PAWEŁ KOWALSKI

Institute of Slavic Studies, Polish Academy of Sciences, c. Warsaw

INFORMATION RETRIEVAL SYSTEM ISYBISLAW AND A SYNERGIC EFFECT: SOME REFLECTIONS ON TERMINOLOGY AND A LANGUAGE OF KEYWORDS FROM THE LINGUISTIC PERSPECTIVE

Стаття подає роздуми над уживанням мовознавчої термінології у цифровому середовищі, де система iSybislaw співіснує з іншими базами даних. Особливу увагу приділено показу того, як у системі розв'язано проблеми термінологічної глобалізації та інтернаціоналізації з метою дати користувачам iSybislaw у світовій мережі ефективний інструмент для роботи з нею, а також як термінологія, уміщена в цій системі, може слугувати об'єктом лінгвістичного дослідження. Висвітлено також деякі можливості для співпраці над термінологічним матеріалом на інституційному рівні.

Ключові слова: інформатика, інформаційно-пошукова система, мовна глобалізація, мовознавча термінологія, славістика.

0. One of the noticeable trends in the development of modern science is its dynamic digitization. The flow of scientific information, which not so long ago was held in printed form, has been largely transferred to the Internet. As a result, a variety of databases and information-retrieval systems that replace traditional bibliographies are created. Thus, we can observe changes of the bibliographic paradigm. In those databases one can observe utilization of the variety of terminological systems (terminology). An example of such a system is the information retrieval (IR) system of Slavic linguistics, which is an initiative of Polish Slavic environment linked institutionally to the Institute of Slavic Studies of the Polish Academy of Sciences. From the beginning the implementation of the project has been carried out in collaboration with linguists and specialists in the field of the IR systems and IR languages from Polish and foreign scientific institutions. The system is represented by a bibliographic database iSybislaw (see. Bibliographic database of world Slavic linguistics publications (www.isybislaw.ispan.waw.pl), which gives formal and substantial characteristics (in the form of keywords) of the recorded documents. A numerous set of linguistic documents and linguistic terminology in the function of keywords is provided by international team. The Ukrainian modul is expanded by the team from the Institute of Ukrainian Language of the National Academy of Sciences of Ukraine [6, s. 151].

0.1. Let us begin with the notion that synergy is „the combined power of a group of things when they are working together that is greater than the total power achieved by each working separately" [4]. In social psychology synergy of activities is determined by the organizational effect – a joint action that gives a greater and better results than individual one; these are activities that complement each other through cooperation and synchronization. It is the effect of organized teamwork which is higher than the sum of the effects of individual actions. As a result of synergy, an organizational effect arises, and that is the average surplus of benefits per member of

the team cooperating with other people, in comparison with the benefit that can be achieved in individual action.

0.2. Not long ago I was asked why in the iSybislaw system one can find the keyword *jezikoslovje* and cannot find the keyword *lingvistika*, although both terms are common in the Slovenian linguistic literature. My answer was – due to the rule: the iSybislaw system as keywords uses only terms from the indexed works – as always it is because we do not have documents with such a Slovenian term in the set of iSybislaw's documents. But it was somewhat misleading, not to say – untruthful. In many Slovenian linguistic works – which have got bibliographic description in the iSybislaw system – these terms were mentioned and used interchangeably, so *lingvistika* should be present in the class of equivalence with keywords 'językoznawstwo' (the meaning of this class is 'documents that refer to the notion of linguistics'. In recent years the set of bibliographic descriptions (bibliographic records and access points to knowledge) in the system has grown. The sets of keywords that reflect the current research field of modern Slavic linguistic must keep up with the increase of these records. It shows that the language of keywords must contain the important linguistic terms, regardless of different denotation – terminology (terms) denote and link elements of linguistic reality (out-of-text), while keywords denote directly the sets of the documents contained in the bibliographic system and directly link certain subsets of them, collections of document features. In fact, the multilingual – in the sense of native languages and not from the perspective of information science theory – sets of keywords in the system in the different Slavic languages might be considered as a tezauri of keywords and a base for multilingual terminological dictionary.

0.3. The paper presents a reflection on the linguistic terminology usage in the iSybislaw system. Especially it shows how the problems of terminological globalization and internationalization are solved to provide the usefull tool for the iSybislaw users worldwide, and how the terminology from the system might be a subject of linguistic research. It also sheds light on some possibilities for cooperation with terminological material and institutions. It is due to the fact that the iSybislaw system as a bibliographic and academic database (information retrieval tool for Slavic linguistics) enriches the different (versatile) cultural institutions [7].

1. Since the 1970s terminology has been associated with the English term LSP (Language for Special Purposes). The term itself was created to separate the language of science and technology from the natural language (LGP). Researchers are stressing that the LSP is a linguistic subsystem for the natural language and is closely related to the structure of the general language, but with its own special phonetic, morphological, derivational, syntactical and semantical characteristics (means) [16; 17].

1.1. Linguistic terminology as a part of LSP is of special character because it is a subject of research and the metalanguage of linguistics. The specificity lies in the fact that it can be simultaneously the object of cognition, the subject of linguistic research and research (academic) apparatus of linguistics (metalanguage of linguistics, language of description); its lexis (terms) can also be included in an artificial interlanguage [18]. In the digital environment of the iSybisaw system terminology

occurs in two functions: retrieval and metainformative: as a metalanguage for the substantial descriptions[1].

1.2. The terminology (technical and scientific/academic) is characterized by globalization processes. Especially it is noticeable in the context of European integration, because the terms related to the discovery of new objects are simultaneously translated into all languages terminologically protected [10, s. 123]. However, in the case of the academic LSP this is not a common practice. Not all linguistic terms are translated into other languages, very often they are just loans from foreign systems. In principle, you can often talk about terminological subsystems: native (national) that is formed in a national language and international, that are in similar form in different terminological systems, but they may occure with different meanings and different references (they refere to different notions). Their signifiants are similar but signiefie can differ[2]. However, for the iSybislaw's users it is not of great importance because they can retrieve information without having known the exact meaning of the notions.

1.3. The growing phenomenon of globalization of particular terminological systems is connected with strong internationalization of terminology, and is a tendency lasting from the very beginning of establishing the terms. Each language has its own terminology system related to the national tradition, word-formation preferences and normative regulations so one can not speak about unified terminological (sub) system of linguistics [13, s. 312]. However, as Zofia Rudnik-Karwatowa emphasizes, linguistics and linguistic terminology is not developed in isolation and autonomically. It has always been under the influence of different languages and different (foreign) terminological systems; moreover, linguistic terminology like the discipline itself is very often of interdisciplinary character. Within these systems one can observe the process of terminological globalization that is considered an integration of common elements. It might be noticed on many different levels of language such as phonological, morphological, and syntactical and is influenced by the tendency toward internationalization. This tendency currently is seen as one of the most basic features of the developement of national languages, affecting a number of other systemic, typological and pragmatic linguistic tendencies [see 11, s. 47; 23, s. 11]; it occurs also in the languages for special purposes (LSP). Internationalization of scientific (academic) communication may be conducive to the exchange of thought and integration of science [18, s. 172], but it may be an obstacle to the vision and linguistic description of language [21, 299-300].

1.4. The occurrence of international and native terms next to each other is one of the sources of terminological synonymy, being characteristic not only for linguistic terminology, but also for other scientific and technical terminological systems. This phenomenon might be as well observed among the synthesized units and the analytical units, and are common in Slavic linguistic terminology, for example Polish: *akronim – skrótowiec, afrykata – spółgłoska zwarto-szczelinowa, leksyka – słownictwo, lingwistyka – językoznawstwo, aksjologizacja – wartościowanie*; Slovene: *dialekt – narečje, fonetika – glasoslovje, leksika – besedje, lingvistika – jezikoslovje, morfologija – oblikoslovje, semantika – pomenoslovje, sintaksa – skladnja, tekstna lingvistika – besediloslovje; konverzija – sprevrženje,*

---

1   More on these issue see for example Rudnik-Karwatowa Z. [4].
2   On the issue of terminnological variability see Karpilovska E. [5].

palatal – mehkonebnik, dialektologija – narečjeslovje, asimilacija – prilikovanje; leksika – besedje, homonim – enakoglasnica. In terms of semantic extents, these synonymous pairs show total (*jezikoslovje – lingvistika*) or partial (aspectual) (*poenobesedenje – univerbacija*) synonymy, or even the word-formation variants *favlirati/favlati igralca, slovenizirati/sloveniti pisavo/imena, perfekcionizirati/ perfekcionirati vlogo* [23, 130 – 131].

1.3. The category of internationalizms in Slavic languages is diverse. Its scope covers such units as quotes (citations), calques as well as units with derivational motivation on the basis of specific languages, eg. derivatives formed from acquired international stems and created by native formants, very often common for Slavic language family (for example, Polish: *auktorialny, anaforyczność, modalność, predykatywność, temporalność*; Slovene: *avktorialnost, afektivnost, anaforičnost, modalnost, predikativnost, temporalnost*;), or units with international affixes (and international stems) very often entirely loaned from foreign (other) languages (Polish: *anglicyzm/anglizm, barbaryzacja, predykator, modulant, detematyzacja*; Slovene *anglizem, barbarizacija, frazeologizem*). These terms might also present total or partial internationalizms. The structure of terminological total internationalizms consists of exclusively international elements (often such units are entirely taken from the donor's language), Polish: *determinologizacja, leksykalizacja, neosemantyzm, perfektywizacja; lingwistyka korpusowa, metoda lingwistyczna, metafora konceptualna, semantyka kognitywna*; Slovene: *determinologizacija, leksykalizacija, neosemantizacija, perfektivizacija; kognitivna semantika, kognitivna sintaksa, korpusna lingvistika, leksikalna semantika;* English: *determinologization, lexicalization, neosemantization, perfectivization, corpus linguistics, linguistic method, conceptual metaphor, cognitive semantics*. Hybrid (partial) terms are those with native as well as and international elements in their structure, Polish: *perfektywność, semantyka poznawcza*; Slovene: *jezikoslovna metoda, perfektivnost, stavčna semantika*. Within the set of analytical terminological hybrids, one can also distinguish a structure in which a native element with an international element occurs (analitical hybrid) Polish: *słownik frazeologiczny, językoznawstwo kognitywne*; Slovene: *denpendenčna slovnica, frazeološki slovar* or only internationalize elements, Polish: *gramatyka funkcjonalna, lingwistyka kognitywna*; Slovene: *korpusna lingvistika*. From the term-formation perspective for such units (synthesized and analitical) one can observe differences in the variety (kinds, type) of linguistic mechanisms. For the synthesized units there are traditional word-formation (morphological) mechanisms such as for example derivation, blending, compounding, and for analitycal units there are derivational mechanisms on lexical (and morphological-lexical) level succesfully applied, for instance, in the phraseology-formation (phraseological derivation).

1.4. The internationalization and globalization of terminology cause changes in the term-formation tendencies. There is – just as in general lexis – a growing number of analytical and agglutination tendencies in the creation of new terminological units. In the derivation of nouns the activity of foreign elements – having the character of a prefix or prefixoid – is increasing. The word-formation nests are created with such international elements as Polish: *anti-, de-, post-, pseudo-, re-, meta-, micro-, multi-, neo-*; Slovene: *anti-, de-, kvazi, post-, pseudo-, re-, meta-, micro-, multi-*, Polish: *antyteza, deadiektywizacja, deapelatywizacja, deleksykalizacja, determinologizacja,*

desemantyzacja, depalatalizacja, makrotoponimia, metajęzyk, metatekst, metaterminologia mikrotoponimia, multilingwalny, neosemantyzm, paleosemantyzm, pseudo-synonim, pseudoanglicyzm, reduplikacja; Slovene: *antiopozicijskost, deleksikalizacija, determinologizacija, desemantizacija, metabesedilo, metajezik, multilingavlen, psevdoanglizem, redefinicija* [see for example 19]. It also reflected on the level of internal syntax of the terms. The structure of analytical units in the linguistic terminology is dominated by terms with two elements that have the form of a noun-adjective syntagma (N Adj) in Polish and in Slovene adjective-noun phrase (Adj N), Polish: *słownik bilingwalny, derywat odczasownikowy*; Slovene: *etimološki slovar, determinativno zlaganje, eksperimentalna fonetika*. There are much less frequent units with the noun-noun model (NN). Polish: *aktualizacja metafory, akwizycja języka, przełączanie kodów, studia płci*. They often constitute hidden internationalisms (international structural calque), English: *language acquisition, metaphor actualization, code switching*. Perhaps some of these types of units can be interpreted as a transitional stage in the process of borrowing an international terminological unit, see eg. *akwizycja językowa, przełączanie kodów*.

1.5. In the synthesized terminological units – what has been previously well acknowledged in the linguistic literature – productivity of Greek and Latin elements (affixes, affixoids and elements of compound units) significantly increases in linguistic terminology in the middle of the 20th century. However, such affixes were present in the linguistic systems much earlier. For example, the description of Polish nouns excerpted from the „Dictionary of keywords in Slavic linguistics", in which linguistic terms in the field of Slavic Studies are presented, showed that derivatives containing international suffixes *-acja, -izacja/-yzacja, -izm/-yzm* constitute 59% of the whole set of terms (including derivatives with *-acja* - 39%) [1]. The number of Slovenian formations with these suffixes (*-acija, -izacija, -izem*) in the „Encyclopedia of the Slovenian language" („Enciklopedija slovenskega jezika") is similar [8]. These international suffixes occur in most languages of the Euro-American dictionary community and create an international terminological system: Russian: *артикуляция, гглютинация, делексикализация*; French: *articulation, agglutination, delexicalisation, lexicalisation, morphologisation, accentuation*; English: *delexication, articulation, agglutination, delexication, lexicalization, morphologization, accentuation*; German: *Artikulation, Agglutination, Lexikalisation, Morphologisation, Akzentuation* [18; 20, 325-326].

In the analytical structures at the lexical level the frequent terminological derivation technique in the globalisation perspective are primarily commutation, that is the exchange of one of the constituent unit of the term, for example, Polish: *korpus dwujęzyczny – korpus bilingwalny, leksykografia dwujęzyczna – leksykografia bilingwalna, lingwistyka poznawcza – lingwistyka kognitywna – językoznawstwo kognitywne, semantyka poznawcza – semantyka kognitywna, metafora pojęciowa – metafora konceptualna, metoda lingwistyczna – metoda językoznawcza; rzeczownik odprzymiotnikowy – rzeczownik odadiektywny, rzeczownik odczasownikowy – rzeczownik dewerbalny, alternacja samogłoskowa – oboczność samogłoskowa, amalgamat konceptualny – amalgamat pojęciowy, dopełniacz adnominalny – dopełniacz odrzeczownikowy*; Slovene: *dvojezični korpus – bilingvalen korpus, kognitivno jezikoslovje – kognitivna lingvistika, deliberativno vprašanje – razmišljalno vprašanje, dependenčna slovnica – odvisnostna slovnica, deverbativni*

samostalnik – izglagolski samostalnik, leksikalna enota – besedna enota, leksikalni zaklad – besedni zaklad. All these terms are useful in establishing classes of equivalence, because they allow to create more access points in the system without any threats to obtain relevant information for the users. At this moment in the iSybislaw in many classes of equivalence one can observe the absense of such formations.

2. In the proces of designing and working on the basic information tool for the iSybislaw system, that is the IR language of keywords such pairs (intrinsic synonymy – international and native terms, and cross-lingual synonymy – international terms) are connected in one class with multi elements [see for example 12]. On the external perspective these classes are multilingual and on the internal perspective they present one level of the information retrieval language. Taking into account the iSybislaw system for multilingual keywords in the structure of the created IR language and in the classes of equivalents there is a need to consider the broad terminological material – native and international – otherwise there is no possibility to provide relevant information and a relevant retrieval tool for the users worldwide. Moreover, a number of equal expressions in Slavic languages and English must be combined and easily switched from one language to another. For the benefit and usefulness of the searching processes (retrieval processes) the equivalence classes should contain all the terms from the system, regardless of presence or absence in the documents. This allows to provide a complete and relevant information retrieval of not only all documents included in the system, but also about the field of Slavic linguistics. This will result in the multilingual terminological dictionary of Slavic linguistic terminology and a keyword dictionary of a relevant topic of Slavic linguistics as an integral part of the iSybislaw system. For example for the Ukrainian modul it is possible to use already an existing tool „для підвищення ефективності пошуку в ІПС iSybislaw можуть бути корисними напрацювання у використанні тезаурусного підходу до впорядкування й систематизації лінгвістичної термінології українських лінгвістів – «Електронний словник лінгвістичної термінології з інформаційно-пошуковою системою (тезаурус)», удоступнений на лінгвістичному порталі лабораторії комп'ютерної лінгвістики Інституту філології Київського національного університету імені Тараса Шевченка[3]. Тезаурус містить 3400 термінів із морфології, синтаксису, лексики, семантики, комп'ютерної лінгвістики, поданих українською, російською й англійською мовами. ІПС може функціонувати в складі іншої інтелектуальної системи, наприклад через систему відсилань [9, 184].

The international team – Belarusian, Bulgarian, Croatian, Czech, Polish, Russian, Slovak, Serbian, Ukrainian and German responsible for Upper and Lower Sorbian – consists of competent specialists in the different fields of linguistics is capable to provide such a tool. In this light the work with the system creates a kind of synergistic effect – the work of linguists, terminologists, specialists in scientific information and IT – results in an effect greater than could be achieved by these individuals separately. It is an ongoing project and one can observe cooperation of tasks that allows both to acknowledge the terminological material from the linguistic perspective and increase and optimise functionality of the system from the information science perspective.

---

[3]   www.mova.info

The keywords might be used by external institutions for instance publishing offices to create a list of keywords to help the author of linguistic papers in creating the abstract and the set of keywords. It shows that bibliographic database nowadays not only inherited functions of traditional paper bibliographies but provide a lot more functionality and are the bibliography of the 21st century.

1. *Banasiak J.* (2017), Procesy nominalizacyjne w tekstach językoznawczych na materiale języka polskiego i bułgarskiego. In. Rudnik-Karwatowa Z. (Red.), Dynamika współczesnego słownictwa słowiańskiego w przestrzeni stylowo-funkcjonalnej. Warszawa: Instytut Slawistyki PAN, S. 157 – 176.
2. *Banasiak J.* (2014), Synonymy and search synonymy in an IR system (on the basis of linguistic terminology and the iSybislaw system). „Studia z Filologii Polskiej i Słowiańskiej". nr 49, S. 176 – 187; https://ispan.waw.pl/journals/index.php/sfps/article/view/sfps.2014.017/543.
3. *Buttler D.* (1979), O wzajemnym oddziaływaniu terminologii i słownictwa ogólnego. „Poradnik Językowy", 2, S. 58 – 66.
4. Cambridge Dictionary online, dostęp: www.dictionary.cambridge.org
5. *Karpilovska E.* (2016), *Варіанти термінів мовознавства в сучасній науково-освітній практиці і в часовій перспективі*. In: Rudnik-Karwatowa (Red.), Dynamika współczesnego słownictwa słowiańskiego w przestrzeni stylowo-funkcjonalnej. Warszawa: Instytut Slawistyki PAN, S. 9 – 26.
6. *Karpilovska E.* (2018), *Перспективи розвитку українського модуля в складі інформаційно-пошукової системи iSybislaw.* Kowalski P. (Red.). „Z zagadnień informacji naukowej, terminoznawstwa i językoznawstwa". Warszawa: Instytut Slawistyki PAN, S. 149 – 164.
7. *Kowalski P., Fastyn M., Banasiak J.* (2018), Wykorzystanie zasobów bibliograficznych przez instytucje kultury czy ich integracja (na przykładzie systemu iSybislaw). „Archiwa – Kancelarie – Zbiory". R.11, nr 9, S. 96 – 108.
8. *Kowalski P.* (2016), Synonimia w polskiej i słoweńskiej terminologii językoznawczej. In: Toporišičeva Obdobja. Kržišnik E., Hladnik M. 9Red.), Ljubljana: Filozofska Fakulteta, S. 99 – 106.
9. *Kysliuk L.* (2018), Розбудова парадигматики лінгвистичної термінології в інформаційно-пошуковій системі iSybislaw: погляд користувача. In: Kowalski (Red.), Z zagadnień informacji naukowej, terminoznawstwa i językoznawstwa. Warszawa: Instytut Slawistyki PAN, S. 181 – 190.
10. *Mazurkiewicz-Sułkowska J.* (2014), Słowiańska terminologia techniczna (na materiale polskim, rosyjskim i bułgarskim). Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
11. *Ohnheiser I.* (2003), Komparacja współczesnych języków słowiańskich. Słowotwórstwo, nominacja. Opole: Wydawnictwo Uniwersytetu Opolskiego.
12. *Romaniuk J.* (2018), Критерії добору ключового слова позиця автора vs. користувача. In: Kowalski P. (Red.), Z zagadnień informacji naukowej, terminoznawstwa i językoznawstwa. Warszawa: Instytut Slawistyki PAN, S. 263 – 274.
13. *Rudnik-Karwatowa Z.* (2012), Rola słowotwórstwa w słowiańskiej terminologii językoznawczej (na materiale języka polskiego, czeskiego i rosyjskiego). In: Dragićević R. (Red.), Творба речи и њени ресурси у словенским језицима : зборник радова са четрнаесте међународне научне конференције Комисије за творбу речи при Међународном комитету слависта. Београд: Филолошки факултет, S. 311 – 318.
14. *Rudnik-Karwatowa Z.* (2016), Упорядкування термінології для потреб інформаційно-пошукової системи iSybislaw. *„ Українська мова"*, nr 3, S. 105 – 117.
15. *Sager J.C.* (1990), A practical course in terminology processing. Amsterdam: John Benjamins.

16. *Sorokina E.* (2018), Особенности проявления синонимии в лингвистической терминологии. In. Kowalski P. (Red.), Z zagadnień informacji naukowej. Warszawa: Instytut Slawistyki PAN, S. 305 – 320.
17. *Sorokina E.* (2017), Формирование теории языка для специальных целей (ясц). In Piper P., Jovanović V. (Red.), Slovenska terminologija danas. Beograd: Institut za srpski jezik SANU, S. 29 – 40.
18. *Sitarski J.* (1991), Internacjonalizacja w zakresie polskiej i rosyjskiej terminologii językoznawczej. „Studia Rossica Posnaniensia”, R. 22, S. 171 – 177.
19. *Stramljič Breznik I.* (2009), Hibridizacija novejših slovenskih tvorjenk. In: Przejawy internacjonalizacji w językach słowiańskich. Ed. E. Koriakowcewa. Siedlce: Wydawnictwo Akademii Podlaskiej, S. 165 – 178.
20. *Taran A.* (2018), Динамика мови в джерелах і ключових словах інформаційно пошукової системи iSybislaw. In: Kowalski (Red.), Z zagadnień informacji naukowej, terminoznawstwa i językoznawstwa. Warszawa: Instytut Slawistyki PAN, S. 321 – 333.
21. *Topolińska Z.* (2010), Patrząc przez cudze okulary (infinitivus ~ subiunctivus). „Slavia Meridionalis”, nr 10, S. 299 – 305.
22. *Waszakowa K.* (2009) Internacjonalizacja polskiej leksyki – stan obecny, prognozy na najbliższą przyszłość. In: Koriakowcewa E. (Red.), Przejawy internacjonalizacji języków słowiańskich. Siedlce: Wydawnictwo Akademii Podlaskiej 2009, S. 11 – 28.
23. *Žele A.* (2009): Pomenotvorne zmožnosti z vidika /de/terminologizacije (v slovenščini). In: Ledinek N., Žagar Karer M., Humar M. (Red.), Terminologija in sodobna terminografija. Ljubljana: Založba ZRC, S. 125 – 139.

Paweł Kowalski

INFORMATION RETRIEVAL SYSTEM ISYBISLAW AND A SYNERGIC EFFECT:
SOME REFLECTIONS ON TERMINOLOGY AND A LANGUAGE OF KEYWORDS
FROM THE LINGUISTIC PERSPECTIVE

Digitization of academic discourse causes changes that influence the work of researchers and flow of information in the scientific community. One can observe strong globalization and internationalization processes in the current terminological system wordlwide. The need to index the emerging works in the bibliographic databases, which have the goal to provide the necessary information for their users and to preserve the reflection of science and culture for future generations, does not disappear. This means that indexers are forced to skillfully maneuver between the quality and the amount of information provided. The paper presents some thougts on including and combining the international and native terms in the classes of equivalence in the iSybislaw system. To provide a complete and relevant information retrieval they both should be considered as an accesses point to the knowledge offered by the system.

K e y w o r d s : information science, information retrieval systems, linguistic globalization, linguistic terminology, Slavic studies.