



Social

**INTERNATIONAL JOURNAL OF RESEARCH –  
GRANTHAALAYAH**  
A knowledge Repository



## THESAURUS MODELLING OF MODERN LEGAL METALANGUAGE

**Svitlana Matvieieva** <sup>\*1</sup>

<sup>\*1</sup> National Pedagogical Dragomanov University, Kyiv, Ukraine

---

### Abstract

The article deals with the problem of creating thesaurus for the English-Ukrainian parallel legal corpus. Such a corpus enables translators and researchers optimize their work with legal texts and their language units – legal terms, choose the best variants for translations, dictionaries, ontologies, etc. The paper describes the problem of a term conceptual structure, its functions, and relations within the special legal texts organized in the sentence-aligned corpus. The author of the manuscript demonstrates the main principles of work with the training parallel legal corpus created by the author.

**Keywords:** Thesaurus; Translation; Metalanguage; Legal Term; Parallel Corpus.

**Cite This Article:** Svitlana Matvieieva. (2020). “THESAURUS MODELLING OF MODERN LEGAL METALANGUAGE.” *International Journal of Research - Granthaalayah*, 8(2), 89-92. <https://doi.org/10.5281/zenodo.3692595>.

---

### 1. Introduction

The need for fast and high-quality translation of a large number of legal documents today requires translators and translation theorists to search for new approaches to improving and optimizing the process of such translation. Corpus-related studies and creation of a parallel legal corpus provide new opportunities in the field of legal translation.

One of the main tasks on the way to creating a parallel corpus is the development and creation of a term system, which includes, among other things, the terminology thesaurus of a certain subject field. Today, thesaurus modelling of a legal texts is considered to be a particularly promising and efficient approach to systematizing the metalanguage of law to create a parallel legal corpus and translate legal documents. As L. Bowker states, the corpus-based translations are of higher quality in respect to subject field understanding, correct term choice, and idiomatic expression [1].

Today thesaurus is understood as a specially compiled normative vocabulary of lexical units of data-base query and natural languages, designed to search for words by their meaning [2], as well as to arrange the conceptual space of a subject area or the lexical space of a natural language according to certain structural and thematic parameters [3]. Basically, thesauri are built manually. Selected language units and constructions, their relationships and principles of interaction become

the basis for the development and creation of a thesaurus model of a particular scientific field. There are monolingual and multilingual, universal and specialized thesauri. Creating the parallel legal corpus requires the creating a bilingual specialized legal thesaurus.

## 2. Materials and Methods

Within the research a parallel legal corpus was created by us. It contains the texts of the decisions of the European Court in English, which are on open access on the “European Court of Human Rights” official website [4], and translations of these judgments into Ukrainian, provided on the “ECHR: Cases, Opinions, Matter. Ukrainian Aspect” website [5]. These texts are organized by us in the sentence-aligned English-Ukrainian parallel legal corpus, which consists of about 550,000 words and is used by us for the purpose of conducting research, namely for the extraction and further analysis of language units of all levels, and creation of a legal thesaurus.

50	The applicant moved out.	Заявниця вийшла.	001_en-001_ukr
51	Mr S. became abusive and threatened to kill the applicant and her son if she refused to come back to live with him.	Пан С. став жорстоким і погрожував вбити заявницю та її сина, якщо вона відмовиться повернутися жити до нього.	001_en-001_ukr
52	2. January 2016: Abduction and assault	2. Січень 2016: ВИКРАДЕННЯ ТА НАПАДИ	001_en-001_ukr
53	On 1 January 2016 the applicant lodged a report with the Ulyanovsk district no. 2 police, complaining that S. had damaged the windscreen of her car and taken her identity papers.	1 січня 2016 року заявниця подала заяву до Ульяновського відділку поліції, поскаржившись, що С. пошкодив лобове скло автомобіля та забрав у неї документи, що посвідчували її особу.	001_en-001_ukr
54	On the following day the applicant withdrew her report, claiming that she had found her papers.	Наступного дня заявниця відкликала свою заяву, заявляючи, що знайшла свої документи.	001_en-001_ukr
55	On 5 January 2016 the police declined to institute criminal proceedings, stating that as the documents had been found and as S. had replaced the broken windscreen, no crime had been committed.	5 січня 2016 року поліція відмовилася порушувати кримінальне провадження, заявивши, що оскільки документи були знайдені та С. замінив розбите лобове скло, то злочину не було вчинено.	001_en-001_ukr
56	On 6 June 2016 a supervising deputy prosecutor ordered an additional inquiry, which ended in the issuance of another decision refusing to prosecute S. on the grounds that his actions had not constituted any offence.	6 червня 2016 року заступник прокурора, що здійснював нагляд, призначив додаткове розслідування, яке закінчилося винесенням рішення про відмову в притягненні до кримінальної відповідальності С. на підставі того, що його дії не становили жодного правопорушення.	001_en-001_ukr

Figure 1: Sample from the English-Ukrainian parallel legal corpus

Creating a bilingual thesaurus of legal texts has two main tasks: selecting texts for further processing with their indexing, and incorporating parallel legal texts into the corpus for improving the quality of translation (including automatic) of legal texts by taking into account all kinds of relationships (cohesion, coherence, associations, etc.) of language units and specifics of their functioning under the influence of contextual environment and extra-lingual factors.

Table 1: Selected sentences from the English-Ukrainian parallel legal corpus with the term *a perpetrator* and its context in English and translation into Ukrainian

Sentence number	Source language (English)	Target language (Ukrainian)
128	They typically should <u>order a perpetrator to vacate</u> the residence of the victim for a sufficient period of time and <u>prohibit the perpetrator from entering</u> the residence or <u>contacting</u> the victim.	Зазвичай вони повинні <u>зобов'язувати винного звільнити</u> місце проживання жертви протягом достатнього періоду часу та <u>забороняти йому в'їжджати</u> до місця проживання або <u>контактувати</u> з жертвою.
387	In terms of content, protection orders may <u>order the perpetrator to vacate</u> the family home, <u>stay</u> a specified distance away from the victim and her	Щодо змісту, захисні приписи можуть <u>наказати кривднику звільнити</u> сімейний будинок, <u>перебувати</u> на певній відстані від потерпілої та її

	children (and other people if appropriate) and some specific places and <u>prohibit the perpetrator from contacting</u> the victim.	дітей (та інших людей, якщо це доречно) та деяких конкретних місць, а також <u>заборонити йому контактувати</u> з жертвою.
--	---	--

The methodology for creating a thesaurus provides for the development of a system of relations between units of a particular term system. To ensure the process of automatic replenishment of the parallel legal corpus with new documents, it is necessary to determine and systematize the criteria for selecting texts to be included in the corpus. Identification of legal descriptors (a word, phrase, or alphanumeric character used to identify an item in an information storage and retrieval system [6]), which ensures an automatic search for a unit, a construction and the whole text, provides the basis for creating the parallel legal corpus. Today, there are three main principles of classifying semantic relations between descriptors: subsumption (or hierarchical), synonymous (or equivalent) and associative [7]. A detailed analysis shows the presence of a large number of classification criteria (for example, *whole – part*, *object – subject*, *object – state*, *object – function*, *object – subject – process*, etc.). Applying the frequency criterion also plays an important role.

### 3. Results and Discussions

The development of search functions of the parallel corpus based on replacing text units with the search descriptors provides new opportunities for the analysis, indexing and selection of texts by using not only direct dictionary equivalents, but also involving subsumption relations and associative links. Thus, there is a search not only for texts containing the indicated lexical units, but also for texts that are in certain semantic relations with these units. Such an approach brings terminologists closer to solving the problem of combining diverse aspects within one term being indexed.

The process of selecting descriptors is preceded by a process of cognitive modeling of thesaurus knowledge of a particular scientific field, analysis of the content and scope of reference concepts, development of a system of their relationships and analysis of the frequency of use of units in authentic specific texts in the source and target languages. In the following steps, the thesaurus system of a certain metalanguage is being built in the aspect of combining the semantic components of linguistic units with their logical-conceptual and functional content.

The difficulty of creating a bilingual thesaurus is the difference in the lexical and semantic content and range of concepts in different languages, in synonymic and subsumption relations between units, in the points of intersection of lexical meanings, in the entry of lexemes and constructions to different numbers of semantic groups and subgroups etc. That is, when creating such a thesaurus, the first task is to analyze the semantics and semantic components of the units and their relationships in the source and target languages. Only a great number of examples of the word usage provides the translator with the material for a deep semantic analysis of related language units and making the right translation decision. In addition, such thesauri have great potential both for researchers of one language and for scholars working in the field of comparative linguistics and translation theory.

#### 4. Conclusions and Recommendations

A significant advantage of the corpus thesaurus in comparison with the traditional dictionary is its openness for constant replenishment and changes, which ensures the timely response of the thesaurus to changes and transformations of the legal metalanguage.

Today, a parallel legal corpus with the Ukrainian language among other corpus languages is at the creation stage. One of the problems faced by the compilers of such a corpus (along with a number of objective problems such as the qualification of legal translators, the complexity, time consuming and costly characteristics, etc.) is the problem of ambiguity, duplicates (often due to the presence of several versions of translation of the same document without coordinating terminology consistency) of legal terms and discrepancies, double meaning and legislation conflicts. All these factors create an urgent basis for further research both in the field of translation of legal texts, and in the field of creating new and powerful tools for performing professional legal translations.

#### References

- [1] Bowker, L. (1998). Using Specialized Monolingual Native-Language Corpora as a Translation Resource: A Pilot Study. *Meta: Translators' Journal*, Vol. 43, No.4, 631-651.
- [2] Diakiva, R. (Ed.) (2012)/ *Entsyklopediia Innovatsii*, Kyiv, Mizhnarodna ekonomichna fundatsiia, 599.
- [3] Shtern, I.B. (1998). *Vybrani Topiky ta Leksykon Suchasnoii Lingvistyky: Entsyklopedychnyi Skovnyk dlia Fahivtsiv z Teoretychnyh Humanitarnyh Dystsyplin ta Humanitarnoi Informatyky*, Kyiv, АртЕк, 335.
- [4] European Court of Human Rights, <https://www.echr.coe.int>.
- [5] ECHR: Cases, Opinions, Matters. Ukrainian Aspect, <https://www.echr.com.ua>.
- [6] Farlex, I. (2001). *TheFreeDictionary.com*. Retrieved 09.08.2011, from Farlex. Inc.: <http://www.thefreedictionary.com>.
- [7] Lacasta, J., Nogueras-Iso, J., Zarazaga Soria, F.J. (2010). *Terminological Ontologies: Design, Management and Practical Application*, Springer Science & Business Media, 198

---

\*Corresponding author.

E-mail address: [sam141175@gmail.com](mailto:sam141175@gmail.com)