Serhii Fokin
PhD in Translation Studies
Associate Professor
Kyiv Taras Shevchenko National University
Kyiv, Ukraine
ORCID ID 0000-0003-3920-1785
sergiyborysovych@ukr.net

ELECTRONIC THESAURUS ON TRANSLATION STUDIES AS DATA MINING TOOL

The article presents the results of the experiment on organizing an electronic thesaurus of translation terms, which was prepared as a part of practical classes on translated computer lexicography and terminography. The thesaurus is supplemented by terms from the field of translation studies in Ukrainian, Spanish and Portuguese. The need to organize the thesaurus is dictated by the significant disorder of translation terminology, the lack of definitions for the terms used or significant differences between the available definitions. Thesaurus allows extracting the following paradigmatic characteristics of the term: hyperonym, hyponym, meronym, holonim, cause, effect, characterization and others. Among the syntagmatic characteristics of the term in the thesaurus adjectives and verb collocations are given. The former can shed light on the qualitative characteristics of the phenomenon / object denoted by the term; the second is able to describe the actions performed by the object or the actions affected by the object. It is stated that the semantic paradigm of a term is based mainly on its connections with other terms; partially paradigmatic relations can be derived from its syntagmatic characteristics: adjective collocations of the term naturally indicate its hyponyms. Characterization of the term according to these characteristics (in particular, hyperonym, as well as other essential seven: cause, effect, meronym, holonim, etc.) can be exploited to develop a definition of the term by intentional. New knowledge can be generated on the basis of links between thesaurus articles, in particular, due to the transitivity of most thesaurus functions, chains of thesaurus links can be automatically created; inversion search allows finding not only the article in which the original term is indicated in the lemma, but also its use in the gloss and the corresponding semantic connections with other terms.

Keywords: thesaurus; semantic paradigm; inversion search; intentional definition; translation studies.

1. INTRODUCTION

Thesaurus is meant to represent data about semantic relations among entries. Although a thesaurus can cover the common vocabulary, it turns out more reasonable to envisage relations among terms rather than among common lexemes. Since the terms are introduced by consensus being mandatorily based on the previously defined lexemes, a term system structure is much more consistent and easier to show as a relation schema than the common vocabulary is. Furthermore, a thesaurus could be seen as one of the best possible ways of illustrating terminology as a system. This circumstance is the reason why a thesaurus is an excellent tool for processing terminology. Among the well-known thesauri, it is worth mentioning the Visual thesaurus (1998–2021). As for thesauri of a narrower domain, closer by their subject to this research, some of them are compiled by the Laboratory of Computational Linguistics of Taras Shevchenko National University of Kyiv: Thesaurus of Computational Lexicography (2003–2021) and Linguistic Thesaurus (2003–2021).

As a relatively new branch of philological sciences, translation studies are characterized by two circumstances: 1) lack of term system uniformity at the international level; 2) absence of translation studies dictionaries with properly defined terms in many languages, Ukrainian including. A consequence of these factors is that terms and pseudo terms are often arbitrarily used in literature and their meaning may rest unclear in a given context. The translation thesaurus under development process is meant to bridge this gap.

Lack of encyclopedic definitions in Ukrainian for many of them requires both contextual (in this case, primarily syntagmatic) and semantically relational approaches for terms explanations. The translation terminology appears defined mostly in doctoral thesis and articles, among which there is only one explicitly dealing with translation studies terminology in Ukrainian (Zorivchak, Dzera, 2001).

2. METHODS

2.1. Thesaurus data

For the sake of performing a series of initial experiments, as well as trying to cover the lack of translatological terminological dictionaries in Ukrainian, a minimum of term entries in *.csv* format was prepared. The major part of the entries was compiled in Ukrainian, partially in Spanish and Portuguese. The algorithm, JavaScript, and HTML codes were developed by the author. The resulting thesaurus received its conventional name, "TransParent" (*Tezaurus,* 2021).

The terms were initially extracted from comparable original texts, mainly articles in the domain of translation studies in Ukrainian, Spanish and Portuguese. Although there exist some techniques that proved successful for mining equivalent terms from parallel texts, and are worth separate research, we could scarcely make use of them due to considerable semantical asymmetry in translation studies among apparently equivalent terms and also given the shortage of parallel texts in the domain of translation studies.

The described thesaurus contains data regarding the terms paradigmatic and syntagmatic relations. It also involves some knowledge inferencing functionality.

2.2. Dictionary structure

The dictionary macrostructure constitutes a relational database presented in the format of a bidimensional table. Each row of the table can be interpreted as an entry where the first column "Term" is equivalent to the lemma and the rest of the columns are given the role of the glossa. The columns cover the following functions: adjectival phrases with the term, prepositional and postpositional verbs with the term, hyperonym, hyponym, holonym, meronym, consequence, cause, definition, characterization of the term, special translation theory covering it. Since one of the prominent features of electronic dictionaries, compared to those in paper format, is the hypertextuality and richness of links among entries and their structural parts (Fokin, 2019, p. 156), we have cared about providing these links in search options. Therefore, a search of a term performed across entries makes interact different access points so that the newly created links contribute to knowledge inference at runtime, as explained in the following item.

2.3. Knowledge inference methods

Apart from representing the explicitly stated syntagmatic and paradigmatic relations of a term, which have been manually introduced into the dictionary relational database, an electronic thesaurus value should be also proved in its capability of mining more data than the explicitly introduced into its base. For the sake of providing this additional functionality, we implemented two more features:

1) reversed search of a term with its relations with other;
2) thesaurus functions chains.

2.3.1. Reversed search

Besides highlighting the relations of a term, the reversed search is also meant to unveil its usages in other entries with the relations to some other core entry term. For example, in table 1, the "transposition" is not only mentioned in its respective entry as a lemma (i.e., in the column "Term") but is also present in the glossa of the entry "transformation":

Table 1. Snippet from the database of "TransParent"

Term	Adjectives with term	Verbs with terms	Verbs subj with term	Hyperonym	Synonyms	Hyponyms
Перекладацька трансформація	перекладацька, прагматична, обов'язкова	застосувати, підлягати, проходити, застосувати	поставати	прийом, стратегія	перекладацьке перетворення	модуляція, перестановка, *транспозиція*, вилучення, додавання, асиміляція, нейтралізація, форенізація, одомашнення, калькування

2.3.2. Building thesaurus functions chains

Building thesaurus functions chains is another automatic data mining procedure implemented in the thesaurus subject to experiment. A chain of thesaurus functions implies using the same recursively applied functions to the results of itself, based on transitivity. The transitivity consists in the fact that if 'a' is related through a relation R to 'b', and 'b' is related through the same relation 'R' to 'c', then 'a' is also related through the relation R to 'c' (Transitive low, 2016). Building chains of relations among terms allows establishing relations among entries or their parts that had not been explicitly related in the manually prepared database.

3. RESULTS AND DISCUSSION

The thesaurus provides paradigmatic and syntagmatic data regarding each term existing in its database.

3.1. Term paradigmatic relations

The term "semantic paradigm" has quite a broad interpretation, e. g., "inter-words paradigm" VS "intra-words paradigm" (Kuznietsova, 1989, p. 29–30). The one definitely valid for covering thesaurus functions is the first one. Let us cursorily notice that different scholars have been assigning different names to the same phenomenon, for instance, Karaulov (1976) calls the relations of this kind "field" (p. 175).

Thus, the semantic paradigm comprises the semantic relations of the term. The most common relations are synonymy, antonymy, hyponymy, hypernymy, partonomy (meronymy) and holonymy. Some metonymic relations are also frequently involved in translation studies terminology, such as 'cause' and 'consequence.' Nevertheless, the most prominent links from the standpoint of the thesaurus explanatory power are the relations of hyponymy/hyperonymy and holonymy/meronymy. Both of them are essential for building definitions by extensions. Let us consider the following statement: ***An adequate translation*** *is a reconstruction of source text textemes and consists of an explicitation of the textual relations and functions of the source text* (Baker, 2008, p. 272).

In the example above, the term "reconstruction" is used as a hyperonym of "adequate translation" and a meronym ("explicitation"). Besides the described functions, the thesaurus includes the following parameters comprising more extended descriptions: definitions, examples of usage and characterizations. Since strict definitions are often difficult to mine from a text unless explicitly stated, characterizations may often come in handy. Characterization is a statement providing some relevant features of a term but which cannot be considered a definition due to insufficient restriction. For example: "adaptation is a procedure which can be used whenever the context referred to in the original text does not exist in the culture of the target text" (Baker, 2008, p. 3–4). This statement is valid only from the left to right since it would be not true to state vice versa: "*a procedure which can be used whenever the context referred to in the original text does not exist in the culture of the target text is an adaptation".

3.2. Term syntagmatic relations

The syntagmatic approach is based on the n-grams analysis. Since the majority of terms in the thesaurus are nouns or nominal syntagms, their typical collocations are expected to be verbs and adjectives (adjectival phrases in the vast sense of their meaning). The verbs can follow the term, designating the action performed by the entity expressed in the term, e.g., " (…) unlike conference interpreting, **courtroom interpreting requires** attention to detail, chronology and facts that may seem redundant to the interpreter" (Baker, 2008, p. 65). The term "action" in this case is used in the broadest possible sense: action, influence, causation, etc. In the case, the verb precedes the term in question, e.g., "a variant of interlingual MT is knowledge-based MT (KBMT), which **produces** semantically accurate translation (…)" (Baker, 2008, p. 163), it denotes the action the entity can be subject to (e.g., "translators resort to neutralization").

Adjectival collocations often designate subtypes of the term, e.g., "stylistic compression" or "syntactic compression." In such a way, the syntagmatic relations can implicitly provide data concerning the lexical-semantic paradigm, i.e., the semantic relations of a lexeme. The latter also appears explicitly represented in the translation thesaurus through the paradigmatic relations functions: synonyms, hyponyms, hyperonyms, holonyms, meronyms, causes and consequence, which may provide additional differential features to the definition.

Hyperonyms of a term can be useful for formulating the definitions by extension. For example, knowing that the hyperonym for adaptation is "procedure," the intensional definition for adaptation can include it as a starting point: "adaptation is a procedure which…." The differential semes for this definition are minable from the adjective and adjectival phrases (expressing types of adaptation), from its verbal collocations, etc.

3.3. Property chains implementation

As already stated, the chain of thesaurus functions is only possible to implement on transitive features: hyperonyms, hyponyms, holonyms, meronyms and synonyms to some extent.

These chains, thus, allow establishing semantic relations between some unrelated entities, for example, between "transposition" and "procedure." The hyperonym for "transposition" is "transformation." To build up a chain, we apply thus the function of hyperonym once again to the "transformation" whose hyperonym is "procedure" in this thesaurus. As a result, we obtain a chain of hyperonyms:

'transposition' is categorized by 'transformation';

'transformation' is categorized by 'procedure.'

Thus, the result of searching "transposition" will also contain this chain, where "transpositions" and "procedure" appear to hold hyper-hyponymic relations of the second level:

Транспозиція *Is categorized by in Ukrainian:* Трансформація, Перекладацька трансформація, Граматична трансформація, Морфологічна трансформація
--------**Трансформація** *Is categorized by in Ukrainian:* прийом, стратегія
--------**Стратегія** *Is categorized by in Ukrainian:* прийом

This new relation provides the possibility of enriching the knowledge base about the term by artificially generating logically true statements such as

'Transposition' is a 'procedure.'

Implementing chains of functions posed a series of technical issues, mainly that of the recursivity looping. In case the initial term appears repeated in other entries (for example, "transformation" is a synonym of "modification" and vice versa), the loop would become infinite,

which would result in the program's breakage. Saving each searched term in a separate array and checking at each step of the recursion whether this array contains the current one allows to avoid this issue.

Some properties are not transitive. Thus, these properties are not valid to build up chains of properties. That is the case, for example, of verbal collocations. At the same time, there are some relatively transitive functions. Synonyms, for instance, can indeed produce a chain of terms, whereas the entities placed too far in this chain would hardly preserve the synonymy relation. That is, however, a debatable question: in the common vocabulary, the synonymy is prone to be interpreted as similarity, while term synonyms are expected to be identical entities. Thereby, the function of synonymy could be indeed transitive in a terminological area.

As a result of data processing, this dictionary allows observing translation studies terms behavior (explicit and implicit semantic features) in their context and with regard to other terms of the same semantic field. At the interface level, the thesaurus for a given concept illustrates its relations to other term entries with the type of the relation shown in italics.

Thus, once introduced the term "транспозиція" (transposition) into the search field, the output result would be as follows (Fig. 1):

Трансформація *Is subdivided into in Ukrainian:* Модуляція, перестановка, калькування, **трансп**озиція, вилучення, додаванн, асиміляція, нейтралізація, форенізація, одомашнення.
Транспозиція *Is categorized by in Ukrainian:* Трансформація, Перекладацька трансформація, Граматична трансформація, Морфологічна трансформація.
Транспозиція *Is subdivided into in Ukrainian:* Повна **трансп**озиція, неповна транспозиція (Siniahovska, 2009).
Транспозиція *Characteristics in Ukrainian:* Транспозиція частин мови – це заміна слова вихідної мови однієї частини мови словом іншої частини мови цільової мови у процесі перекладу або набуття ним іншої морфологічної категорії (Siniahovska, 2009).
Транспозиція *Consequence in Ukrainian:* За повної **трансп**озиції, окрім переходу слова внаслідок перекладу з однієї частини мови в іншу, відбуваються також лексико-семантичні та синтаксичні трансформації відповідних мовних одиниць. (Siniahovska, 2009).
Транспозиція *Cause in Ukrainian:* Граматичні, словотвірні, лексико-семантичні, жанрово-стилістичні та ситуативно-контекстуальні причини застосування **трансп**озиції зумовлені системними відмінностями мов (Siniahovska, 2009).
Транспозиція *Definition in Ukrainian:* Транспозиція - це прийом перекладацької трансформації, що полягає в заміні однієї частини мови на іншу. (Siniahovska, 2009).
Транспозиція *Studied within special theory of translation in Ukrainian:* Лінгвістичної теорії перекладу
-----**Транспозиція** *Synonyms in Ukrainian:* Повна транспозиція
-----**Транспозиція** *Is subdivided into in Ukrainian:* Повна транспозиція, неповна транспозиція
-----**Транспозиція** *Consequence in Ukrainian:* За повної транспозиції, окрім переходу слова внаслідок перекладу з однієї частини мови в іншу, відбуваються також лексико-семантичні та синтаксичні трансформації відповідних мовних одиниць. (Siniahovska, 2009).
-----**Транспозиція** *Cause in Ukrainian:* Граматичні, словотвірні, лексико-семантичні, жанрово-стилістичні та ситуативно-контекстуальні причини застосування транспозиції зумовлені системними відмінностями мов. (Siniahovska, 2009).
-----**Транспозиція** *Studied within special theory of translation in Ukrainian:* Лінгвістичної теорії перекладу

Fig. 1. Search results for the term "transposition" in Ukrainian

The first paragraph gets included in the thesaurus thanks to the reversed search function since "transposition" does not appear in the lemma.

Although the thesaurus is multilingual, we did not provide translations for terms due to a significant asymmetry between some apparently equivalent concepts. For example, the term "addition" is sometimes treated as an error (Delisle, 1999, p. 115), whereas Neliubin (2009) focuses the respectively equivalent Russian lexeme "добавление" as amplification or a necessary procedure in translation (p. 49).

In the lack of a consistent terminological dictionary, a thesaurus could be used for mining the essential term features. Since definitions by intension predominantly rely on hyperonyms followed by some differential features of a term, these definitions are minable from a thesaurus provided that the entry has a non-empty field for hyperonyms and features (such as actions, causes,

consequences). Such a definition could be generated automatically and claims for a human review.

4. CONCLUSIONS AND SCOPE FOR FURTHER RESEARCH

Thesauri can be compiled either manually or automatically. In this proposal, we have used a hybrid approach for preparing the thesaurus of translation studies terms. The following data were manually introduced for each term entry: hyperonyms, hyponyms, holonyms, meronyms, synonyms, verbal and adjectival collocations, as well as definitions and other term's characteristics. The rest of the thesaurus data gets inferred in real time using two techniques: 1) reversing the left and the right part of the entries; 2) establishing new relations based on transitivity. Besides the described data, the former also allowed looking for the usages of the term in question in other entries and find out its relations to other entities. Thus, the ideographic data output for the user got enriched with inferred knowledge. The latter was used to establish direct logical relations among terms previously related only indirectly through other entities. The aforementioned features predetermine the following properties of the dictionary: 1) some data is not contained in the database and gets generated in real time; 2) as a result of additional data inference, a manually added entry automatically contributes more volume of data than this single entry holds; which is the scope for further research.

REFERENCES

Baker, M. (2008). *Routledge encyclopedia of translation studies*. London: Routledge.

Delisle, J. (1999). *Terminologie de la traduction / Translation Terminology / Terminología de la traducción / Terminologie der Übersetzung*. John Benjamins. [in English, German, French and Spanish]

Fokin, S. (2019). Neural network pattern for enhancing functionality of electronic dictionaries. *Advanced Education*, *6*(12), 150–158. https://doi.org/10.20535/2410-8286.132940 [in English]

Karaulov, Yu. N. (1976). *Obshchaya i russkaya ideografiya* [General and Russian ideography]. Nauka. [in Russian]

Kuznietsova, E. V. (1989*). Leksikologiya russkogo yazyka* [Lexicology of the Russian language]. Vysshaya shkola. [in Russian]

Linhvistychnyi tezaurus [Linguistic thesaurus] (2003-2021). http://www.mova.info/mov_thes.aspx?l1=68# [in Ukrainian]

Neliubin, L. L. (2009). *Tolkovyi perevodovedcheskii slovar* [Translation studies dictionary]. Nauka. [in Russian]

Siniahovska, I. Yu. (2014). Vyznachennia ta klasyfikatsiia perekladatskyh transformatssy u protsess khudozhnoho perekladu tekstu [Definition and classification of grammar transformations in the process of fiction translation]. *Scientific works. Philology. Linguistics, 209*(221), 89–93. http://irbis-nbuv.gov.ua/cgi-bin/irbis_nbuv/cgiirbis_64.exe?C21COM=2&I21DBN=UJRN&P21DBN=UJRN&IMAGE_FILE_DOWNLOAD=1&Image_file_name=PDF/Npchdufm_2014_221_209_20.pdf [in Ukrainian]

Tezaurus perekladoznavchykh terminiv [Thesaurus of translation terms]. (2021). Retrieved from http://postup.zzz.com.ua/trans-parent.html [in Ukrainian]

Tezaurus z kompiuternoi leksykohrafii [Thesaurus on computational lexicography]. (2003-2021). Retrieved from http://www.mova.info/Page3.aspx?l1=188&vocid=1 [in Ukrainian]

Transitive law. (2016). Encyclopedia Britannica. https://www.britannica.com/topic/transitive-law [in English]

Visual Thesaurus. (1998–2021). https://www.visualthesaurus.com [in English]

Zorivchak, R., & Dzera, O. (2001). Hlosariy perekladoznavchyh terminiv [Glossary of translation terms]. (2001). *Notes of the Translation Workshop, 1,* 72–74. [in Ukrainian]

Сергій Фокін. Електронний перекладознавчий тезаурус як знаряддя добування даних. У статті представлено результати експерименту з упорядкування електронного тезаурусу перекладознавчих термінів, що підготовлено у межах практичних занять з перекладної комп'ютерної лексикографії і термінографії. Тезаурус поповнюється термінами з галузі перекладознавства українською, іспанською і португальською мовами. Потреба в упорядкуванні тезаурусу продиктована значною невпорядкованістю перекладознавчої термінології, відсутністю дефініцій для використовуваних термінів або ж суттєвими розбіжностями між наявними дефініціями. Тезаурус дозволяє добувати такі парадигматичні характеристики терміна: гіперонім, гіпонім, меронім, голонім, причина, наслідок, характеризація та інші. З-поміж синтагматичних характеристик терміна у тезаурусі надаються прикметникові і дієслівні колокації. Перші можуть пролити світло на якісні характеристики явища/предмета, позначуваного терміном; другі здатні охарактеризувати дії, виконувані предметом або ж дії, впливу яких зазнає цей предмет. Зазначено, що семантична парадигма терміна ґрунтується переважно на його зв'язках з іншими термінами; частково парадигматичні відношення можуть бути виведені з його синтагматичних характеристик: прикметникові колокації терміна закономірно вказують на його гіпоніми. Характеризація терміна

за вказаними характеристиками (зокрема, гіперонім, а також інші суттєві семи: причина, наслідок, меронім, голонім та інші) можуть бути використані для розробки дефініції терміна за інтенсіоналом. На підставі зв'язків між тезарурсними статтями можуть генеруватися нові знання, зокрема, завдяки транзитивності більшості тезаурусних функцій можуть автоматично створюватися ланцюги тезаурусних зв'язків; завдяки інверсійному пошуку можна знаходити не лише статті, в яких вихідний термін зазначено у лемі, а й його вживання у глосі й відповідні семантичні зв'язки з іншими термінами.