

DOI: 10.11649/sfps.2014.006

Zofia Rudnik-Karwatowa  
(Instytut Sławistyki PAN, Warszawa)

## **Wielojęzyczny system informacyjno-wyszukiwawczy językoznawstwa sławistycznego iSybisław: wyzwania, osiągnięcia i możliwości**

1. Powstanie Internetu i możliwości wykorzystania w nim technik przetwarzania informacji dały impuls do działań innowacyjnych w różnych dyscyplinach naukowych. Efektem takich twórczych działań w językoznawstwie sławistycznym było opracowanie koncepcji nowoczesnego systemu informacji sławistycznej (Bojar & Rudnik-Karwatowa, 1998, ss. 42–47). W wyniku kilkuletnich prac stworzony został zgodnie z podstawowymi założeniami metodologicznymi system informacyjno-wyszukiwawczy iSybisław, prezentowany przez bibliograficzną bazę danych (Bibliograficzna baza danych światowego językoznawstwa sławistycznego iSybisław, b.d.). W bazie iSybisław zapewniono otwarty dostęp do bibliograficznych metadanych, a także – na razie w nielicznych przypadkach – do pełnych tekstów dokumentów.

System iSybisław ma charakter specjalistyczny: udostępnia szczegółowe informacje o polskim i światowym piśmiennictwie z zakresu językoznawstwa sławistycznego, w tym językoznawstwa konfrontatywnego i językoznawstwa

This is an Open Access article distributed under the terms of the Creative Commons Attribution 3.0 PL License ([creativecommons.org/licenses/by/3.0/pl/](http://creativecommons.org/licenses/by/3.0/pl/)), which permits redistribution, commercial and non-commercial, provided that the article is properly cited. © The Author(s) 2014.

Publisher: Institute of Slavic Studies, PAS & The Slavic Foundation  
[Wydawca: Instytut Sławistyki PAN & Fundacja Sławistyczna]

kontrastywnego słowiańsko-niesłowiańskiego, od 1992 r. po dzień dzisiejszy. Do zbioru informacyjnego systemu wchodzi również prace językoznawcze slawistów dotyczące problemów teoretycznych, ważne dla slawistyki językoznawczej. Tak wyznaczony zbiór informacyjny jest systemem informacji kierującej użytkowników do dokumentów zawierających relewantne dla nich informacje. Dążenie do precyzji odwzorowywania treści dokumentów w charakterystykach wyszukiwawczych dokumentów pochodnych realizowane jest w mających odpowiedni stopień szczegółowości słowach kluczowych. Stanowią one główne narzędzie indeksacyjne i wyszukiwawcze w systemie. Dodatkowo do wyrażenia treści dokumentów stosuje się także klasyfikację oraz – od niedawna – swobodne słowa kluczowe, zwane tagami.

2. Użytkownikami systemu informacyjno-wyszukiwawczego językoznawstwa slawistycznego iSybislaw są przede wszystkim specjaliści z tej dziedziny wiedzy: pracownicy naukowcy, doktoranci, studenci slawistyki i polonistyki oraz – w znacznie mniejszym zakresie – badacze języków niesłowiańskich, historycy, a także literaturoznawcy. System informacji slawistycznej powinien więc zaspokajać relewantne potrzeby informacyjne tej grupy użytkowników. Osiągnięcie tego celu będzie bardzo trudne z uwagi na wielojęzyczność zarówno zbioru informacyjnego systemu, jak i użytkowników oraz indeksatorów. Zapewniliśmy dotychczas użytkownikom komunikowanie się z systemem i korzystanie z niego za pomocą wielojęzycznego interfejsu – w języku polskim, rosyjskim i angielskim – zastosowaliśmy także rozbudowane opcje wyszukiwawcze w systemie<sup>1</sup>. Przed nami stoi pionierskie zadanie: stworzenie docelowo, zgodnie z projektem, systemu umożliwiającego jednoczesne wyszukiwanie informacji w dowolnym języku słowiańskim i angielskim. Zbudowanie takiego systemu, zapewniającego relewantne, dokładne i kompletne informacje, stanowi ogromne **wyzwanie** dla międzynarodowego zespołu slawistów zaangażowanych w to przedsięwzięcie<sup>2</sup>.

---

<sup>1</sup> Ze szczegółowymi informacjami na temat funkcjonowania systemu można zapoznać się w pracy Zenona Mikosa zamieszczonej w tym tomie czasopisma „Studia z Filologii Polskiej i Słowiańskiej” (Mikos, 2014).

<sup>2</sup> Projekt jest związany instytucjonalnie z Instytutem Slawistyki PAN (zainicjowano tu prace metodologiczne nad nowoczesnym systemem informacji dokumentacyjnej; IS PAN jest też koordynatorem prowadzonych w sposób ciągły prac nad informacją bibliograficzną językoznawstwa slawistycznego) i realizowany przez zespół współpracowników z renomowanych polskich i zagranicznych ośrodków slawistycznych: z Uniwersytetu Wrocławskiego,

**3.1.** Zrealizowanie wyznaczonych w projekcie zadań wymaga rozwiązania szeregu problemów metodologicznych. Problem zasadniczy wiąże się z zapewnieniem efektywności podstawowego narzędzia informacyjnego systemu iSybislaw, jakim jest język informacyjno-wyszukiawczy: **język słów kluczowych**. Zastosowanie w systemie iSybislaw języka słów kluczowych mającego charakter **paranaturalny**, a nie sztuczny, umożliwia użytkownikom, mającym kompetencję językową w zakresie języków naturalnych, samodzielne i wygodne wyszukiwanie oraz samodzielną ocenę relewancji wyszukanej informacji, indeksatorom zaś ułatwia opracowanie informacji dla systemu. Ten walor języka paranaturalnego był wyraźnie podkreślony w koncepcji systemu informacji sławistycznej (Bojar & Rudnik-Karwatowa, 1998, s. 45).

Słowa kluczowe to jednostki leksykalne paranaturalnego języka informacyjno-wyszukiawczego (języka słów kluczowych). Odnoszą się bezpośrednio do rzeczywistości dokumentacyjnej, odwzorowując elementy treści dokumentów – ich znaczenie odpowiada równokształtnym z nimi wyrażeniom języka naturalnego w funkcji metainformacyjnej. Innymi słowy: „w procesach informacyjnych słowa kluczowe to wyrazy (wyrażenia) języka naturalnego, użyte do opisu tekstów dokumentów w celu późniejszego ich wyszukiwania” (Babik, 2010, s. 77).

**3.2.** W dziedzinowych językach słów kluczowych istnieje tendencja do opierania się wyłącznie na systemie terminologicznym danej dziedziny. W systemie informacji sławistycznej nie zastosowano takiego ograniczenia. Zakres tematyczny systemu iSybislaw wyznacza najliczniejszą reprezentację słów kluczowych bazujących na terminach językoznawstwa sławistycznego, a także teorii języka i językoznawstwa ogólnego. Słowa kluczowe opierają się również, choć w mniejszym zakresie, na innych wyrażeniach języka naturalnego, odwzorowujących elementy treści dokumentów, które wyznaczają pola tematyczne ważne dla organizacji i wyszukania informacji o zbiorze reprezentatywnych dokumentów z językoznawstwa sławistycznego. Do

---

Uniwersytetu im. Adama Mickiewicza w Poznaniu, Uniwersytetu Marii Curie-Skłodowskiej w Lublinie, Instytutu Języka Czeskiego ANRCz w Pradze, Instytutu Słowianoznawstwa RAN w Moskwie, Państwowego Uniwersytetu Moskiewskiego im. Łomonosowa, Instytutu Języka Ukraińskiego NANU w Kijowie, Centrum Badań Kultury, Języka i Literatury Białoruskiej NANB w Mińsku, Instytutu Łużyckiego w Budziszynie, Centrum Badawczego Lingwistyki Arealnej ANiS w Skopju i innych. Dużą część zespołu stanowią członkowie Komisji Bibliografii Lingwistycznej afiliowanej przy Międzynarodowym Komitecie Sławistów.

takich wyrażen należą: nazwy osobowe, geograficzne i korporatywne oraz nazwy pospolite, np. nazwiska językoznawców, których dorobek naukowy jest przedmiotem badań relacjonowanych w dokumentach, miejsca i rok odbycia się konferencji będących przedmiotem omówienia w dokumentach, nazwy towarzystw naukowych, tytuły dokumentów itp. oraz wyrażenia języka ogólnego stanowiące przedmiot badań, jak np. leksemy, frazeologizmy, morfemy itp., wyrażenia mające dużą frekwencję w zbiorze dokumentów, np. w języku polskim *nazwy barw*, *nazwy części ciała*, nazwy zgromadzeń naukowych: *kongres*, *konferencja* i inne.

Zastosowanie w funkcji słów kluczowych, reprezentujących treść dokumentów, oprócz terminów językoznawczych także niektórych wyrażen języka ogólnego, wchodzących w zakres tematyczny systemu informacyjno-wyszukiawczego, pozwala w sposób adekwatny opisać wszystkie istotne elementy treści każdego dokumentu oraz poszerzyć możliwości wyszukiawcze. Dzięki przyjętemu rozwiązaniu metodologicznemu powiększa się możliwość wieloaspektowego, fasetowego oglądu treści dokumentów, zaspokajającego potrzeby informacyjne zarówno językoznawców, jak i użytkowników reprezentujących inne dziedziny wiedzy, a także niespecjalistów, np. studentów i uczniów.

**3.3.** Formą organizacji i prezentacji leksyki języka słów kluczowych oraz relacji między wyrażeniami tego języka i wyrażeniami języka naturalnego w funkcji metainformacyjnej jest *Słownik słów kluczowych językoznawstwa sławistycznego* oparty na terminologii językoznawczej w języku polskim (Rudnik-Karwatowa & Karpińska, 1999; 2006; wersja demo [http://ispan.waw.pl/cnis/slownik\\_slow\\_kluczowych.htm](http://ispan.waw.pl/cnis/slownik_slow_kluczowych.htm)). W wersji elektronicznej tego słownika z 2006 r. zaktualizowano słownictwo przedstawione w wydaniu książkowym. Uwzględnione zostały nieujęte w pierwszym wydaniu słownika terminy już stosowane, lecz utrwalone i rozpowszechnione w ostatnim okresie albo niewystępujące w dotychczas badanym korpusie tekstów reprezentatywnych dla współczesnego językoznawstwa sławistycznego<sup>3</sup>.

Kontynuowana jest systematyczna aktualizacja słownictwa, niezbędna dla podniesienia jakości systemu informacyjno-wyszukiawczego. Do rozbudowywanego systemu leksykalnego języka słów kluczowych weszły terminy nieuwzględnione w polskim słowniku: pol. *chrematonim*, *chrematonimia* 1

---

<sup>3</sup> O aktualizacji słownictwa w słowniku elektronicznym zob. w pracy: Rudnik-Karwatowa, Mikos & Bojar, 2007, s. 24–25.

'zbiór chrematonimów' i *chrematonimia* 2 'dyscyplina', *frazematyka*, *kaszubistyka*, *pragmalingwistyka*, *frazematyka* i inne, a także terminy z innych języków słowiańskich, np. cz. *genderová lingvistika*, *pedolingvistika*, ros. *когнитивная лингвистика*, *лингвокультурология* i wiele innych. Znaczną grupę wśród nich stanowią terminy nazywające subdyscypliny językoznawcze i teorie naukowe.

Bardzo ważnym zadaniem dla zespołu współpracowników systemu iSybil-sław jest opracowanie słowników prezentujących zbiory podstawowej leksyki języka słów kluczowych w każdym języku słowiańskim oraz porządkujących te zasoby (przede wszystkim rozwiązujących problem synonimiczności i wieloznaczności jednostek leksykalnych). Będą one stanowić pomocne narzędzie indeksowania na odpowiednim i pożądanym poziomie szczegółowości oraz pomoc dla użytkowników w uzyskaniu kompletnej i trafnej informacji.

**3.4.** System informacji specjalistycznej powinien uwzględniać informacje relewantne dla użytkowników, konieczne jest zatem zapewnienie odpowiedniej szerokości i głębokości indeksowania. Inaczej mówiąc, informacje zawarte w dokumencie powinny być odwzorowane w charakterystykach wyszukiawczych w sposób wyczerpujący i na odpowiednim stopniu szczegółowości. Dążymy do tego, aby relewancję i kompletność informacji zapewnił system stosujący jako podstawowe narzędzie informacyjne język słów kluczowych. Słowa kluczowe, charakteryzujące się elastycznością i otwartością<sup>4</sup>, umożliwiają bowiem odzwierciedlenie w procesie indeksowania wieloaspektowej i istotnej treści dokumentu oraz wyszukanie informacji relewantnej, szczegółowej i kompletnej. Informacji takiej nie zapewniają inne języki informacyjne, np. język klasyfikacyjny, który ze względu na płytką klasyfikację rzeczową jest stosowany w naszym systemie tylko pomocniczo.

Jakość informacji wprowadzanej do systemu zależy w dużym stopniu od kompetencji (wiedzy) językoznawczej indeksatorów. Doświadczeni badacze z polskich i zagranicznych ośrodków slawistycznych zapewniają przygotowanie informacji relewantnej dla użytkowników.

**3.5.** Stosowanie języka naturalnego, głównie terminologii, w funkcji słów kluczowych obok wskazanych wyżej korzyści dla użytkownika stwarza także szereg problemów niepożądanych w systemie informacyjnym. Jak wiadomo,

---

<sup>4</sup> Zalety języka słów kluczowych szczegółowo omówił Wiesław Babik (Babik, 2010, ss. 112–113, 213).

mimo postulowanej przez większość językoznawców precyzji terminów, czyli ścisłego wyznaczania ich zakresu i jasnej treści, w terminologii językoznawczej występuje **synonimia** i wieloznaczność.

W poszczególnych językach słowiańskich funkcjonują pary terminów mających identyczne lub bliskie znaczenie, a także szeregi tworzone przez więcej niż dwa równoznaczne lub bliskoznaczne terminy. Zjawisko to wiąże się z łacińsko-grecką tradycją terminologii językoznawstwa europejskiego, z internacjonalizacją terminologii, a także z wyraźnie wzmożoną od lat 80. XX wieku aktywnością różnych szkół i kierunków, tworzących odpowiedniki terminów istniejących już w pracach językoznawczych, por. pary i szeregi synonimiczne w polskiej literaturze językoznawczej: *verbum* – *czasownik*, *spółgłoska szczelinowa* – *spirant* – *spółgłoska trąca* – *spółgłoska frykatywina*, *spółgłoska zwarto-szczelinowa* – *spółgłoska zwarto-trąca* – *afrykata*, *tekstem* – *model tekstu* – *wzorzec tekstowy* – *typ tekstu* i w języku czeskim: *akuzativ* – *4. pád*, *singulár* – *číslo jednotné*, *frikativa* – *spiranta* – *třená souhláska*, *derivát* – *odvozenina* – *odvozené slovo*, *kompozitum* – *složenina* – *složené slovo* i inne. Zuzanna Topolińska przywołuje serię terminów odnoszących się do kategorii imperceptiwu (w starszej polskiej tradycji terminologicznej jest to tzw. *strona nieświadka*) obecnej w języku tureckim i w językach ligi bałkańskiej, w tym w słowiańskich językach Bałkanów: *dubitativus*, *(non)confirmativus*, *inferentialis*, *evidentialis*, *admirativus*, *mediativus*, *prekažanost* i inne. Terminy te kształtowały się w warunkach lokalnych tradycji gramatycznych na różnych etapach rozwoju, zmieniając się tylko w sferze konotacyjnej, a nie denotacyjnej (Topolińska, 2010, s. 300).

Wśród powyższych przykładów wyodrębnić można dwie grupy synonimów: terminy równoznaczne (synonimy absolutne, dublety), np. *verbum* – *czasownik*, *akuzativ* – *4. pád*, *singulár* – *číslo jednotné* itp., oraz terminy aspektowe, oznaczające różne aspekty tych samych zjawisk, procesów i obiektów rzeczywistości językowej, np. ze względu na przypisanie funkcji formantu słowotwórczego paradygmatowi fleksyjnemu lub brakowi sufiksu, tj. sufiksowi zerowemu: pol. *derywacja paradygmatyczna* – *derywacja zerowa*, *derywacja bezsufiksalna*. W drugiej grupie sytuują się wspomniane wyżej terminy reprezentujące kategorię imperceptiwu w językach południowosłowiańskich na Bałkanach.

**3.6.** Dzięki specjalnym procedurom zastosowanym w procesie przygotowania metainformacji w charakterystykach wyszukiwawczych dokumentów system iSybisław zapobiega niewyszukaniu informacji wyrażonej synonimicznymi terminami, a więc stracie informacji. W języku słów kluczowych nie eliminujemy

relacji synonimiczności między wyrażeniami leksykalnymi, lecz ją wskazujemy i wykorzystujemy do efektywnego wyszukiwania informacji. Przyjmując rozwiązanie polegające na ustanowieniu **klas ekwiwalentów**, w które wchodzi pary lub ciągi (szeregi) synonimiczne terminów klasycznych i aspektowych, jednocześnie zapewniamy użytkownikom dotarcie do tej samej informacji według dowolnego wyrażenia z danej klasy. Użytkownicy mają możliwość korzystania z dowolnego ekwiwalentu jednostek językowych zapisanych w systemie i mogą uzyskać w ten sposób kompletną informację, zawartą we wszystkich różnojęzycznych dokumentach i zapisanych w różnych alfabetach. Nie muszą wykazywać się kompetencją językową, aby zgodnie z potrzebami informacyjnymi podanymi w zapytaniu zostać skierowani przez system do wszystkich dokumentów zawierających potrzebną informację. Na przykład informację o derywacji paradygmatycznej (por. p. 3.5.) uzyskają, posługując się jakimkolwiek wyrażeniem całej klasy ekwiwalentów z różnych języków naturalnych/etnicznych: pol. *derywacja paradygmatyczna*, *derywacja fleksyjna*, *derywacja zerowa*, *derywacja bezsufiksalna*, ros. *нулевая деривация*, *нулевая аффиксация*, *нулевая суффиксация*, ukr. *безафіксне словотворення*, *нульсуфіксація*, *нульова суфіксація*, *нульсуфіксальний спосіб словотворення*<sup>5</sup> i in.

Łączenie ekwiwalentów językowych wyrażeń wszystkich języków słowiańskich w klasy (ciągi) ekwiwalentów umożliwi równorzędne traktowanie w wyszukiwaniu informacji nie tylko terminów o równorzędnym zakresie, lecz także innych, różnych jednostek językowych, ekwiwalentnych w funkcji denotacyjnej. Na przykład użytkownik będzie mógł dotrzeć do wszystkich dokumentów zawierających informację o autorze, którego imię i nazwisko jest zapisane w zbiorze wyszukiwawczym w różnych językach i alfabetach oraz w różnych wariantach w tym samym alfabecie (por. w systemie iSybislaw: *Shevelov George Y.*, *Szerech Jurij*, *Szewelow Jurij*, *Шевелев Юрий Владимирович*, *Шевельов Юрій*), jeśli zastosuje dowolne wyrażenie z klasy ekwiwalentów. Przyjęte rozwiązanie metodologiczne ma istotne znaczenie dla osiągnięcia efektywności wielojęzycznego systemu informacji sławistycznej.

Należy zwrócić uwagę na to, że w językach informacyjno-wyszukiwawczych synonimia może być traktowana szerzej niż w języku naturalnym. Za synonimy uważane są nie tylko „klasyczne” wyrażenia równoznaczne, lecz

---

<sup>5</sup> Prof. Jevhenii Karpilovskiej serdecznie dziękuję za podane przykłady terminów z języka ukraińskiego.

także wyrażenia, które w języku naturalnym nie są synonimami (Bojar, 2005). W przypadku języka słów kluczowych językoznawstwa sławistycznego relacja synonimii wyszukiwawczej występowałyby więc nie tylko między klasycznymi synonimami, jak np. w języku polskim: *językoznawstwo* – *lingwistyka*, *język polski* – *polszczyzna*, w czeskim: *derivát* – *odvozenina* – *odvozené slovo*, lecz także między takimi wyrażeniami, jak np. w języku polskim: *określoność* – *nieokreśloność*, *derywacja* – *kompozycja*, *dokonaność* – *niedokonaność*, *żywołność* – *nieżywołność*. Na obecnym etapie prac nad językiem słów kluczowych systemu iSybislaw nie stosujemy synonimii wyszukiwawczej w powyższej postaci. Wyszukiwanie uwzględnia reprezentowanie jednego pojęcia przez wiele ekwiwalentnych semantycznie wyrażen, zarówno klasycznych (absolutnych) synonimów (*językoznawstwo*, *lingwistyka* i ich ekwiwalenty innosłowiańskie), jak i synonimów aspektowych (pol. *predykatyw*, *czasownik nieosobowy*, *czasownik niefleksyjny*, *czasownik niewłaściwy*, *kategoria stanu* (rzadko), ros. *предикатив*, *безлично-предикативное слово*, *предикативное слово*, *предикативное наречие*, *безличное наречие*, *категория состояния*, zob. także pol. *derywacja paradigmatyczna*, ros. *нулевая деривация*, *нулевая аффиксация*, *нулевая суффиксация* i in.). Wartość zastosowanej metody jest tym większa, że wyrażenia są ekwiwalentne wyszukiwawczo w różnych językach, np. dyscyplina językoznawcza ma w systemie wyszukiwawczym różnojęzyczną reprezentację: ekwiwalentne terminy w różnych językach słowiańskich i w języku angielskim, reprezentowane w systemie przez słowa kluczowe (pol. *językoznawstwo*, *lingwistyka*, cz. *lingvistika*, *jazykověda*, chorw. *jezikoslovlje*, głuż. *rečespyt*, brus. *мовазнаўства*, *лінгвістыка*, ros. *лингвистика*, *языкознание*, ukr. *мовознавство*, *лінгвістика*, ang. *linguistics* itd.).

3.7. W strukturalizacji systemu leksykalnego języka słów kluczowych przyjęto także inne ważne rozwiązanie, a mianowicie dotyczące prezentowania **wieloznacznych<sup>6</sup> terminów**. W terminach tych zostaną wydobyte typowe kombinacje znaczeń umożliwiające konstruowanie w języku słów kluczowych odrębnych jednostek leksykalnych. Wyodrębnione słowa kluczowe będą opatrzone dopowiedzeniami słownymi i notacją 1, 2, 3 itd. Przykładowo wieloznaczny termin językoznawczy w języku polskim *gramatyka* jest prezen-

---

<sup>6</sup> Termin wieloznaczność stosowany jest w pracy jako synchroniczne określenie dla rozróżnianych w planie diachronii zjawisk polisemii i homonimii odniesionych do poziomu wyrażenia (znaku) językowego (por. Polański, 1993, s. 408).



towany w języku słów kluczowych przez trzy klasy ekwiwalentnych jednostek z wyspecyfikowanymi znaczeniami w postaci dopowiedzeń słownych i indeksami 1, 2, 3, por. *gramatyka 1* ‘zespół reguł funkcyjnych jednostek tworzących system językowy’, *gramatyka 2* ‘dział językoznawstwa opisujący zespół reguł funkcyjnych jednostek tworzących system językowy’<sup>7</sup> i *gramatyka 3* ‘dokument zawierający opis reguł funkcyjnych jednostek tworzących system językowy’; wieloznaczny polski termin *morfologia* jest prezentowany przez dwie klasy ekwiwalentnych jednostek leksykalnych, por. w języku polskim: *morfologia 1* ‘budowa wyrazów i form wyrazowych oraz odmiana wyrazów’ i *morfologia 2* ‘dyscyplina językoznawcza zajmująca się budową wyrazów i form wyrazowych oraz odmianą wyrazów’. Podobną strukturę semantyczną w tej grupie wieloznacznych terminów mają także m.in. pol. *słotwórstwo 1 i 2*, *etymologia 1 i 2*, *fonetyka 1 i 2*, *fonologia 1 i 2*, *fleksja 1 i 2*, *morfonologia 1 i 2*, *składnia 1 i 2* itp. Jak widać, dekompozycja znaczeń takich terminów pozwala wykryć typowe węższe znaczenia, a także relacje semantyczne między nimi.

Procedury te stanowią dopiero początek konstruowania języka słów kluczowych. Porównywane i jednocześnie łączone są ze sobą w klasy ekwiwalentów terminy równoważne w różnych językach słowiańskich i w języku angielskim. W niektórych językach naturalnych to nie jednostkowe ekwiwalenty, lecz pary jednostek lub większa ich liczba. Ze względu na zaawansowanie prac nad polskojęzyczną częścią systemu i języka informacyjno-wyszukiwawczego jednostki leksykalne języka polskiego mają pierwszeństwo we włączaniu do klas ekwiwalencji w strukturze tworzonego języka informacyjno-wyszukiwawczego. Trzeba podkreślić, że proces tworzenia klas ekwiwalentów nie ma nic wspólnego z przekładem terminów w jednym języku na inne języki.

Relacje między wydobytymi znaczeniami takich par lub trójek jednostek leksykalnych (*składnia 1 – składnia 2*, *gramatyka 1 – gramatyka 2 – gramatyka 3* itp.) mają regularny charakter, chodzi o realizowanie pewnego modelu semantycznego<sup>8</sup>. Bliskość rozpatrywanego zjawiska i pochodności słotwórczej, oczywiście pod względem semantycznym, a nie formalnym, podkreślona została w nazwie *derywacja semantyczna*<sup>9</sup>. W przywołanej grupie terminów

---

<sup>7</sup> Zuzanna Topolińska proponuje wprowadzenie do polskiej terminologii językoznawczej terminu *morfosyntaksa*, równorzędnego z *gramatyką 2* (Topolińska, 2009, s. 210).

<sup>8</sup> Szczegółowo o tym pisze Jurij D. Apresjan (Апресян, 1974, ss. 187–193).

<sup>9</sup> Jest to termin Dimitrija N. Šmeleva (Шмелев, 1968) rozpowszechniony w językoznawstwie dzięki pracom Jurija D. Apresjana. Inne rozumienie terminu *derywacja semantyczna*

i ich ekwiwalentów innosłowiańskich wieloznaczność można przedstawić jako typową relację semantyczną: ‘zjawisko’ – ‘dyscyplina językoznawcza zajmująca się badaniem danego zjawiska’ (Апресян, 1974, s. 204).

Innego typu relację można wyodrębnić w strukturze semantycznej takich wieloznacznych terminów, jak pol. *nominacja*, *kompozycja*, *kontaminacja* itp. Charakterystyczna dla tej wieloznaczności jest regularność relacji semantycznej ‘czynność/proces’ – ‘rezultat’: *kontaminacja 1* ‘zespolenie różnych elementów językowych w jeden wyraz lub związek wyrazowy’ – *kontaminacja 2* ‘wyraz lub związek wyrazowy’ i inne. Podobne zjawisko obserwujemy w innych językach, np. czeskim: *nomínace 1* i *nomínace 2*, rosyjskim: *номинация 1* i *номинация 2*.

Istotnym czynnikiem sprzyjającym wieloznaczności terminów są różnice ujęć badawczych, stosujących przyjęte powszechnie terminy w nowym znaczeniu. Często ma miejsce sytuacja polegająca na tym, że ten sam fragment rzeczywistości, zjawisko, obiekt opisywany jest przez różne szkoły i teorie lub ich przedstawicieli. Różna baza pojęciowa i inne rozumienie terminów może powodować nieporozumienia w komunikowaniu się językoznawców. Przykład stanowi polemika między Zygmuntem Salonim i Adamem Przepiórkowskim, którzy w swoich opisach dotyczących przypadku liczebników w polskim zdaniu reprezentują różne szkoły metodologiczne i używają w odmienny sposób podstawowych terminów *przypadek* i *podmiot* (Saloni, 2005, ss. 27–48).

Rozwiązanie w języku informacyjno-wyszukiawczym problemu wieloznaczności terminów, przynajmniej tzw. wieloznaczności klasycznej, polegającej na powtarzaniu się tego samego typu relacji między dwoma znaczeniami wieloznacznych terminów (Gajda, 1990, s. 71), wzbogaci strukturę semantyczną tego języka o relacje kojarzeniowe, które zapewnią większą efektywność, zwłaszcza trafność wyszukiwania informacji w systemie iSybisław. Relacyjna organizacja paradygmatyki języka słów kluczowych uwzględniająca relację synonimiczności i wyspecyfikowane niektóre relacje kojarzeniowe ma dużą wartość wyszukiawczą<sup>10</sup>.

---

znaleźć można m.in. w pracach słowotwórczych o podstawach semantycznych Stanisław Karolaka i Viary Maldjiejew.

<sup>10</sup> Szerzej o problemie synonimii i wieloznaczności terminów językoznawczych oraz o procedurach eliminowania ich skutków w języku słów kluczowych zob. w pracy: Rudnik-Karwatowa, Banasiak & Mikos, 2013.



Prace nad aparatem pojęciowo-terminologicznym systemu informacyjno-wyszukiawczego iSybisław przebiegają wieloetapowo. Opracowywane są jednojęzyczne słowniki słów kluczowych w każdym z uwzględnionych w systemie języków słowiańskich. Podobnie jak przy zbieraniu i opracowywaniu terminów językoznawczych do słownika słów kluczowych w języku polskim, tak i w tym przypadku stosowana jest metoda indukcyjno-dedukcyjna, polegająca na gromadzeniu i indeksowaniu słownictwa z tekstów reprezentatywnych dla językoznawstwa slawistycznego, a następnie sprawdzaniu i porównywaniu z istniejącymi zbiorami leksyki (słowniki, encyklopedie, gramatyki) oraz uzupełnianiu zasobu leksykalnego.

W procesie opracowywania słowników rozwiązywane są problemy relacji semantycznych synonimii i bliskoznaczności, a także wieloznaczności terminów występujących w systemie terminologicznym każdego z języków słowiańskich. Eliminowane są skutki tych zjawisk charakterystycznych dla języków naturalnych poprzez tworzenie w każdym słowniku klas ekwiwalentów, w które wchodziły słowa kluczowe, i umożliwienie użytkownikom dotarcie do informacji według dowolnego wyrażenia znajdującego się w takiej klasie ekwiwalentów.

Słowa kluczowe z poszczególnych języków łączone są ze sobą w klasy ekwiwalencji w strukturze tworzonego języka informacyjno-wyszukiawczego, umożliwiając wyszukiwanie informacji przy użyciu dowolnego wyrażenia wyszukiawczego z danej klasy ekwiwalencji. Przyjęto jednocześnie rozwiązanie polegające na tym, że językowi polskiemu – ze względu na inicjatywę i zaawansowanie prac nad systemem informacyjno-wyszukiawczym przez polskie środowisko slawistyczne – przyznano pierwszeństwo we włączaniu jego jednostek leksykalnych do klas ekwiwalencji w strukturze tworzonego języka informacyjno-wyszukiawczego. Takie preferowane wyrażenia w języku polskim reprezentują w obecnie tworzonym języku słów kluczowych całe klasy jednostek leksykalnych w językach słowiańskich i języku angielskim. Nie powoduje to jednak ograniczenia możliwości odwzorowywania treści różnojęzycznych dokumentów ani procedur strategii wyszukiawczej, bo dowolne wyrażenie z klasy ekwiwalentów służyć może jako narzędzie wyszukiawcze. W docelowym języku informacyjno-wyszukiawczym zniknie potrzeba utrzymania preferencji jednostek leksykalnych w języku polskim. Język informacyjno-wyszukiawczy będzie językiem paranaturalnym z wejściem we wszystkich językach słowiańskich i w języku angielskim.

Zbudowany na tej podstawie słownik umożliwi użytkownikom stosowanie strategii wyszukiwawczej ułatwiającej dostęp do wielojęzycznego zbioru informacji i przeszukiwanie zbioru równocześnie w wielu językach. Słownik będzie uwzględniał paradygmatykę języka słów kluczowych: relację ekwiwalencji i wieloznaczność, dzięki czemu jako narzędzie wyszukiwawcze w systemach informacyjno-wyszukiawczych umożliwi zwiększenie efektywności wyszukiwania informacji, a także zapewni dokładność i kompletność wyszukiwania relewantnej informacji w tej dziedzinie.

Funkcjonujący obecnie system informacyjno-wyszukiawczy iSybisław jest efektem realizacji początkowego etapu prac nad projektem, zaś osiągnięcie zamierzonego rezultatu docelowego wymaga jeszcze dużej pracy intelektualnej i organizacyjnej oraz działania praktycznego międzynarodowego zespołu językoznawców slawistów z polskich i zagranicznych placówek naukowych, specjalistów z informacji naukowej i informatyki.

Lawinowy przyrost piśmiennictwa z dziedziny językoznawstwa wymaga usprawnienia działań przyspieszających proces opracowywania coraz większego zbioru opisów tych dokumentów do zbioru wyszukiwawczego. Może dobrym rozwiązaniem będzie włączenie autorów prac do współtworzenia języka wyszukiwawczego w systemie iSybisław. W tym celu realizowany jest projekt rozbudowy systemu o swobodne słowa kluczowe (tagi). Efekt tego stanowi zwiększenie możliwości wyszukiwawczych: użytkownicy mogą korzystać ze słownika jawnego w postaci kartoteki wzorcowej słów kluczowych oraz ze słownika niejawnego swobodnych słów kluczowych (tagów).

Cennym *novum* w systemie iSybisław jest rozbudowa jego struktury hiper tekstowej, polegająca na połączeniu opisów bibliograficznych z pełnymi tekstami, których one dotyczą. Dostęp do dokumentów opisanych w bazie systemu realizuje się za pomocą linków. W systemie istnieje też możliwość zapisywania tekstów bezpośrednio na serwerze. Problemy o charakterze intelektualno-prawnym nie pozwalają na razie szeroko udostępniać pełnych tekstów, należy jednak spodziewać się w niedługim okresie zmiany tej sytuacji w związku z opracowaniem przez Ministerstwo Nauki i Szkolnictwa Wyższego planu wdrożenia otwartego dostępu do treści naukowych w Polsce. Podobne działania obejmą wszystkie kraje członkowskie Unii Europejskiej. Z satysfakcją możemy podkreślić, że system iSybisław od momentu jego powstania zapewniał użytkownikom otwarty dostęp do bibliograficznych metadanych dokumentów.

Podsumowując, można stwierdzić, że na obecnym etapie prac międzynarodowego zespołu tworzony jest spójny wielojęzyczny język informacyjno-

-wyszukiawczy. Jednocześnie wieloaspektowe prace teoretyczne i metodologiczne w zakresie systemów terminologii językoznawstwa sławistycznego w językach słowiańskich dają mocne podstawy do budowy wielojęzycznego słownika terminologii językoznawczej.

## Bibliografia

- Babik, W. (2010). *Słowa kluczowe*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Bibliograficzna baza danych światowego językoznawstwa sławistycznego iSybisław, b.d.; Pobrano 15 marca 2013, z <http://www.isybislaw.ispan.waw.pl>
- Bojar, B. (2002). *Słownik encyklopedyczny informacji, języków i systemów informacyjno-wyszukiawczych*. Warszawa: Wydawnictwo SBP. (Nauka, Dydaktyka, Praktyka; 56).
- Bojar, B. (2005). *Językoznawstwo dla studentów informacji naukowej*. Warszawa: Wydawnictwo SBP.
- Bojar, B. (2009). Języki informacyjno-wyszukiawcze wczoraj, dziś... czy jutro?, *Zagadnienia Informacji Naukowej*, (1), 3–24.
- Bojar, B., & Rudnik-Karwatowa, Z. (1998). Koncepcja nowoczesnego systemu informacji sławistycznej. W Dalewska-Greń, H., Rusek, J. & Siatkowski, J. (Red.), *Językoznawstwo. Prace na XII Międzynarodowy Kongres Sławistów w Krakowie 1998* (ss. 42–49). Warszawa: Energeia. (Z Polskich Studiów Sławistycznych; seria 9).
- Gajda, S. (1990). *Wprowadzenie do teorii terminu*. Opole: WSP im. Powstańców Śląskich w Opolu. (Studia i Monografie; 162).
- Mikos, Z. (2014). Rozwój funkcjonalności systemu informacji bibliograficznej językoznawstwa sławistycznego – od SYBISLAWa do iSybisława, *Studia z Filologii Polskiej i Słowiańskiej*, 49 [w druku].
- Polański, K. (Red.) (1993). *Encyklopedia językoznawstwa ogólnego*. 1993. Wrocław: Zakład Narodowy im. Ossolińskich.
- Rudnik-Karwatowa, Z., Banasiak, J., & Mikos, Z. (2013). Językowe problemy optymalizacji wyszukiwania informacji w systemie iSybisław. *Slavia Orientalis*, 62, (4), 631–646.
- Rudnik-Karwatowa, Z., & Karpińska, H. (1999). *Słownik słów kluczowych językoznawstwa sławistycznego*. Warszawa: Sławistyczny Ośrodek Wydawniczy.
- Rudnik-Karwatowa, Z., & Karpińska, H. (2006). *Słownik słów kluczowych językoznawstwa sławistycznego*. Warszawa: Sławistyczny Ośrodek Wydawniczy [CD-ROM].
- Rudnik-Karwatowa, Z., Mikos, Z., & Bojar, J. (2007). Nowoczesny system informacji sławistycznej: zadania, dotychczasowe wyniki i perspektywy, *Zagadnienia Informacji Naukowej*, (2), 19–40.
- Saloni, Z. (2005). O przypadkach w języku polskim : na marginesie artykułu Adama Przepiórkowskiego. *Biuletyn Polskiego Towarzystwa Językoznawczego*, 61, 27–48.

- Topolińska, Z. (2009). Kilka słów o naszej siatce pojęciowej i terminologicznej. W J. Mindak-Zawadzka, & I. M. Doliński, *Językowy świat Słowian. Zjawiska, interpretacje, znaki zapytania. Księga jubileuszowa na 70-lecie Profesora Włodzimierza Pianki*. Warszawa: Instytut Sławistyki Zachodniej Uniwersytetu Warszawskiego.
- Topolińska, Z. (2010). Patrząc przez cudze okulary (infinitivus – subiunctivus). *Slavia Meridionalis*, 10, 299–305.
- Апресян, Ю. Д. (1974). *Лексическая семантика. Синонимические средства языка*. Москва: Наука.
- Шмелев, Д. Н. (1968). *Лексика современного русского языка*. Москва: Наука.

### Bibliography (transliteration)

- Apresian, IU. D. (1974). *Leksicheskaia semantika. Sinonimicheskie sredstva iazyka*. Moskva: Nauka.
- Babik, W. (2010). *Słowa kluczowe*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Bibliograficzna baza danych światowego językoznawstwa slawistycznego iSybislaw (n.d.). Retrieved March 15, 2013, from <http://www.isybislaw.ispan.waw.pl>
- Bojar, B. (2002). *Słownik encyklopedyczny informacji, języków i systemów informacyjno-wyszukiawczych*. Warszawa: Wydawnictwo SBP. (Nauka, Dydaktyka, Praktyka; 56).
- Bojar, B. (2005). *Językoznawstwo dla studentów informacji naukowej*. Warszawa: Wydawnictwo SBP.
- Bojar, B. (2009). Języki informacyjno-wyszukiawcze wczoraj, dziś... czy jutro? *Zagadnienia Informacji Naukowej*, (1), 3–24.
- Bojar, B., & Rudnik-Karwatowa, Z. (1998). Koncepcja nowoczesnego systemu informacji slawistycznej. In H. Dalewska-Greń, J. Rusek, & J. Siatkowski (Eds.), *Językoznawstwo. Prace na XII Międzynarodowy Kongres Sławistów w Krakowie 1998* (pp. 42–49). Warszawa: Energeia. (Z Polskich Studiów Sławistycznych; seria 9).
- Gajda, S. (1990). *Wprowadzenie do teorii terminu*. Opole: WSP im. Powstańców Śląskich w Opolu. (Studia i Monografie; 162).
- Mikos, Z. (2014). Rozwój funkcjonalności systemu informacji bibliograficznej językoznawstwa slawistycznego – od SYBISLAWa do iSybislaw, *Studia z Filologii Polskiej i Słowiańskiej*, 49 [w druku].
- Polański, K. (Ed.). (1993). *Encyklopedia językoznawstwa ogólnego*. Wrocław: Zakład Narodowy im. Ossolińskich.
- Rudnik-Karwatowa, Z., & Karpińska, H. (1999). *Słownik słów kluczowych językoznawstwa slawistycznego*. Warszawa: Sławistyczny Ośrodek Wydawniczy.
- Rudnik-Karwatowa, Z., & Karpińska, H. (2006). *Słownik słów kluczowych językoznawstwa slawistycznego*. Warszawa: Sławistyczny Ośrodek Wydawniczy [CD-ROM].

- Rudnik-Karwatowa, Z., Banasiak, J., & Mikos, Z. (2013). Językowe problemy optymalizacji wyszukiwania informacji w systemie iSybisław. *Slavia Orientalis*, 62(4), 631–646.
- Rudnik-Karwatowa, Z., Mikos, Z., & Bojar, J. (2007). Nowoczesny system informacji sławistycznej: zadania, dotychczasowe wyniki i perspektywy. *Zagadnienia Informatyki Naukowej*, (2), 19–40.
- Saloni, Z. (2005). O przypadkach w języku polskim: na marginesie artykułu Adama Przepiórkowskiego. *Biuletyn Polskiego Towarzystwa Językoznawczego*, 61, 27–48.
- Shmelev, D. N. (1968). *Leksika sovremennogo russkogo iazyka*. Moskwa: Nauka.
- Topolińska, Z. (2009). Kilka słów o naszej siatce pojęciowej i terminologicznej. In J. Mindak-Zawadzka & I. M. Doliński, *Językowy świat Słowian. Zjawiska, interpretacje, znaki zapytania. Księga jubileuszowa na 70-lecie Profesora Włodzimierza Pianki* (pp. 209–214). Warszawa: Instytut Sławistyki Zachodniej Uniwersytetu Warszawskiego.
- Topolińska, Z. (2010). Patrząc przez cudze okulary (infininitivus – subiunctivus). *Slavia Meridionalis*, 10, 299–305.

## **Multilingual information retrieval system of Slavic linguistics iSybisław: challenges, achievements and opportunities**

### **Summary**

In multilingual information retrieval (IR) system iSybisław we introduce a number of solutions enlarging its efficiency and effectiveness with the main tool, which presents the keyword language. In this paper the applicability and consistency of the accepted approach in the presentation of the content of documents and IR is being discussed. We use natural language (professional language based on linguistic terminology) in the construction of the keyword language and implementation of the tools of semantic organization and access to multilingual informational database. The core task is to make allowance in the retrieval database for relations of equivalence between the keywords in different Slavic languages and in the English language, which will enable the simultaneous search in different languages. Implementation of the natural language enables optimization of the IR language but on the other hand it creates problems connected with synonymy and polysemy of the

terms used as keywords. Such problems are being solved with implementation of the classes of equivalent keywords, pairs and more complicated units of monosemantic keywords presenting the terms linked with each other by semantic relations.

Burst-type increase of written documents in the field of linguistics requires taking effort in speeding up the processing of the growing number of representations of these documents into search database. So we decided to call upon the users (the authors of the works considered in the system) to take part in creation of the IR language in the iSybislaw system through tagging.

**Keywords:** database; iSybislaw; keyword; multilingual information retrieval system; Slavic languages; Slavic linguistics

**Słowa kluczowe:** baza danych; iSybislaw; języki słowiańskie; językoznawstwo sławistyczne; słowo kluczowe; wielojęzyczny system informacyjno-wyszukiawczy