

Zofia Rudnik-Karwatowa, Jakub Banasiak, Zenon Mikos
Warszawa

Językowe problemy optymalizacji wyszukiwania informacji w systemie *iSybislaw*

Nowoczesny system informacji slawistycznej *iSybislaw*, prezentowany przez *Bibliograficzną bazę danych światowego językoznawstwa slawistycznego* (<http://www.isybislaw.ispan.waw.pl>), rozwija się, w głównych zarysach, zgodnie z koncepcją przedstawioną na XII Międzynarodowym Kongresie Slawistów w Krakowie¹. Realizowane są możliwości funkcjonalne systemu, powiększa się dynamicznie jego zbiór informacyjny, udostępniając obecnie w Internecie szczegółowe informacje o polskim i światowym dorobku naukowym językoznawstwa slawistycznego od roku 1992 (przede wszystkim od roku 1998) do chwili obecnej².

Nieograniczony zasięg **językowy** dokumentów uwzględnianych w zbiorze informacyjnym (w systemie informacyjno-wyszukiwawczym znajdują się informacje o dokumentach we wszystkich językach słowiańskich, w językach międzynarodowych, a także innych, np. estońskim, węgierskim) wymaga od twórców systemu rozwiązania szeregu problemów metodologicznych, zapewniających jego efektywność, przede wszystkim sprawność podstawowego narzędzia informacyjnego systemu *iSybislaw*, jakim jest język informacyjno-wyszukiwawczy: język słów kluczowych³. Zastosowanie w systemie *iSybislaw* języka mającego charakter naturalny (notację i strukturę naturalną, lub paranaturalną) umożliwi użytkownikom samodzielne i wygodne wyszukiwanie, a także samodzielną ocenę relewancji otrzymanej informacji, indeksatorom zaś ułatwia opracowywanie informacji dla systemu⁴. Elementarnymi jednostkami leksykalnymi języka słów kluczowych są słowa kluczowe. Informując o treści dokumentów i umożliwia-

¹ Zob. B. Bojar, Z. Rudnik-Karwatowa, *Koncepcja nowoczesnego systemu informacji slawistycznej*, [w:] *Językoznawstwo. Prace na XII Międzynarodowy Kongres Slawistów w Krakowie 1998* (Z Polskich Studiów Slawistycznych; seria 9), Warszawa 1998, s. 42–47.

² Obecnie udostępnionych jest w Internecie ponad 18 tysięcy rozbudowanych rekordów zawierających szczegółowe informacje o formie i treści dokumentów.

³ Rolę słów kluczowych we współczesnych systemach informacyjnych podkreśla W. Babik w książce *Słowa kluczowe*, Kraków 2012.

⁴ Zob. B. Bojar, Z. Rudnik-Karwatowa, *Koncepcja...*, s. 45.

jąc wyszukanie informacji, stanowią główne narzędzie indeksacyjne i wyszukiwawcze.

Różnojęzyczny jest nie tylko zbiór informacji o dokumentach, różnojęzyczni są także użytkownicy systemu. W związku z tym zapewniliśmy użytkownikom bezpośrednie komunikowanie się z systemem i korzystanie z niego za pomocą interfejsu w języku polskim, rosyjskim i angielskim.

Istotnym zadaniem (i celem) jest znalezienie i zastosowanie w **wielojęzycznym systemie informacyjno-wyszukiwawczym** rozwiązań, które umożliwią użytkownikom efektywne, tj. dokładne i kompletne, wyszukiwanie informacji. Zgodnie z przyjętą koncepcją takim rozwiązaniem w systemie *iSybislaw* jest ustanowienie **klas ekwiwalencji** słów kluczowych. Wejdą do nich ekwiwalentne jednostki językowe należące do różnych języków a pełniące tę samą funkcję denotacyjną (w językach informacyjno-wyszukiwawczych denotacją bezpośrednią jest zbiór dokumentów, do którego odsyła dane wyrażenie, a denotacją pośrednią rzeczywistość pozadokumentacyjna), np. ekwiwalentne terminy w różnych językach słowiańskich i w języku angielskim, reprezentowane w systemie przez słowa kluczowe (pol. *językoznawstwo, lingwistyka*, cz. *lingvistika, jazykověda*, chorw. *jezikoslovlje*, głuź. *rečespyt*, brus. *мовазнаўства, лінгвістыка*, ros. *лингвистика, языкознание*, ukr. *мовознавство, лінгвістика*, ang. *linguistics* itd.). Ze względu na zakres tematyczny systemu *iSybislaw* najliczniej reprezentowane są w zbiorze wyszukiwawczym ekwiwalentne terminy językoznawstwa sławistycznego, inne to m.in. nazwy własne: osobowe, geograficzne, korporatywne itp.

Z budowaniem klas ekwiwalencji wiążą się różnego typu transformacje polegające na tworzeniu informacji wyjściowych z informacji źródłowych, które tylko częściowo były omówione w innych pracach⁵. W niniejszym artykule skoncentrujemy się przede wszystkim na problemach **synonimii** i **wieloznaczności** terminologii językoznawczej oraz na rozwiązaniach dotyczących strukturalizacji języka informacyjno-wyszukiwawczego, które zapewniają odpowiednią reprezentację informacji terminologicznej i odpowiednie procedury wyszukiwawcze. Przedstawimy w związku z tym językowe problemy dotyczące podsystemu informacji terminologicznej językoznawstwa sławistycznego, zawierającego ekwiwalentne terminy w różnych językach słowiańskich i terminy wieloznaczne, oraz procedury eliminowania w języku słów kluczowych skutków synonimii i wieloznaczności terminów.

Prace nad aparatem pojęciowo-terminologicznym systemu informacyjno-wyszukiwawczego *iSybislaw* przebiegają wieloetapowo. W każdym z uwzględnionych w bazie języków słowiańskich, a także w języku angielskim, opracowywany jest jednojęzyczny słownik słów kluczowych. Przy gromadzeniu terminów języko-

⁵ Zob. m.in. Z. Rudnik-Karwatowa, Z. Mikos, J. Bojar, *Nowoczesny system informacji sławistycznej: zadania, dotychczasowe wyniki i perspektywy*, „Zagadnienia Informacji Naukowej” 2007, nr 2 (90), s. 19–40.

znawczych do słowników 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).

Na kolejnym etapie prac rozwiązywane są w każdym z opracowywanych słowników problemy relacji semantycznych synonimii i bliskoznaczności oraz wieloznaczności terminów, występujących w systemie terminologicznym każdego z języków naturalnych. Zadanie to jest realizowane poprzez eliminację skutków tych zjawisk, charakterystycznych dla języków naturalnych, i uwzględnienie w każdym z opracowywanych słowników klas ekwiwalencji, w które wchodzi słowa kluczowe, i umożliwienie użytkownikom dotarcie do informacji według dowolnego wyrażenia znajdującego się w klasie ekwiwalentów. Następnie słowa kluczowe z poszczególnych języków łączone są ze sobą w klasy ekwiwalencji w strukturze tworzonego języka docelowego. Prace te przebiegają wielokierunkowo, ponieważ każdy język naturalny traktowany jest równorzędnie. Pozwala to uniknąć błędów wynikających m.in. z asymetrii leksykalnej. Proponowane rozwiązania metodologiczne służą głównie optymalizacji wyszukiwania informacji w bazie.

Wartość użytkową systemu *iSybislaw* już na obecnym etapie podnosi zastosowanie układu deskryptorowo-askryptorowego jednostek leksykalnych języka słów kluczowych, który – umożliwiając wyszukiwanie informacji przy użyciu dowolnego terminu wyszukiwawczego z danej klasy ekwiwalencji – wskazuje jednocześnie preferencję wybranego terminu języka naturalnego. Przyjęte rozwiązanie służy ponadto jako środek organizacji leksyki tworzonego słownika słów kluczowych w każdym języku naturalnym, a przede wszystkim docelowego języka informacyjno-wyszukiwawczego. Jak wiadomo, językowi polskiemu – ze względu na inicjatywę i zaawansowanie prac nad systemem informacyjno-wyszukiwawczym przez polskie środowisko slawistyczne – przyznano pierwszeństwo we włączaniu jego jednostek leksykalnych do klas ekwiwalencji w strukturze docelowego języka informacyjno-wyszukiwawczego. Właśnie preferowane terminy języka polskiego reprezentują w obecnie tworzonej strukturze słów kluczowych całe klasy terminów w językach słowiańskich i języku angielskim.

Ponieważ system jest dostępny w wersji elektronicznej, użytkownik nie musi zaznajamiać się z całym aparatem pojęciowo-terminologicznym ujętym w języku słów kluczowych, każde słowo kluczowe z danej klasy ekwiwalencji jest bowiem z użytkowego punktu widzenia funkcjonalnym odpowiednikiem terminu.

Pewna utrata informacji związana z językowymi preferencjami autorów zaindeksowanych dokumentów jest minimalizowana poprzez zawarcie w opisie dokumentu streszczeń (w których aparat pojęciowo-terminologiczny autora pierwotnego tekstu jest w pełni zachowany i widoczny).

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-wyszukiwawczych umożliwi zwiększenie efektywności wyszukiwania informacji, zapewni dokładność i kompletność wyszukiwania relewantnej informacji w tej dziedzinie.

Uwzględnienie relacji ekwiwalencji (oraz rozwiązanie problemu wieloznaczności⁶) zbliża nas do tzw. wyszukiwania semantycznego, gdzie w istocie o odpowiedzi na zapytanie decyduje nie obecność danej konkretnej formy językowej, lecz obecność znaczenia, które ona posiada. Podejście takie jest zgodne z nowymi tendencjami w zakresie technologii informacji, m.in. w pracach nad siecią semantyczną.

Terminologia językoznawcza, stanowiąca trzon języka słów kluczowych w systemie *iSybislaw*, podlega na poziomie swojej funkcji nominatywnej prawidłowościom i przemianom analogicznym do tych, które zachodzą w słownictwie ogólnym. Z punktu widzenia zastosowań metainformacyjnych w systemach informacyjno-wyszukiwawczych stanowi to poważny problem zarówno dla użytkowników, jak i indeksatorów. System leksykalny języka podlega w czasie licznym procesom, które skutkują występowaniem zjawisk takich jak synonimia i wieloznaczność. W systemach informacyjno-wyszukiwawczych zjawiska takie są niepożądane, ponieważ obniżają skuteczność wyszukiwawczą: synonimia powoduje niekompletność wyszukiwania, wieloznaczność przyczynia się do powstawania szumu informacyjnego.

Ostre rozgraniczenie zjawisk synchronii i diachronii (chodzi o względnie duży przedział czasowy) jest utrudnione w obszernych bankach danych, w których prace starsze współistnieją z nowszymi. Niemożliwe jest zastosowanie spójnego języka słów kluczowych opartego na spójnym systemie pojęciowo-terminologicznym danej dyscypliny naukowej. Systemy pojęciowo-terminologiczne odzwierciedlają bowiem relacje synchroniczne między ich składnikami i zmieniają się wraz z rozwojem tej dyscypliny czy poddyscypliny naukowej. Nie ma mowy o całkowitej przebudowie systemu; procesy innowacyjne przejawiają się w różnym stopniu w różnych podpolach semantycznych i są zróżnicowane w poszczególnych językach. Ważnym czynnikiem sprzyjającym zachowaniu terminów w dynamicznie zmieniającym się systemie pojęciowo-terminologicznym jest tradycja nazewnicza.

Encyklopedia językoznawstwa ogólnego podaje następującą definicję synonimii: „Wyrażanie tej samej treści za pomocą dwu lub więcej różnych form języko-

⁶ Termin *wieloznaczność* stosowany będzie w pracy jako synchroniczne określenie dla rozróżnianych w planie diachronii zjawisk polisemii i homonimii odniesionych do poziomu wyrażenia (znaku) językowego (por. *Encyklopedia językoznawstwa ogólnego*, pod red. K. Polańskiego, Wrocław 1993, s. 408).

wych⁷. Co prawda, swoista ekonomia językowa, charakterystyczna również dla języków specjalistycznych, nieraz prowadzi do dyferencjacji znaczeń pierwotnie synonimicznych wyrażen⁸, jednak w przypadku tworzenia słownika będącego narzędziem informacyjno-wyszukiwawczym konieczne jest połączenie wyrażen synonimicznych i bliskoznacznych lub usunięcie tych z nich, które twórcy systemu z różnych względów uznałyby za redundantne. Drugie z powyższych rozwiązań wymagałoby jednak od użytkownika systemu informacyjno-wyszukiwawczego dokładnego zaznajomienia się z aparatem pojęciowo-terminologicznym zastosowanym w indeksacji, co utrudniłoby proces wyszukiwania informacji. Oczywiście twórcy systemu *iSybislaw* zdają sobie sprawę ze złożoności zjawisk i przemian typowych dla podsystemu terminologicznego i w miarę możliwości je uwzględniają.

W pracy abstrahujemy w pewnym stopniu od problemu istnienia i ewentualnego zakresu synonimii właściwej, choć wydaje się, że w ramach podsystemu terminologicznego jest to mniej problematyczne niż dla pozostałych podsystemów językowych. Za podstawową własność znaku językowego uznaje się w pracy posiadanie zarówno konotacji (sensu, intensji), jak i denotacji (zakresu, odniesienia, oznaczenia w rozumieniu Frege'owskim). Ze specyfiki językoznawstwa wynika, że (pomiając terminy metanaukowe – nazwy metod i prądów badawczych) podstawowym odniesieniem terminów są szeroko rozumiane zjawiska z rzeczywistości językowej. Trzon terminów językoznawczych pełni zatem funkcję metajęzykową. Użyte w systemie informacyjno-wyszukiwawczym pełnią one funkcję metainformacyjną, dostarczając danych na temat informacji metajęzykowej i metanaukowej, zawartej w opisywanych pracach. W informacji naukowej często akcentuje się konieczność takiego doboru słownictwa w ramach słownika słów kluczowych, aby było ono maksymalnie bogate informacyjnie (znaczące) i tym samym miało zakres wyznaczony w sposób możliwie jednoznaczny. Nie ulega wątpliwości, że ścisłe definicje stanowią niezwykle ważny składnik dobrego warsztatu naukowego. Należy jednak zwrócić uwagę, że nawet w obrębie jednego języka możliwe jest wyznaczenie tego samego zakresu na różne sposoby, poprzez różne nazwy definiowane i rozumiane na rozmaite sposoby. Zjawisko to w języku ogólnym opisywane jest jako tzw. profilowanie⁹, jednak warto zaznaczyć, że w przypadku terminologii często problem stanowią nie tyle uwzględnione w danej definicji cechy przypadkowe, co właśnie cechy istotne (*differentia specifica*).

Praca związana z tworzeniem systemu informacyjno-wyszukiwawczego wymaga rozwiązania wielu problemów teoretycznych i metodologicznych, ale przyświe-

⁷ *Encyklopedia językoznawstwa...*, s. 533.

⁸ Z oczywistych przyczyn abstrahujemy tu od względów stylistycznych.

⁹ Por. prace J. Bartmińskiego, w których profile pojęć traktowane są jako sposoby organizacji treści semantycznej w ramach podstawowego znaczenia, m.in. *O profilowaniu pojęć w słowniku etnolingwistycznym*, [w:] *Profilowanie pojęć. Wybór prac*, zest. J. Bartmiński, Lublin 1993, s. 8.

ca jej inny cel niż w rozważaniach stricte teoretycznych (naukowych czy wręcz metanaukowych). Ponieważ za najważniejsze uznajemy potrzeby informacyjne użytkownika-językoznawcy lub osoby o podobnych zainteresowaniach, założyliśmy, że szuka on głównie treści metajęzykowych (informacji na temat zjawisk rzeczywistości językowej). W pracach nad jednojęzycznymi słownikami słów kluczowych językoznawstwa slawistycznego za decydujące o powiązaniu wyrażenia relacją synonimii wyszukiwawczej uznaliśmy odsyłanie terminów metajęzykowych do tego samego zbioru elementów językowych. To, co dane terminy orzekają o danym wycinku rzeczywistości językowej, uznaliśmy w tym przypadku za drugorzędne wobec samego faktu identyczności wyznaczanych przez nie zbiorów. Dla przykładu nie ulega wątpliwości, że wszystkie stosowane w polskich pracach slawistycznych terminy: *tryb nieświadka*, *narrativus/narratyw*¹⁰, *imperceptivus* oraz *tryb imperceptywny* odsyłają do tego samego zestawu form czasownika bułgarskiego lub macedońskiego, lecz robią to w odmienny sposób. Chaos terminologiczny, jaki panuje w tym zakresie, skłania niektórych badaczy do preferowania „cytowania” terminów pochodzenia obcego (oczywiście można polemizować co do samego podanego niżej ciągu „synonimicznego”), por. „Категория прекажаност, będąca przedmiotem niniejszego opracowania, przez językoznawców nazywana jest: trybem imperceptywnym, kategorią postawy, stroną świadka i nieświadka, dubitativem; terminy te będą więc traktowane równorzędnie”¹¹. Zróżnicowanie sensów terminów językoznawczych o tym odniesieniu w języku polskim, rosyjskim i bułgarskim przedstawia tabela 1¹².

Wszystkie przedstawione w tabeli terminy, zdefiniowane w ramach prac nad wielojęzycznym systemem informacyjno-wyszukiwawczym *iSybislaw* realnie i/ lub ostensywnie (tj. w sposób wyznaczający ich zakres jednoznacznie przy jednoczesnym abstrahowaniu od pewnych różnic konotacyjnych), po przetransformowaniu w słowa kluczowe zostały połączone relacją ekwiwalencji wyszukiwawczej (na poziomie poszczególnych języków relacją synonimii wyszukiwawczej).

¹⁰ Terminy te są zresztą wieloznaczne, mogą bowiem odnosić się do przypadku gramatycznego.

¹¹ M. Ledzion-Jelen, *Sposoby oddawania macedońskiej kategorii прекажаност w języku polskim i niemieckim*, [w:] *Kategorie w języku, język w kategoriach*, pod. red. M. Cichońskiej, Katowice 2009, s. 130.

¹² Por. np.: V. Koseska-Toszewa, M. Korytkowska, R. Roszko, *Polsko-bułgarska gramatyka konfrontatywna*, Warszawa 2007, s. 71–88; В. Косеска-Тошева, *Функции и форми на перфекта в българските диалекти*, [w:] *Българският език през XX век*, отг. ред. В. Радева, София 2001, s. 130–138; M. Ledzion-Jelen, *Sposoby oddawania macedońskiej...*, s. 130–144; В. Аврамова, *Строение человека в болгарской языковой картине мира (на фоне русской языковой картины мира)*, [w:] *Проблемы когнитивного и функционально-коммуникативного описания русского и болгарского языков 5*, науч. ред. Т. Иванова, Шумен 2007, s. 153–169; A.B. Letuchiy, *Сравнительные конструкции, ирреалис и эвиденциальность*, [w:] *Lexikalische Evidenzialitätsmarker im Slavischen*, eds. V.A. Plungjan, B. Wiemer, Wien 2008 (= *Wiener Slawistischer Almanach; Sonderband 72*), s. 215–238.

Tabela 1.

Składnik sensu	język polski	język bułgarski	język rosyjski
element renarracji immanentnie obecny w nazwie trybu	(re)narrativus/ narratyw	преизказно наклонение	пересказывательное наклонение
element niebycia świadkiem immanentnie obecny w nazwie trybu	tryb nieświadka	несвидетелско наклонение	несвидетельское наклонение
element niepercypowania immanentnie obecny w nazwie trybu	imperceptivus/ tryb imperceptywny	имперцептив	заглазное наклонение / имперцептив

W bazie konsekwentnie przestrzega się rozróżniania dwóch poziomów rzeczywistości językowej – planu wyrażania i treści (mimo że jeśli konsekwentnie traktować rzeczywistość językową znakowo, rozróżnienie to uznać by można za drugorzędne)¹³. Skutkuje to rozłącznym traktowaniem nazw jednostek semantycznych, takich jak *imperceptywność*, i nazw ich wykładników jak *imperceptivus*. Taki podział szczególnie kłopotliwy okazuje się wobec pewnych zjawisk związanych z regularnym charakterem różnicy *treść – wykładnik* wykorzystywanym czasem w pracach językoznawczych. Obraz wyłaniający się w tym względzie z poszczególnych języków wynika w dużym stopniu z uzusu i tradycji. Zarówno w języku polskim i bułgarskim, jak i rosyjskim termin *predykat* funkcjonuje zarówno jako nazwa składnika struktury semantycznej, jak i syntaktycznej. Chcąc zachować w bazie konsekwencję, uznaliśmy to drugie znaczenie predykatu za sekundarne i opatrzyliśmy subskryptem. Otrzymane dla badanych języków ciągi synonimiczne połączone w jedną klasę ekwiwalencji wyszukiwawczej przedstawia tabela 2.

Nie ulega wątpliwości, że zamienne stosowanie wszystkich podanych terminów w jednej pracy naukowej (czy nawet szerzej, w jednym idiolekcie terminologicznym) prowadziłyby do niespójności. Analiza materiału wykazała, że preferencje autorskie w tym względzie są bardzo różnorodne i różnie motywowane (przy czym derywat terminologiczny *predyktor* z formantem *-or* reprezentuje produktywny w Słowiańszczyźnie typ słotwórczy o charakterze międzynarodowym)¹⁴. Warto

¹³ K. Bogacki, S. Karolak, *Założenia gramatyki o podstawach semantycznych*, [w:] *Podstawy metodologiczne semantyki współczesnej*, pod red. I. Nowakowskiej-Kempnej, Wrocław 1992, s. 157–187.

¹⁴ Termin *wyrażenie predykatywne* bardzo konsekwentnie stosuje np. Z. Topolińska, z kolei A. Kiklewicz redefiniuje go i stosuje w innym znaczeniu (por.: Z. Topolińska, *Język, człowiek, prze-*

Tabela 2.

język polski	język bułgarski	język rosyjski	język ukraiński
<i>wyrażenie predykatywne</i>	<i>предикативен израз</i>	<i>предикативное выражение</i>	<i>предикативне висловлювання</i>
<i>predykat 2</i>	<i>предикат 2</i>	<i>предикат 2</i>	<i>предикат 2</i>
<i>predyktor</i>	<i>предикатор</i>	<i>предикатор</i>	<i>предикатор</i>
<i>predykat składniowy</i>	<i>синтактичен предикат</i>	<i>синтаксический предикат</i>	<i>синтаксичний предикат</i>
<i>orzeczenie</i>	<i>сказуемо</i>	<i>сказуемое</i>	<i>присудок</i>

również zaznaczyć, że niektóre z podanych terminów stosuje się niezwykle rzadko (np. ukr. *предикативне висловлювання* czy ros. *предикатор*, spotykane głównie w tłumaczeniach)¹⁵. Podobnie rozbudowany ciąg synonimiczny można utworzyć dla rosyjskiego terminu *предикатив*¹⁶. W języku rosyjskim stosowane są następujące terminy: *категория состояния*, *безлично-предикативное слово*, *безличное наречие*, *предикативное наречие*, *предикативное слово*, *предикатив*, *присвязка*, *присвязочное безличное сказуемое*. W języku polskim z kolei: *czasownik nieosobowy*, *czasownik niefleksyjny*, *czasownik niewłaściwy*, *predykatyw*, *kategoria stanu*. Należy zwrócić jednak uwagę, że klasyfikacje części mowy stosunkowo rzadko konstytuują zbiory nierozmyte i często możliwe są złożone relacje wiążące poszczególne terminy. Tak np. w polskim systemie terminologicznym możliwe jest stosowanie nazwy *predykatyw* w szerokim sensie, synonimicznie z szeroko rozumianym czasownikiem (a zatem *predykatyw 2*, czyli *predykatyw* w powyższym sensie wyznaczałby zbiór, względem którego zbiór wyznaczony terminem *predykatyw 1* (w rozumieniu wąskim) byłby w relacji inkluzji)¹⁷. Obecnie rzadsze jest rozumienie *predykatywu* jako jednego z przypadków gramatycznych, por. „ПРЕДИКАТИВ. Периферийный падеж, выражающий класс объектов, свойства которых приписываются другому имени (работать контролером;

strzeń, Warszawa; Kraków, 1999, s. 57; A. Kiklewicz, *Podmiot i orzeczenie jako kategorie gramatyki funkcjonalnej*, „Prace Językoznawcze” 2003, nr 5, s. 117–139).

¹⁵ Por. np. Ч. Филлмор, *Дело о падеже*, [w:] *Новое в зарубежной лингвистике*, Вып. 10: *Лингвистическая семантика*, сост. В.А. Звегинцев, Москва 1981, s. 369–495.

¹⁶ В.Н. Немченко, *Введение в языкознание*, Москва 2008, s. 417.

¹⁷ Z. Kubiszyn-Mędrala, *Czasowniki nieosobowe (niewłaściwe), ich miejsce w polskim systemie morfologiczno-syntaktycznym*, „Biuletyn Polskiego Towarzystwa Językoznawczego” 2000, z. 58, s. 95–104.

выглядеть фразом)''¹⁸. Co ciekawe, termin *predykatyw* w powyższym sensie nie jest obcy również polskiej tradycji językoznawczej (gdzie konkuruje on z terminem *orzecznik*), por. „Potrzeby konfrontacji gramatycznej obu języków, którymi się zajmuję, zmusiły mnie z kolei do rewizji statusu orzecznika polskiego. W poziomie semantycznym i syntaktycznym paralela z sytuacją macedońską jest niewątpliwa. W poziomie morfologicznym polszczyzna pokazuje synkretyzm predykatywu z instrumentalem''¹⁹.

Złożone relacje semantyczne zachodzące między terminami oraz występowanie rozbieżnych konwencji terminologicznych nie zmieniają faktu, że leksykalny podsystem terminologiczny charakteryzuje dążenie do systemowości. Terminy, których znaczenie uległo nadmiernemu rozmyciu, „zużywają się” i stopniowo wychodzą z użycia; por. np. rezygnacja ze stosowania terminu *agens* w pracach M. Korytkowskiej²⁰. Z kolei tzw. puste miejsca (jednostki potencjalne) często pustymi pozostają wobec braku wyraźnej potrzeby nazewnicznej. Obserwacja tego stanu rzeczy prowadzi do trywialnego wniosku, że od językoznawców oczekuje się kompetencji w zakresie terminologii lingwistycznej. Języki różnią się znacznie między sobą i wnioski oparte na materiale jednojęzycznym często nie są reprezentatywne dla zbioru wielojęzycznego. Nawet blisko spokrewnione języki charakteryzuje wspomniana już asymetria leksykalna. Tradycyjne podejście od języka wejściowego do docelowego może nie dawać pełnego obrazu języka docelowego. Jak wspomniano wyżej, istotnym novum metodologicznym prac nad systemem *iSybislaw* jest odrzucenie takiego podejścia i przyjęcie równoległego badania poszczególnych języków i łączenia wyrażen z poszczególnych języków relacjami ekwiwalencji. Poniższa tabela przedstawia powiązane relacją synonimii ciągi terminów wielocłonowych dla trzech różnych języków. W ciągach tych wyróżnić można dodatkowo pary terminów będące swoistymi wariantami kombinatorycznymi. W tabeli 3 uwzględniono również jednostki potencjalne (wyrażenia przekreślone).

Jak widać, dla każdego z badanych języków zbiór wyrażen jest inny i specyficzny. Zwróćmy uwagę, że terminy w tabeli zostały pogrupowane na zasadzie analogii członów określanych i określających (rodzimych i obcych). Układ taki nie jest jednak decydujący, jeśli chodzi o kwestie związane z dystrybucją (funkcjonowaniem) jednostek w poszczególnych językach, która mimo ekwiwalencji semantycznej i podobieństwa formalnego może być różna.

Z informacyjno-wyszukiwawczego punktu widzenia rozróżnianie zjawisk synonimii i wariacji jest bezzasadne. W obu wypadkach różne formy językowe wyrażają te same treści i optymalizacja wyszukiwania informacji wymaga połą-

¹⁸ Т.В. Жеребило, *Словарь лингвистических терминов*, Назрань 2010, s. 283.

¹⁹ Z. Topolińska, *Czy orzecznik to przypadek*, [w:] *Podstawy metodologiczne semantyki współczesnej*, pod red. I. Nowakowskiej-Kempnej, Wrocław 1992, s. 192.

²⁰ M. Korytkowska, *Typy pozycji predykatowo-argumentowych*, [w serii:] *Gramatyka konfrontatywna bułgarsko-polska*, t. 5, cz. 1, Warszawa 1992, s. 155–191.

Tabela 3.

język polski	język bułgarski	język rosyjski
<i>grupa nominalna</i>	<i>номинална група</i>	<i>номинальная группа</i>
<i>grupa imienna</i>	<i>именна група</i>	<i>именная группа</i>
<i>fraza nominalna</i>	<i>номинална фраза</i>	<i>номинальная фраза</i>
<i>fraza imienna</i>	<i>именна фраза</i>	<i>именная фраза</i>
<i>syntagma nominalna</i>	<i>номинална синтагма</i>	<i>номинальная синтагма</i>
<i>syntagma imienna</i>	<i>именна синтагма</i>	<i>именная синтагма</i>

czenia wariantów i/lub synonimów w jedną klasę ekwiwalencji. Istnieją różne ujęcia opisujące wariantywność na poziomie morfemów, uwzględnienie wariantów słowotwórczych z kolei zmusza do zadania pytania o charakter relacji wiążącej równoznaczne terminy złożone, w których jeden z członów został zamieniony na identyczny funkcjonalnie (i semantycznie) element (por. wyżej). Pewną różnicę stanowi, pozwalająca przewidzieć tzw. miejsca puste, systemowość wariacji oraz obecność elementu (elementów) wspólnego w wariantach. Pod tym względem synonimię określić można jako mniej systemową i bardziej nieregularną. Osobny problem stanowi również możliwość wystąpienia w kilku językach naraz wariantów wykładników ekwiwalentnych pojęć, które mają jednak odmienny charakter (np. fonetyczny vs fleksyjny) por. ros. *алломорф / алломорфа* i pol. *allomorf / alomorf*. Znajomość faktów językowych pozwala na stwierdzenie, że typowa wariantywność jest w terminologii zjawiskiem stosunkowo rzadkim, jednak synchroniczna dynamika języka przejawia się i tu, choć głównie w konkurencji wyrażen i formantów rodzimych i pochodzenia obcego.

W organizacji struktury systemu leksykalnego języka słów kluczowych przyjęto także inne ważne rozwiązanie – dotyczące prezentowania **wieloznacznych terminów**. W terminach tych zostaną wydobyte typowe kombinacje znaczeń umożliwiające konstruowanie w języku słów kluczowych odrębnych jednostek leksykalnych prezentujących w języku naturalnym (paranaturalnym) relewantne informacje zawarte w dokumentach. Wyodrębnione słowa kluczowe będą opatrzone dopowiedzeniami słownymi i notacją 1, 2, 3 itd. Na przykład wieloznaczny termin językoznawczy w języku polskim *gramatyka* jest prezentowany w języku słów kluczowych przez trzy klasy ekwiwalentnych jednostek – we wszystkich językach słowiańskich i języku angielskim – z wyspecyfikowanymi znaczeniami w postaci dopowiedzeń słownych i indeksami 1, 2, 3, por. w języku polskim: *gramatyka 1* ‘zespół reguł funkcyjnych jednostek tworzących system językowy’, *gramatyka 2* ‘dział językoznawstwa opisujący zespół reguł funkcyjnych jednostek tworzących

system językowy’ i *gramatyka 3* ‘dokument zawierający opis reguł funkcyjnych jednostek tworzących system językowy’; polski termin *morfologia* jest prezentowany przez dwie klasy ekwiwalentnych jednostek leksykalnych w językach słowiańskich i języku angielskim, 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’; wieloznaczny termin pol. *słotwórstwo* jest prezentowany przez dwie klasy ekwiwalentnych słów kluczowych w językach słowiańskich i języku angielskim, por. pol.: *słotwórstwo 1* ‘tworzenie nowych wyrazów’ i *słotwórstwo 2* ‘dyscyplina językoznawcza zajmująca się tworzeniem nowych wyrazów’. Ze względu na podobieństwo struktury semantycznej w tej grupie wieloznacznych terminów znajdują się także m.in. pol. *etymologia*, *fonetyka*, *fonologia*, *fleksja*, *morfologia*, *składnia* itp. wraz z ekwiwalentami innosłowiańskimi i w języku angielskim. Jak widać, dekompozycja znaczeń takich terminów pozwala wykryć typowe węższe znaczenia, a także relacje semantyczne między nimi.

Opisane procedury stanowią zaledwie początek konstruowania języka słów kluczowych, bowiem do wyrażen w języku polskim dodawane są w następnej kolejności ekwiwalenty w innych językach. Trzeba zaznaczyć, że proces ten nie ma nic wspólnego z przekładem terminów w jednym języku na inny język (inne języki). Porównywane są i jednocześnie łączone ze sobą w klasy ekwiwalencji terminy równoważne w różnych językach słowiańskich i w języku angielskim. W niektórych językach naturalnych będą to nie jednostkowe ekwiwalenty, lecz pary lub więcej jednostek (terminy synonimiczne mogą sąsiadować ze sobą w tekście, np. ukr. *словотвір* i *словотворення* ‘tworzenie nowych wyrazów’: Спираючись на результати значного за обсягом картографування явищ *словотвору* в польских говірках [...] А. Ковальська так само дійшла висновку про високу ареологічну цінність *словотворення*: „Така кількість з проблематики виключно *словотворення* [...] уповажнює на твердження: факти *словотворення* [...] мають не менше значення...”²¹, por. np. dwie klasy ekwiwalentów powstałe po wyeliminowaniu wieloznaczności terminów:

- pol. *słotwórstwo 1* ‘tworzenie nowych wyrazów’, cz. *tvoření slov 1*, *słotvorba 1*, ros. *словообразование 1*, ukr. *словотвір 1*, *словотворення 1*, brus. *словаўтварэнне 1*, ang. *word-formation 1* itd.;
- pol. *słotwórstwo 2* ‘dyscyplina językoznawcza zajmująca się badaniem tworzenia nowych wyrazów’, cz. *słotvorba 2*, *tvoření slov 2*, ros. *словообразование 2*, ukr. *словотвір 2*, *словотворення 2*, brus. *словаўтварэнне 2*, ang. *word-formation 2* itd.

²¹ П.Ю. Гриценко, *Лінгвістичне картографування і проблеми словотворення*, [w:] *Відображення історії та культури народу в словотворенні*, упор. та ред. Н.Ф. Клименко і Є.А. Карпіловської, Київ 2010, s. 366.

Jak już wspomniano, w pierwszej kolejności tworzony jest język informacyjno-wyszukiwawczy oparty na języku polskim i to właśnie tylko w tym języku ze względów metodologicznych wskazuje się preferencję słów kluczowych w obrębie poszczególnych klas ekwiwalentów. W docelowym języku informacyjno-wyszukiwawczym preferencja nie będzie uwzględniana, ponieważ dowolne słowo kluczowe z klasy ekwiwalentów w dowolnym języku jest funkcjonalnym odpowiednikiem terminu i wyszukiwanie można przeprowadzać za pomocą któregośkolwiek z nich.

Językoznawcy już pod koniec lat 60. XX wieku odnotowali, że relacje między wydobytymi znaczeniami takich par lub trójek jednostek leksykalnych (*składnia 1 – składnia 2, gramatyka 1 – gramatyka 2 – gramatyka 3* itp.) mają regularny charakter i że chodzi o realizowanie pewnego modelu semantycznego²². Bliskość rozpatrywanego zjawiska i pochodności słowotwórczej, oczywiście, pod względem semantycznym, a nie formalnym, podkreślona została w nazwie *derywacja semantyczna*²³. W przytoczonej wyżej grupie terminów i ich ekwiwalentów inności słowiańskich, opatrzonych indeksem 1 i 2, wieloznaczność można przedstawić jako typową relację semantyczną: ‘zjawisko’ – ‘dyscyplina językoznawcza zajmująca się badaniem danego zjawiska’²⁴.

Podobny typ relacji semantycznej ujawnia się w grupie wieloznacznych terminów z zakresu onomastyki, np. w polszczyźnie: *antroponimia 1* ‘zbiór, zasób, system antroponimów’ i *antroponimia 2* ‘dyscyplina językoznawcza zajmująca się badaniem antroponimów’, *toponimia 1* ‘zbiór, zasób, system toponimów’ i *toponimia 2* ‘dyscyplina językoznawcza zajmująca się badaniem toponimów’, *zoonimia 1* ‘zbiór, zasób, system zoonimów’ i *zoonimia 2* ‘dyscyplina językoznawcza zajmująca się badaniem zoonimów’, *chrematonimia 1* ‘zbiór, zasób, system chrematonimów’ i *chrematonimia 2* ‘dyscyplina językoznawcza zajmująca się badaniem chrematonimów’ itp.

Wraz z rozwojem onomastyki w poszczególnych krajach słowiańskich udoskonalane są funkcjonalnie systemy terminologiczne tej dyscypliny, usuwana jest na peryferie wieloznaczność terminów. Pomocne okazuje się w tym zakresie słowotwórstwo dzięki zastosowaniu mechanizmów słowotwórczych. Proces eliminowania wieloznaczności przebiega niejednakowo w językach słowiańskich, najbardziej zaawansowany jest w pracach czeskich i słowackich onomastów, także w pracach

²² Szczegółowo o tym pisze J.D. Apresjan, zob. Ю.Д. Апресян, *Лексическая семантика. Семантические средства языка*, Москва 1974, s. 187–193.

²³ Termin D.N. Šmeleva (Д.Н. Шмелев, *Лексика современного русского языка*, Москва 1968) rozpropagowany został w językoznawstwie dzięki pracom J.D. Апресяна. Inne rozumienie terminu *derywacja semantyczna* znaleźć można m.in. w pracach słowotwórczych o podstawach semantycznych S. Karolaka i V. Maldjiewej.

²⁴ Por. Ю.Д. Апресян, *Лексическая семантика...*, s. 203.

rosyjskich (rosyjskojęzycznych), natomiast w polskim piśmiennictwie z zakresu onomastyki aktywność jego nastąpiła stosunkowo później. Przyczyny należy upatrywać w niejednakowym poziomie rozwoju dyscypliny onomastycznej w krajach słowiańskich. Na przykład w polskiej literaturze językoznawczej terminy *antroponimia*, *toponimia* jeszcze do niedawna były stosowane w dwóch znaczeniach: *antroponimia 1* ‘zbiór, zasób, system antroponimów’, *antroponimia 2* ‘dyscyplina / dział językoznawstwa zajmujący się badaniem antroponimów’, *toponimia 1* ‘zbiór, zasób, system toponimów’ i *toponimia 2* ‘dyscyplina / dział językoznawstwa zajmujący się badaniem toponimów’. Znaczenia te i relację między nimi można uogólnić w postaci typowej relacji ‘zjawisko’ – ‘dyscyplina’, którą przedstawiliśmy wcześniej.

Rozwój badań onomastycznych i ich specjalizacja spowodowały potrzebę stosowania jednoznacznych terminów oznaczających poddyscypliny onomastyki. Dzięki zastosowaniu międzynarodowego modelu dostosowanego do możliwości systemu słowotwórczego polszczyzny szeroko upowszechnia się nowy typ struktur, dający początek serii nazw dyscyplin: *antroponomastyka*, *toponomastyka*. Na ich wzór powstały neologizmy *zoonomastyka* ‘dyscyplina / dział językoznawstwa (onomastyki) zajmujący się badaniem zoonimów’, *chrematonomastyka* ‘dyscyplina / dział językoznawstwa (onomastyki) zajmujący się badaniem chrematonimów’ i in. Pierwowzór stanowił zapewne model czeski o charakterze międzynarodowym, por. *antroponomastika*, *toponomastika*, *chrematonomastika*, a także niemiecki, por. *Anthroponomastik*, *Toponomastik* itp., bowiem w tych środowiskach naukowych od ponad pół wieku odnotowuje się najwyższy poziom rozwoju onomastyki i jej terminologii²⁵.

Inne rozwiązanie przyjęto w rosyjskich pracach językoznawczych, gdzie nazwy dyscyplin onomastycznych tworzone są wyspecjalizowanym sufiksem *-ика*: *антропонимика*, *топонимика*, *зоонимика*, *ойконимика*, *оронимика*, *этнонимика* itp. Nadal jednak niektórzy językoznawcy używają terminów: *антропонимика*, *топонимика*, *зоонимика* itp. zarówno w znaczeniu ‘zbiór, zasób, system’, jak i ‘dyscyplina onomastyczna’. W przypadku terminów wieloznacznych ustanowienie modeli i wykładników słowotwórczych, wyspecjalizowanych pod względem semantycznym, zależy od wielu czynników, m.in. preferencji słowotwórczych, uzusu, tradycji nazewniczych, i może być procesem długotrwałym.

Badając zjawisko wieloznaczności terminów, nie można pominąć innej spójnej, choć mniej licznej, grupy terminów, jak: pol. *nominacja*, *kompozycja*, *kontaminacja* itp. W ich strukturze semantycznej wyodrębnić można węższe znaczenia, które tworzą między sobą relację: ‘czynność / proces’ – ‘rezultat czynności / procesu’, np. pol. *kontaminacja 1* ‘zespolenie różnych elementów językowych w jeden wyraz lub związek wyrazowy’ – *kontaminacja 2* ‘wyraz lub związek wyrazowy po-

²⁵ V. Šmilauer, *Úvod do toponomastiky*, Praha 1963.

wstały z zespolenia różnych elementów językowych²⁶ i innośłowiańskie ekwiwalenty: cz. *kontaminace 1 – kontaminace 2*, ros. *контаминация 1 – контаминация 2* itd. W przypadku wieloznacznego terminu pol. *kompozycja* i jego ekwiwalentów w ich strukturze semantycznej wyodrębniamy znaczenia ‘sposób czynności, procesu’ – ‘rezultat’. Terminy opatrzone indeksem 2 interpretujemy jako mające wtórne znaczenie, rezultatywne, jednostek terminologicznych opatrzonych indeksem 1 i posiadających znaczenie abstrakcyjnej czynności / procesu lub sposobowe.

O tym, że mamy do czynienia ze zjawiskiem semantycznym, a nie słowotwórczym, chociaż bliskim słowotwórstwu, świadczy np. termin *nominacja 2* i cała klasa innojęzycznych ekwiwalentów, które powstały bez użycia środków słowotwórczych. Znaczenie rezultatu wyrażone jest czysto semantycznie, a podstawę derywacji semantycznej stanowi termin nie pochodny synchronicznie (*nominacja 1* ‘procedura’). Zilustrujemy to przykładami z tekstów językoznawczych: *nominacja 1* („Reinterpretacja danego obiektu w jego nowej nazwie polega na wyborze w procesie *nominacji* innej cechy tego obiektu niż ta, którą wyzyskano w już istniejącej nazwie”²⁷) – *nominacja 2* („powstawanie złożonych *nominacji* znajduje się na granicy dwóch systemów – zaimkowego i słów pełnoznaczących...”²⁸), cz. *nominace 1* („*Nominace* některých nových výrobků z oblasti elektrotechniky v ruské a české reklamě”²⁹) – *nominace 2*³⁰, ros. *номинация 1* („Акты прямой (прямой) номинации при помощи слов примечательны тем, что они одновременно сопряжены с опредмечением человеком объективного мира...”³¹) – *номинация 2* („*номинации*, соответствующие нормам и осознаваемые в момент своего

²⁶ Znaczenie rezultatywne tej struktury podaje K. Waszakowa w pracy *Słowotwórstwo współczesnego języka polskiego. Rzeczowniki sufiksalne obce*, Warszawa 1994, s. 84.

²⁷ T. Smółkowa, *Neologizmy we współczesnej leksyce polskiej*, Kraków 2001.

²⁸ О.П. Ермакова, *Взаимодействие двух систем частей речи (местоименной и знаменательной) при образовании составных номинаций*, [w:] *Słowotwórstwo a inne sposoby nominacji. Materiały z 4 konferencji Komisji Słowotwórstwa Międzynarodowym Komitecie Słowistów. Katowice 27–29 września 2000 r.*, red. tomu K. Kleszczowa, Ludwig Selimski, Katowice 2000, s. 152.

²⁹ L. Mrovčová, *Nominace některých výrobků z oblasti elektroniky v ruské a české reklamě*, „Opera Academiae Paedagogicae Liberecensis”, 2005, vol. 3, s. 48.

³⁰ W przebadanym dotychczas czeskim materiale językowym brak potwierdzenia użycia tego terminu w znaczeniu rezultatu procesu nadawania nazwy. Być może, w związku z ustabilizowaniem się czeskiego terminu *pojmenování 2* w danym znaczeniu nie istnieje potrzeba do alternatywnego wyrażania tych samych treści. Przywołane przykłady ekwiwalentnych terminów w języku polskim i rosyjskim są nielicznie potwierdzone (por. synonimiczne nazwy w języku polskim: *nazwa, jednostka nominatywna / jednostka nominacyjna* i w rosyjskim: *название, именованье, наименованье, номинативная единица*) w porównaniu z szeroko stosowanymi terminami w znaczeniu procesu nadawania nazwy.

³¹ А.А. Уфимцева, *Лексическая номинация (первичная нейтральная)*, [w:] *Языковая номинация. (Виды наименования)*, Москва 1977, s. 8.

появления в речи как новые”³²; „Взаимодействие двух систем частей речи (местоименной и знаменательной) при образовании составных *номинаций*”³³).

Rozpatrywana kombinacja znaczeń (pojęć) ‘procedura’ – ‘rezultat’ może być również realizowana za pomocą środków słowotwórczych. Przykład derywacji słowotwórczej stanowi operacja semantyczna i formalna w parach ros. *аббревиация 1 – аббревиатура* i cz. *abreviace 1 – abrewiatura*. Bliskość derywacji semantycznej i derywacji słowotwórczej jest szczególnie widoczna w klasach ekwiwalentów, u podstaw których w istocie leżą klasy pojęć, np. ‘rezultat procesu, procedury’, por. ros. *аббревиатура*, cz. *abrewiatura, abreviace 2*; pol. *nomincja 2, nazwa*, ros. *название, номинация 2, наименование*.

Rozwiązanie w języku informacyjno-wyszukiwawczym problemu wieloznaczności terminów wzbogaci strukturę semantyczną tego języka o relacje kojarzeniowe, zapewniające większą efektywność, zwłaszcza trafność, wyszukiwania informacji w systemie *iSybislaw*. Relacyjna organizacja paradygmatyki języka słów kluczowych uwzględniająca relację synonimiczności i wyspecyfikowane niektóre relacje kojarzeniowe ma istotną wartość wyszukiwawczą.

SUMMARY

Linguistic Problems of Information Retrieval Optimization in *iSybislaw* System

During the past dozen or so years there have occurred large changes in the information society and computer technologies. An increase in the information needs of the users of information retrieval systems (IRS) has been observed. The methodology of information processing has changed: user unfriendly and complicated information retrieval languages (IRL) have been abandoned and the use of natural languages in their metalinguistic function has been reestablished with the use of keywords being its symptom.

In the commonly used IRS the structure of the keyword language is flat, the keywords are nonsubordinate. The keyword language, used in *iSybislaw* (in the bibliographical information retrieval system of the world Slavic linguistics to be more exact), **is non-standard in its character**. It differs from other keyword languages with its paradigmatic semantic structure. The methodology of processing multilingual vocabulary takes into consideration the **relations of equivalence and polysemy**. Establishing the equivalence classes (consisting of terminological equivalents and other type equivalents including: personal, geographic and corporative names among others) in an IRS allows the retrieval of relevant information by means of any given search term: in any language and alphabet. A certain type

³² А.Ф. Журавлев, *Технические возможности русского языка в области предметной номинации*, [w:] *Способы номинации в современном русском языке*, Москва 1982, s. 47.

³³ О.П. Ермакова, *Взаимодействие двух систем...*, s. 147.

of preference will be introduced in the case of equivalence in the form of descriptors (and non-descriptors), whereas in the case of polysemy a system of (meta)metalinguistic appositions will be used. The use of a descriptor system in an IRS such as iSybislaw ensures the completeness of retrieved information, resolving the issue of polysemy on the other hand allows for a more effective information retrieval measured by the precision and relevancy of retrieved data.