (Weber 1990: 96). Among other results, this analysis was expected to provide information on the most common tokens that moved to the category of terms the semantics of which represent the development of new weapons, reorganization of troops and changes in the military tactics, foreign trade policy of member countries, trade aspects of intellectual property rights, regional trade agreements, regulations on protection and preservation of the environment, protection of personal rights and military security, legal concepts and regulation of the official business relations of international organizations. Based on the research of official documents (Gratta et al. 2020), our results confirmed the dominant

position of one-component WTO terms (see below). In an attempt to provide a deeper understanding of the specialized official terminology of international organizations, we traced groups of documents with the help of word spacing. For this purpose, we used the methodology of Weber (1990) that describes the evolution of a document primarily as a carrier of the retrospective information. This approach contributed to OD clustering (C1 – official WTO documents; C2 – official NATO and UN documents) and formation of an inverted sampling (NATO, UN, WTO terms). The sample generated in the Orange software showed how often keywords that identify an official NATO, UN or WTO document by name are displayed. In order to make the obtained volume of data better controlled in texts of the text corpus, the search limit point was raised to 3 words (exactly this fixed number of content units is sufficient to determine the text title).

4 Data Analysis

4.1 *Qualitative Research Method according to Bowen (2009)*

Examples of the selected documents and analyzed data are shown in Table 1.

Table 1: A Sampling of documents and data analyzed

Documents selected	Data analyzed
Assistance for the refugee and migrant crisis in the Aegean Sea	NATO helps to overcome the outcomes of the humanitarian crisis
The United Nations Stabilization Mission in Haiti	The UN Peacekeeping Mission aimed to overcome the outcomes of the devastating earthquake in Haiti in 2010
Agreement on import licensing procedures	Terms of import of goods into the customs territory of the importing member
Homeward bound? Questions on promoting the reintegration of returning migrants	The strategy is aimed at expanding the rights of the host country to govern and control the borders
System of environmental and economic accounting, 2012. Central Framework	New developments in the ecological and economic accounting and achievements in methodological researches
The regional trade agreements	WTO members involved in trade agreements and negotiations aimed to clarify and improve WTO disciplines
United Nations Mission in Kosovo	The mission is aimed to promote a peaceful and normal life for all people in Kosovo
Statement of the North Atlantic Council on the democratic people′s republic of Korea	Peaceful denuclearization of the Korean Peninsula
Report of the Republic of Trinidad and Tobago. Third International Conference on Small Island Developing States	The report summarizes the results of implementation of the Barbados Plan of Actions and the Mauritius Strategy of Implementation by Trinidad and Tobago
Agriculture committee adopts report on implementation of Bali decision on tariff quotas	Tariff quota administration procedure

In this investigation, we have used text matrices to illustrate categories consisting of two columns: one column for the category name and a second one for quotes illustrating the category as designed by Schreier (2012). Within Schreier (Schreier 2012: 223) this type of text matrix is especially useful for summarising information about coding frame. In this respect, the systematic review of official documents provided the background information that helped us to understand the military, political, and economic context. We processed the information contained in the documents based on the LSP principle (language for special purposes). Given the specifics of the LSP, we identified three approaches to the start of the analysis: 1) the use of the authentic material (official documents of NATO, UN, and WTO); 2) coverage of common terminological constructions by each of the organizations; 3) description of the basic conceptual characteristics regardless of the number of formed tokens, phrases, regular expressions, etc. In particular, the following fragment from the official text demonstrates its affiliation with the WTO:

(1)   ***WTO members*** *continue to conduct negotiations to reform* ***agricultural trade***. *These talks began in early 2000 under the original mandate of the* ***Agriculture Agreement*** *and became part of the Doha Round at the 2001 Doha Ministerial Conference. At the 2013 Bali Ministerial Conference, ministers adopted* ***important decisions*** *on agriculture. More recently, at the 2015 Nairobi Ministerial Conference,* ***WTO members*** *agreed on a historic decision to eliminate* ***agricultural export*** *subsidies, the most important reform of* ***international trade rules*** *in agriculture since the WTO was established* (WTO OD 1).

In the cited text fragment, there are several manifestations that emphasize the affiliation with the WTO. This is reflected in the terminological constructions: *WTO members, agricultural trade, Agriculture Agreement, important decisions, agricultural export, international trade rules.* This choice of the research material allowed us to cover a large layer of the specialized OD vocabulary that takes a proper place in the international process of the world community. The scope of the context is determined by the limits that prevent double interpretation of the meaning of a word impossible, thus helping to interpret it correctly.

According to the LSP analysis, the highly specialized terminology of the UN is often used in a form of stable constructions that reflect the legal sphere of the activity:

(2)   *Since the beginning of the year, the* ***UN Migration Agency*** *has carried out over one million* ***health consultations*** *for displaced and conflict-affected Yemenis and* ***migrants*** *struggling* ***to reach the help*** *they need, the organization announced on Friday* (UN OD 2).

The text fragment is saturated with the official UN terminology: *UN Migration Agency, health consultations, migrants.* An essential aspect for the content of the official UN text is the use of a stable structure *(to reach the help)* that demonstrates *the need to help victims.* The combination of the preposition '*since*' mentioned in the text that is part of the fixed expression *(since the beginning)* and the descriptive phrase "*the organization announced*" gives grounds to conclude that this is a fragment of the official text. Also, there is no doubt about the systematic nature of the official text, as indicated by Kukharenko: "the text presents a structured system where all components are interconnected and interdependent" (Kukharenko

2004: 13). Thus, the fragment of the text reveals the semantic load of not only individual tokens, but also of phrases, sentences and supraphrasal units.

The difference of the prism of vision of the established, highly specialized NATO's terminology from the common vocabulary is projected onto the text, what distinguishes it, for example, from the fiction one. Highly specialized terms are intended to identify the international organization and its official documents as confirmed by the LSP principle:

(3)   *In practice, the Chairman of the **Military Committee** presides over the Military Committee, where each member country has a **military representative** (or Milrep) for his/her **Chief of Defence**. This committee – NATO's most senior **military authority** – provides the **North Atlantic Council** and the **Nuclear Planning Group** with consensus-based **military advice** – that is, advice agreed to by all of NATO's Chiefs of Defense* (NATO OD 3).

Generalized concepts of NATO's governing bodies: *Military Committee, military authority, North Atlantic Council, Nuclear Planning Group* are designed to combine the military context in an official document where military terminological combinations of an adjective with a noun *(military representative, military authority, military advice) serve as identifying features.* These features are endowed with a respective military meaning, and a set of them is the most universal method of describing semantics in the official text.

In an effort to report absolute frequencies we integrated them into a text matrices which we have created for presenting the categories in our coding frame. To do so, the methodology of Schreier (2012) was employed which has led to integrating the frequency information.

The tables below demonstrate thirty-two most common term combinations for each international organization.

Table 2: Identified UN term combinations

Identified term combinations	Words in the document	The number of occurrences of the term combination in a document	The number of occurrences of the term combination in all documents	Collocation from which the term combination was singled out
perplexing environment	3303	4	10	*perplexing environment* of the Middle East
natural environment	3812	5	12	respect the *natural environment*
international environment	2536	6	9	that can afflict the *international environment*
political environment	2536	4	11	the current highly charged *political environment*
external environment	2020	3	8	the country's *external environment*
safe environment	1840	3	5	a *safe environment* for

				Syrian refugees
Average value	2674.50	4.2	9.2	

Table 3: Identified NATO terms

Identified term combinations	Words in the document	The number of occurrences of the term combination in a document	The number of occurrences of the term combination in all documents	Collocation from which the term combination was singled out
intense attack	919	1	1	a much more *intense attack*
Russia attack	969	2	4	*attack* by *Russia*
successfully attack	4 252	2	2	*successfully attack* America and Europe
nuclear attack	1364	1	4	possible *nuclear attack*
chlorine attack	1399	2	2	*chlorine attack* in eastern Ghouta
area attack	5283	3	6	the type used in the *area attack*
terror attack	1922	1	4	fatal *terror attack* by a follower
first attack	47,472	6	8	when the *first attack* had already begun
missile attack	4004	2	3	warning of a *missile attack*
German troops attack	240,007	3	3	*German troops attack* successfully on Northern and Southern
air attack	240,007	2	6	it was the *air attack* on Pearl Harbor
night attack	67,683	4	6	the *night attack* had exhausted Krymov
abrupt attack	215,248	5	5	but any *abrupt attack*
Average value	63,886.9	2.6	4.2	

Table 4: Identified WTO terms

Identified term combinations	Words in the document	The number of occurrences of the term combination in a document	The number of occurrences of the term combination in all documents	Collocation from which the term combination was singled out
World Trade	2784	2	9	*World Trade* Organization
trade-war	832	1	5	is Reborn as *Trade-War* Refuge
global trade	1913	2	5	*global trade* framework
momentum trade	1715	1	1	Would call a *momentum trade*
free trade	3072	3	5	the UK *free trade* area
China trade	2985	2	3	U. S. – *China trade* relations
international trade	1819	2	4	to impact *international trade*
automotive trade	1291	1	1	NAFTA's rules on *automotive trade*
simmering trade	1198	1	1	*simmering trade* war with China
dairy trade	1426	1	3	U. S. – Canada *dairy trade*
agricultural trade	1705	2	5	it won't go any further in opening *agricultural trade*
bilateral trade	2286	3	4	nonetheless related *bilateral trade*
foreign trade	2067	4	6	structural and *foreign trade* policies
Average value	1930.2	1.92	4	

In our research, each individual term combination carries certain key information that does not go beyond the narrow scope of issues and embodies a set of formulated views on a particular international organization. As a repeater of the conceptual content, the identification and wording of dominant positions (broader context, valuable syntactic constructions, influence of other elements of a text on the semantic identification) (Bowen 2009: 27–40) occur in the OD texts. NATO/WTO/UN official texts present one of the aspects of such finding. The consistency of use of the complex methodology was aimed at identification of a term combination, its differentiation and delineation of textual boundaries, what reveals the meaning of the studied phenomenon and proposes to consider the international organization in a format

of the official business text. In addition, the inventory of empirical materials envisaged correlation of the NATO, UN, WTO term combination with a common phrase in the OD. Using the methodology of Schreier (2012), we created the chart where the abstract frequency information was transformed into graphic and this graphic shows what the coding frequency for one category is in comparison to another.

The influence of the text corpus on the transformation processes of the triad of international organizations is systematically described in Figure 2.
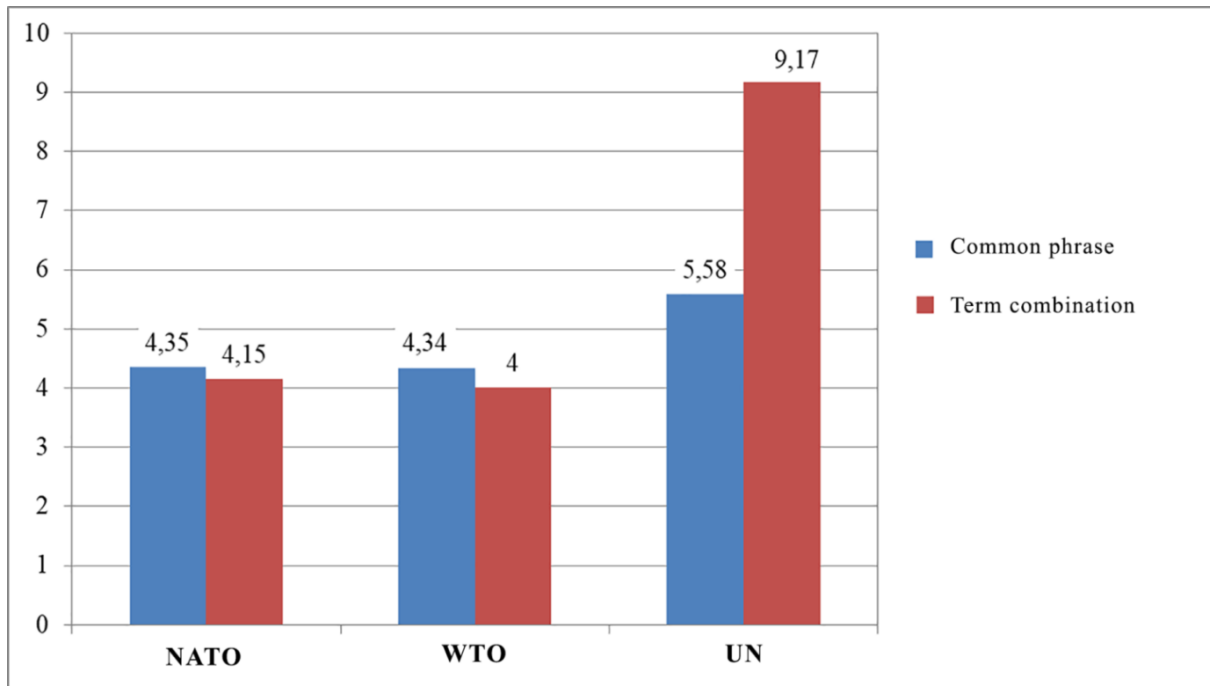

Figure 2: Indicator of inclusion of the NATO, UN, WTO term combination and a common phrase in the OD

Our calculations show that the most productive term combinations of the OD collection form fragments of the official text of the UN and fragments of the official text of NATO. In view of the presented results, though the occurrence of the term combination in the official texts of the WTO is less, the formal and logical presentation of facts and linguistic phenomena does not raise any doubt among the readership and does not complicate the procedure of interpretation of the official text.

Thus, the LSP became a dynamic tool of the high-quality text research method that served as a means of establishing the specialized terminology within a particular professional environment. In our work, this is represented by: 1) terminological units that reveal the content of WTO economic agreements and rules of the international trade; 2) clearly expressed terminology of the UN official text that reproduces the meaning of the legal sphere of the international organization, where the economic and environmental security, protection of individual rights and military security are emphasized; 3) the structure of the NATO official terminology that consists in its uniformity (military one-component lexemes, military terminological phrases, military fixed constructions) on the one hand and the uniqueness on the other hand, what is a valuable asset of an international organization. This spectrum of the functional load of the authentic OD material is enhanced by tokens with the specifying

semantics, expressed with the help of language means. The unifying link in most official texts of NATO, UN, and WTO is their highly specialized military, economic, and legal terminology that ensures the unity of the content plan and expression. The parenthetical constructions (lexemes, phrases, term combinations) identified during the analysis of the text array are not exhaustive and envisage a possibility of further deepening into the problem of reproduction of the OD with the help of a qualitative method of text research.

4.2 Quantitative Research Method according to Weber (1990)

The official documents selected for the quantitative analysis were classified into 3 groups: 1) official texts that disclose the linguistic features of the modern business communication (Himmelmann 2006), language formulas in diplomatic texts (Mahmoud 2020), acquired knowledge about the OD fixed with the help of linguistic tools and objectified by lexico-semantic and syntactic units and creating an integral semantic whole (Reiman 2009); 2) official texts, the content of which reflects the problems of international organizations (F. Cripps, A. Izurieta (Cripps, Izurieta *The UN global policy model (GPM): technical description,* Burnard *The text encoding initiative: an overview,* Dahlburg *NATO creates 'spearhead' to deter Russia,* Donnelly *Universal human rights in theory and practic,* Farley *Military-to-civilian career transition guide: the essential job search handbook for service members,* Grainger *Trade facilitation: a review,* Holt *Women in conflict in the Middle East: Palestinian refugees and the response to violence*, Ignatieff *Human rights as politics and idolatry,* Kugler *The NATO response force 2002–2006: Innovation by the Atlantic Alliance,* Vincent *Human rights and international relations,* Weinrod *NATO command structure: considerations for the future. Center for technology and national security policy* and others); 3) official texts in the political discourse (Auboin *World Trade Organization / Economic research and statistics division,* Jeker *Agreement on preshipment inspection,* Salama *NATO science for peace and security series C: environmental security,* Heuser *The development of NATO's nuclear strategy,* Ruys *The use of force in international law: a case-based approach* and others). The proposed classification significantly exceeds the scope of Kuchyk's classification of official documents (Kuchyk 2016: 503–507). This choice of the research material, first of all, served as a trigger for the modern implementation of the NATO, UN, and WTO terminology in the experimental text corpus, secondly, indicates the intertextual links between official documents of the triad of international organizations, and thirdly, shows that the problem of semantic OD units fits into the existing system of search for a new meaning, and terms take a proper place in information processes of the corpus linguistics.

As mentioned before, Word Cloud tool displays a cloud of words that are most common in texts. The data (see Fig.3) show a significant predominance of terms in official WTO documents in the experimental text corpus. In general, our analysis excludes punctuation, prepositions, pronouns and conjunctions and provides grounds for conclusions about the possibility of identification of an international organization with the help of nouns and descriptive compounds. The correlation of the official terminology in the text corpus is systematized in Figure 3.

Figure 3: Frequency of occurrence of the official terminology in the experimental TC

In our research, the highlighted words in the larger font occur in the text corpus more often than others. It is also important to note that the center of Word Cloud contains the word with the highest occurrence among all official documents of the experimental text corpus ('trade' – 538 times). In an attempt to calculate how often each category was coded across all units of analysis (Mayring 2022) we presented absolute frequencies and integrated it into the chart. The Figure below shows 28 most common terms in the experimental text corpus out of 197 words.
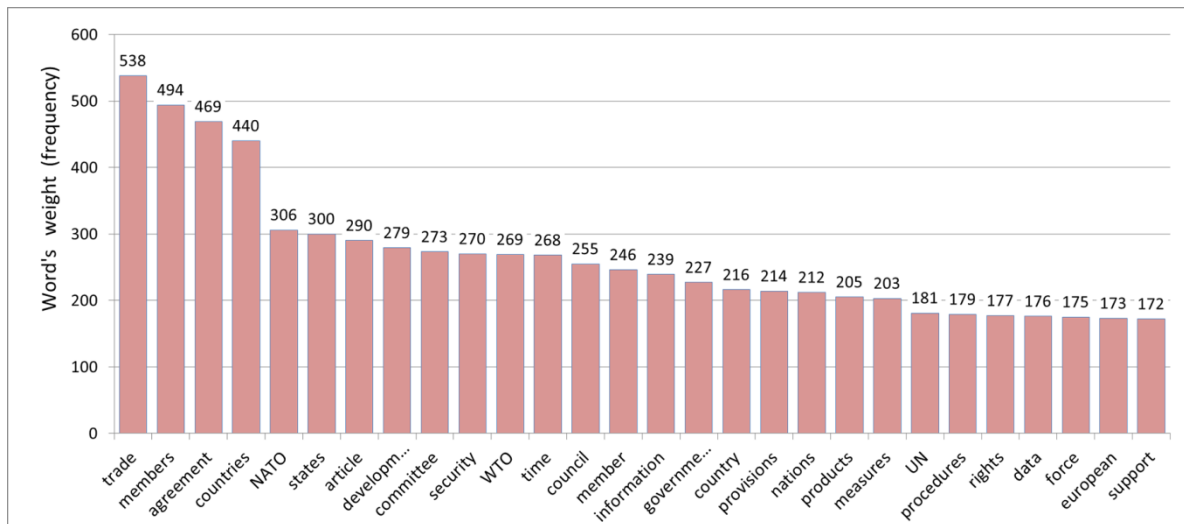

Figure 4: Twenty-eight most common terms in the experimental text corpus

In the process of analysis, we created a data metric that plays an important role in the program and allows tracking groups of documents with the help of word spacing, as well as offers the additional research capacity in the field of terminology research in the official NATO, UN and WTO texts. In our research, we present the results of assessment of vector representations based on the OD. This allows calculating the semantic similarity between words, sentences and

documents, and the processing of the natural language is based namely on these possibilities (Weber 1990). The way of representation of a word in input data and in language models consists in using the cosine distance. In this experimental text corpus, we also applied the calculated distances to the hierarchical clustering of documents. In compliance with Schreier (2012), in terms of the hierarchical levels in our coding frame, the main categories constitute three, higher levels and two subcategories, lower levels. In the course of this analysis we have made use the inferential statistics (Mayring 2022) in order to tell any differences between categories that we have observed. Figure 5 shows the results of the calculation in the process of comparing documents containing the terminology of NATO, UN, and WTO.
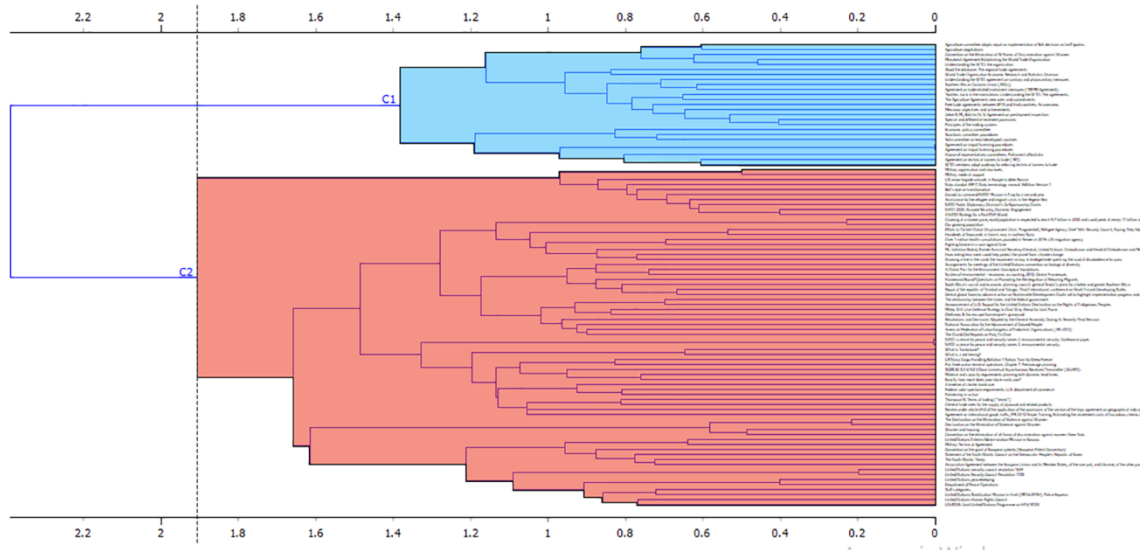

Figure 5: Cosine Distance between NATO, UN, WTO documents

The data obtained as a result of clustering were shown by 2 groups of documents: C1 – 25 documents with the keywords that identify the official WTO documents by name, and C2 – 75 documents that identify NATO and the UN.

Thus, the research clearly demonstrates that the formalized analysis (Weber 1990) is an important step in the quantification processing of a text with the further interpretation of results. This indicates the need for an empirical content-analytical analysis of other official documents.

5 Discussion

As the results of the present study show, the segmentation of our material by qualitative research means dividing it into units such that each segment/unit fits into one category of our coding frame. All over coders are important in qualitative content analysis. This finding coincides with the position of Bowen (2009), who expresses that in some way, each one of the coders and the themes generate serve to integrate data gathered by different methods and they are useful in providing a "behind-the-scenes look" at some sub-projects and follow up activities that were not observed. Furthermore, the segmentation ensures that coders are interpreting and referring to the same parts of our material. In the terminology of quantitative research, we compare the main categories to variables, and the subcategories to the levels of these variables. The coding frame acts like a filter: material that is not covered by the main categories is no

longer visible after we have conducted our analysis. Besides, the distinctions in the material that are not covered by our sub-categories are also invisible. Consequently, we consider every single element from qualitative content analysis and refuse each and every excessive element after conducted quantitative analysis.

As a matter of fact, in qualitative research, case selection, segmentation of the material, methods for data collection, and quantitative research as the method for data analysis are reviewed separate steps within our investigation.

In line with research studies dealing with qualitative content analysis (Schreier 2012; Tomankova 2016; Mayring 2022) the results confirmed the reliability and validity are rooted in a quantitative, positivist conception of research. The authors found out that both methods of research are necessary to use in order to present reliability and validity of information. As claimed by researchers, there is no clear dividing line between qualitative and quantitative content analysis and they are closely related. Additionally, there is no obvious alternative to quantitative research. Within the quantitative research, we check all positions concerning qualitative quality criteria, such as credibility, trustworthiness, auditability and authenticity.

6 Conclusions

Given the dominant position of the term combination in official texts (Azzopardi 2015) and according to Stubbs (2001), who identifies the text corpus as an important basis that can be used to document the language use norms and interpret individual texts, the purpose of the present research was to clarify the specifics of the official NATO, UN, and WTO documents and their content units.

The qualitative analysis was aimed at a systematic review of documents, reflection of the frequency of use of terminological units and their correlation with the most common phrases in the OD. On the other hand, the second quantitative analysis was designed to provide an alternative view on the study of official documents that remain unnoticed in the process of the qualitative analysis (Tomankova 2016: 75–94). As noted before, the second analysis was aimed only at identification of the most common terms that occur in the experimental text corpus, as well as at creation of a hierarchical sequence of official documents. Both types of the analysis are complementary and make up for the shortcomings of each other (Flyvbjerg 2004: 15–18).

Having analyzed official documents by using the methods of qualitative and quantitative analysis, we identified their advantages and disadvantages depending on the type of the research (Table 5).

Table 5: Advantages and disadvantages of the qualitative and quantitative methods of the research

Qualitative method of the OD analysis		Quantitative method of the OD analysis in the experimental TC	
Advantages	*Disadvantages*	*Advantages*	*Disadvantages*
	mechanical processing of a text	automatic processing of a text	
identification of a term		identification of a one-	does not

combination, common phrase		component term	identify a term combination
lexico-semantic analysis → analysis of the whole fragment of a text		quantitative analysis → excludes punctuation marks, prepositions, pronouns, conjunctions	
determination of the frequency of occurrence of a term combination in selected fragments of texts		determination of the frequency of occurrence of one-component terms in the whole text corpus	
correlation of a term combination with a common phrase			the ratio between a one-component term and a token is not envisaged
Hierarchical clustering			
		monitors the OD texts → directs to the NATO, UN, WTO text	

In general, the qualitative and quantitative approaches indicate certain empirical patterns that have been identified by selecting the most common terms and term combinations from official documents. For example, in the analyzed collection of 40 official texts, the occurrence of WTO and NATO term combinations is the highest, as well as in the experimental text corpus containing a total of 100 units. However, the result of comparison of the NATO, UN, and WTO term combination with a common phrase was the opposite. The results showed that the occurrence of the WTO term combination in the official document is the lowest, while the occurrence of the common phrase is the highest, what was confirmed by means of the qualitative analysis.

Sources

WTO OD 1 – *Agriculture negotiations.* (https://www.wto.org/english/tratop_e/agric_e/negoti_eNEW.htm) (Accessed 2022-02-4).
UN OD 2 – Over 1 million health consultations provided in Yemen in 2019: UN migration agency. (https://news.un.org/en/story/2019/11/1051521) (Accessed 2022-01-20).
NATO OD 3 – Military organization and structures. (https://www.nato.int/cps/en/natohq/topics_49608.htm) (Accessed 2021-12-10).

Abbreviations

LSP – Language for special purposes
NATO – North Atlantic Treaty Organization
OD – Official documents

TC – Text corpus
UN – United Nations Organization
WTO – World Trade Organization

References

Arivazhagan, Naveen & Bapna, Ankur & Firat, Orhan & Aharoni, Roee & Johnson, Melvin & Macherey, Wolfgang. 2019. The missing ingredient in zero-shot neural machine translation. (Paper presented at the International conference on learning representations, ICLR 2019. March, 2019.) (https://doi.org/10.48550/arXiv.1903.07091) (Accessed 2022-02-14).

Azzopardi, Annalies. 2015. Dominant position: a term in search of meaning. *Global Antitrust Review* 6. 6–26.

Bau, Anthony & Belinkov, Yonatan & Sajjad, Hassan & Durrani, Nadir & Dalvi, Fahim & Glass, James. 2019. Identifying and controlling important neurons in neural machine translation. (Paper presented at the International conference on learning representations). (https://doi.org/10.48550/arXiv.1811.01157) (Accessed 2022-04-14).

Blumer, Herbert. 1937. Social Psychology. Chapter 4. In Schmidt, Emerson Peter (ed.), Man and Society: A Substantive Introduction to the Social Science, 144–198. New York: Prentice-Hall, Inc.

Borysova, Yuliia. 2012. Kilkisni ta yakisni aspekty vymiriuvannia v sotsialnykh doslidzhenniakh. [Quantitative and qualitative aspects of measurement in social research]. *Philosophy and political science in the context of modern culture* 4(3). 34–37.

Bowen, Glann. 2009. Document analysis as a qualitative research method. *Qualitative research journal* 9(2). 27–40.

Cabre Castellvi, M. Teresa. 2003. Theories of terminology: their description, prescription and explanation. *Terminology* 9(2). 163–199.

Castells, Manuel. 2009. *Communication power*. Oxford: Oxford University Press.

Flyvbjerg, Bent. 2004. Case study in the context of qualitative and quantitative issues. *Sociological research* 7. 15–19.

Goldman, Emily. 2008. *Strategic communications. Theory and application.* Washington, DC: Department of Defense.

Graber, Doris A. 2010. *Mass media and American politics*. Washington, D.C.: CQ Press.

Del Gratta, Riccardo & Boschetti, Federico & Bambaci, Luigi & Sarnari, Francesco. 2020. Approaching document analysis with a formal model. (Paper presented at the 6th IEEE Congress on Information Science and Technology (CiSt), Agadir-Essaouira, Morocco, 5–12 June 2020).

Himmelmann, Nikolaus P. 2006. Chapter 1 Language documentation: What is it and what is it good for? In Gippert, Jost & Himmelmann, Nikolaus P. & Mosel, Ulrike (eds.). *Essentials of Language Documentation*. Berlin, New York: De Gruyter Mouton.

Kuchyk, Halyna. 2016. Stylistychni rysy movy ustanovchykh dokumentiv mizhnarodnykh orhanizatsii. [Stylistic features of the language of the constituent documents of international organizations]. *Young Scientist* 5(32). 503–507.

Kugler, Richard L. 2007. *The NATO response force 2002–2006: Innovation by the Atlantic Alliance. Case studies in defense transformation*. Washington D.C: National Defense University, Center for Technology and National Security Policy.

Kukharenko, Valeriia. 1988. *Interpretatsyia teksta*. [Text interpretation]. Moskva: Prosvieshchieniie.

Mahmoud, Altarabin. 2020. *On translating Arabic and English media texts*. UK: Cambridge Scholars Publishing.

Mayring, Philipp. 2022. *Qualitative content analysis: A step-by-step guide*. Thousand Oaks, CA: Sage Publications Lt.

Morris, Trevor & Goldsworthy, Simon. 2008. *Public relations for the new Europe*. London: Palgrave Macmillan.

Page, Benjamin I. & Shapiro, Robert Y. & Dempsey, Glenn R. 1987. What moves public opinion? *American political science review* 81(1). 23–43.

Raganato, Alessandro & Tiedemann, Jörg. 2018. An analysis of encoder representations in transformer-based machine translation. (Paper presented at the 2018 EMNLP workshop blackbox NLP: Analyzing and interpreting neural networks for NLP, Brussels, Belgium, November 2018). 287–297.

Reiman, D. Will. 2010. Basic oral language documentation. *Basic oral language documentation* 4. 254–268.

Roelcke, Thorsten. 1991. Das Eineindeutigkeitspostulat der lexikalischen Fachsprachensemantik. *Germanistische Linguistik* 19. 194–208.

Sanchez, Christian & Martınez, Paloma. 2018. A simple method to extract abbreviations within a document using regular expressions. (Paper presented at the third workshop on evaluation of human language technologies for Iberian languages, IberEval 2018, Sevilla, Spain, 18 September 2018). 102–107.

Schreier, Margrit. 2012. *Qualitative content analysis in practice.* Thousand Oaks, CA: Sage Publications Lt.

Speed, Jno. Gilmer. 1893. Do newspapers now give the news? *The Forum*. 705–711.

Sterling, Christopher H. (ed.). 2008. *Military communications: From ancient times to the 21st century*. Santa Barbara: ABC CLIO.

Stubbs, Michael. Words and phrases. Corpus studies of lexical semantics. (http://www.uni-trier.de/fileadmin/fb2/ANG/Linguistik/Stubbs/stubbs-2001-words-phrases-ch-1.pdf) (Accessed 2021-12-15).

Temmerman, Rita. 2000. *Towards new ways of terminology description: the sociocognitive approach*. Amsterdam-Philadelphia: John Benjamins Publishing Company.

Tomankova, Veronika. 2016. Lexical bundles in legal texts corpora – selection, classification and pedagogical implications. *Discourse and interaction* 9(2). 75–94.

Thomas, William I. & Znaniecki, Florian. 1918–1920. The polish peasant in Europe and America. Monograph of an immigrant group. Volume I. Primary-group organization. Boston: Richard G. Badger The Gorham Press.

Weber, Robert P. 1990. *Basic content analysis*, 2nd ed. SAGE Publications, Inc.

West, Mark D. (ed.) 2001. *Theory, method and practice in computer content analysis*. Westport, CT: Ablex.

Yuliya Demyanchuk
Foreign Languages and Translation Studies Department
Educational and Scientific Institute of Psychology and Social Protection
Lviv State University of Life Safety
y.demianchuk@gmail.com