
УДК 047.31+811.6

Павел Ковальський, Якуб Банасяк

Інститут славістики Польської Академії наук, м. Варшава

КЛЮЧОВІ СЛОВА ТА КЛАСИ ЕКВІВАЛЕНТНОСТІ В СИСТЕМІ ISYBISLAW ЯК ЗНАРЯДДЯ ТА ПРЕДМЕТ ДОСЛІДЖЕНЬ

У статті обговорено окремі етапи побудови класів еквівалентності у межах багатомовного масиву ключових слів у системі бібліографії світового славістичного мовознавства iSybislaw. Окреслено найважливіші теоретичні та практичні проблеми формування таких класів, а також методологічні підходи, що їх використовують розробники системи. Розглянуто питання, пов'язані з інформаційними потребами користувачів системи, зокрема пропозиції, спрямовані на підвищення пошукової прагматики бази iSybislaw.

Ключові слова: інформаційно-пошукова система, інформаційно-пошукова мова, ключове слово, мовознавча термінологія, iSybislaw.¹

0. Популяризація в науковому середовищі Інтернету як комунікаційного простору та употужнення протягом останніх років комп'ютеризації наукової діяльності викликають динамічні зміни у функціонуванні традиційної бібліографії, яка дедалі більше перетворюється на сучасні інформаційно-пошукові системи (далі — ІПС). Завданням таких систем є удоступнення користувачам релевантної інформації в ефективний і зручний («дружній») спосіб. Як зауважує Е. Артович, «сучасний розвиток інформаційних сервісів, доступних онлайн через телеінформаційні мережі, ставить перед розробниками й експлуатувальниками

¹ Статтю підготовлено в рамках міжнародного наукового проекту «Онлайн-бібліографія світового славістичного мовознавства», який Інститут славістики Польської Академії наук виконує разом з Інститутом української мови НАН України за Угодою про наукове співробітництво між ПАН і НАНУ. З польської мови текст переклали Є.А. Карпіловська та Ю.В. Романюк.

ІПС, які послуговуються з дедалі складнішого програмного забезпечення, нові вимоги до лінгвістичної основи систем. Здійснений нещодавно аналіз залежності між технологією удоступнення інформації та способами подання знань у традиційних і сучасних, інтерактивних ІПС наочно продемонстрував, що розвиток інтерактивних способів пошуку інформації й системного програмного забезпечення для їх застосування залишається в прямій кореляції з оснащенням системи дедалі потужнішим, належним чином структурованим знанням про природну мову та інформаційно-пошукові мови, представленим окремо від знання про дійсність»² [6: 61]. Перенесення бібліографії в цифровий простір, відповідне перетворення форми її подання, або її ремедіація, яка полягає у представленні одного засобу інформації (медіа) в іншому³ [11: 45], дозволяє особам, які готують бібліографічні описи (предметні та формальні описи документів), застосовувати нові методи та інструменти в зовсім іншому середовищі. Такі інструменти дають можливість створювати нові пункти доступу до інформації, уміщеної в документах. Вони уможливають інші, ніж у традиційній паперовій бібліографії, організацію інформації та її подання, а також швидку, просту та ефективну доступність такої інформації. Усі ці заходи спрямовано на задоволення інформаційних потреб користувачів ІПС, тобто вони на перший план виводять пошукову прагматику, практичні проблеми пошуку.

0.1. У статті маємо намір представити міркування стосовно функціонування одиниць мови ключових слів у бібліографічній базі даних світового славистичного мовознавства iSybislaw, доступній за інтернет-адресою www.isybislaw.ispan.waw.pl. Ми також хочемо показати окремі етапи побудови класів внутрішньо- та міжмовної еквівалентності у межах багатомовного масиву ключових слів бази iSybislaw. Ця база даних є інформаційно-пошуковою системою, яку було удоступнено користувачам у мережі Інтернет у 2007 р. Ідею її створення вперше було презентовано ширшому грону науковців на XII Міжнародному з'їзді славистів у Кракові в 1998 р. [9: 41—47]. Над розбудовою інформаційної системи, яка на сьогодні налічує понад 23 тис. повних описів документів (записів), працює міжнародний колектив мовознавців-славистів з різних слов'янських країн (Польщі, України, Росії, Білорусі, Чехії, Словаччини, Хорватії, Македонії, раніше — з Сербії та Словенії), а також з країн за межами Славії (Німеччини, Японії, раніше — з Франції та Естонії)⁴. Осмислюючи завдання, пов'язані з ІПС, та описуючи їхню прагматику, практичний аспект, слід взяти до уваги дві перспективи розгляду цієї проблеми. Перша перспектива, сфокусована на користувача такої системи, це — зовнішня перспектива. Ко-

² Українські переклади цитат погоджено з авторами статті (*перекл.*).

³ Див. також на цю тему працю М. Мариля [15], який розглядає ремедіацію в контексті функціонування сучасної літератури.

⁴ Докладніше про систему iSybislaw див. праці: З. Рудник-Карватової, З. Мікоса, Б. Бояр [20], М. Пшастек-Самокової [17], З. Рудник-Карватової, Я. Банасяка, З. Мікоса [18], З. Рудник-Карватової [5].

ристувач найчастіше не повинен знати процедур створення структури системи, складних технічних питань, пов'язаних з організацією знань у цій системі. Друга перспектива — внутрішньо-зовнішня, яку також можна назвати перспективою індексатора-користувача. Індексатор (особа, яка створює пункти доступу до знань у конкретній системі) мусить мати знання про структуру всієї системи, брати активну участь у розбудові інформаційно-пошукової мови і водночас використовувати функціональні можливості системи. У цій статті ми представляємо проблеми з урахуванням обох перспектив; методологічні міркування споряджаємо прикладами, дібраними з бази iSybislaw.

1. Одним із найважливіших інформаційно-пошукових інструментів, застосованих у системі iSybislaw, є мова ключових слів. Ця мова в iSybislaw виконує дві основні функції: метаінформаційну та пошукову. Метаінформаційна функція мови ключових слів полягає у відображенні змісту документів, розміщених у базі, натомість її пошукова функція уможливорює для користувачів пошук релевантної інформації в усьому масиві документів. Для того, щоб мова ключових слів була інструментом для пошуку точної і повної інформації, яка задовольнить інформаційні потреби користувачів, належить з великою увагою розширювати, впорядковувати й вибудовувати її лексичний ресурс (масив ключових слів). Надто загальні характеристики пошуку можуть спричинити інформаційні лакуни, так звані пошукові «білі плями», тобто порожні простори, не заповнені потенційні пункти доступу до інформації, результатом чого буде брак отримання відповідної інформації. Пропуск під час індексування потрібних користувачеві проблем і тем спричинюватиме втрату інформації (інформаційну «тишу»). Непродуктивним в інформаційному процесі є також узагальнення інформації під час індексування документа. Звертає на себе увагу, наприклад, відсутність кореляції між частотою вибирання в інформаційно-пошуковій системі класифікації *składnia zdanie proste / zdanie złożone* (синтаксис *просте речення/складне речення*) й додаванням в описі документа багатозначного ключового слова *zdanie* (речення). Так само, утім, занадто глибоке індексування, надлишковість інформації (іноді корисна з погляду як розробника, так і користувача системи, наприклад, у працях, пов'язаних з перетворенням інформації, її конденсацією в процесі анотування, реферування та ін.), можуть призвести до непотрібної деталізації змістового опису й спричинити інформаційний «шум», а через це — й проблеми з пошуком релевантної інформації. Як бачимо, однією з найважливіших ознак мови ключових слів повинна бути пертиненція, або відповідність надаваної інформації інформаційним потребам користувачів. Указані проблеми доводять, що конструювання і розбудова мови ключових слів є завданням, яке вимагає вирішення ряду питань методологічного та практичного (технічного) характеру. Вони засвідчують також необхідність роботи з системою міждисциплінарного колективу, який об'єднує фахівців з різних галузей мовознавства, наукової інформації та інформатики. У процесі накопичування та впорядкування

ключових слів слід оперативно розв'язувати проблеми багатозначності, синонімії і варіантності слів природної мови, або термінів, що становлять ключові слова. На початковому етапі роботи — в площині однієї мови, на її наступних етапах — у площині двох або більше мов. Розробники системи iSybislaw прийняли новаторське методологічне розв'язання тих проблем, які спрямовані на створення у межах всього масиву ключових слів класів внутрішньо- та міжмовної еквівалентності [9: 45—47]. До складу одного класу входять різномовні одиниці, які виконують ту саму денотативну функцію. Їхній безпосередній денотат становить сукупність документів, до яких вони відсилають, а опосередкований — елементи позадокументної дійсності [10: 48—49; пор. також нижче].

1.1. Мова ключових слів iSybislaw є паранатуральною інформаційно-пошуковою мовою, яка відображає інформацію, уміщену в документах [7]. Як уже було підкреслено вище, її використовують у метайнформаційній функції та у функції пошуку інформації в інформаційному масиві (масиві документів) системи iSybislaw. Одиниці мови ключових слів в iSybislaw, як правило, сумірні з мовознавчими термінами, відповідають їм. Слід, проте, підкреслити різницю між двома масивами. Терміни позначають та уточнюють елементи мовної (позатекстової) дійсності, натомість ключові слова безпосередньо позначають масиви документів, уміщених в системі, і безпосередньо уточнюють певні їх угруповання, сукупності ознак документів. Семантику ключових слів не подано користувачеві експліцитно, однак припущено, що у випадку ключових слів (особливо у мовах ключових слів паранатурального типу), які становлять трансформації термінів, користувач інтуїтивно домислить і зрозуміє змістовий обсяг такої одиниці на основі розуміння змісту терміна. Не завжди, однак, слід спиратися на значення термінологічних одиниць, оскільки семантику ключових слів визначає штучна парадигматика, яка може відрізнитися від парадигматики певної термінологічної системи [24]. Це відбувається, зокрема, у разі виділення у межах термінології дуже деталізованих значень/уживань термінів, коли доцільним є спрощення, мотивоване інформаційними потребами користувача бази та його компетенцією. Візьмімо для прикладу ключові для семантичного синтаксису терміни: пол.: *wyrażenie predykatywne, predykator, predykat₂, predykat składniowy*; болг.: *предикативен израз, предикатор* (рідковживане), *предикат₂, синтактичен предикат*; рос.: *предикативное выражение, предикат₂, предикатор* (рідковживане), *синтаксический предикат*. Крім очевидних ідіолектних преференцій, виявляється також різниця в значенні, якої надають терміну *predykator* у певних контекстах [8]. Цей термін тоді не є рівнозначним терміну *wyrażenie predykatywne*. По відношенню до вербально-субстантивних аналітичних конструкцій *predykator* є частиною *предикативного виразу*. Інакше кажучи, *предикативний вираз* у вербально-субстантивних аналітичних термінах складається з *предикатора* і *вербалізатора* [пор. п. 2.3.]. Такі нюанси слід було б узяти до уваги у термінологічному банку. В інформаційно-пошуковій системі, адресованій широкій аудиторії користувачів, вони нерелевантні.

1.2. Розробники iSybislaw для накопичення термінології використовують корпус текстів, репрезентативних для сучасної літератури з певної галузі мовознавчої науки, та їхні бібліографічні описи, а також доповнюють масив спеціалізованої лексики на основі авторитетних академічних граматики, мовознавчих енциклопедій, тезаурусів і словників. Реалізовано індукційно-дедукційний метод, застосований для збирання лексичного масиву «Словника ключових слів славистичного мовознавства» [20: 6—8]. Крім мовознавчої славистичної термінології (галузевих термінів), одиницями мови ключових слів системи iSybislaw виступають також загальнонаукові, загальномовознавчі терміни. До цієї групи входять терміноодиниці, які стосуються теорії мови і загального мовознавства, а також контрастивного слов'янсько-неслов'янського мовознавства. Численну групу у масиві ключових слів системи iSybislaw становлять також нетермінологічні одиниці. Вони базуються на словах природної мови та відображають елементи змісту документів, важливі з погляду пошукової прагматики. До таких одиниць мови ключових слів належать: власні імена: *Mickiewicz Adam*, *Słowacki Juliusz*, також прізвища мовознавців, науковий доробок яких є предметом досліджень, обговорюваних у документах: *Sławski Franciszek*, географічні: *Kijów*, *Małopolska*, *Ukraina* та корпоративні назви (закладів, установ): *Polska Akademia Nauk*, *Instytut Sławistyki*, *Uniwersytet Łódzki*. Численними є також одиниці загальнонаціональної мови, які є предметом досліджень, як, наприклад, лексеми, фразеологізми, морфеми тощо: *dusza* (душа), *kwiatek* (квітка), *serce* (серце), *słońce* (сонце), *stawiać zamki na lodzie* (букв. будувати замки на льоду), **-ować**, **-owo**; слова, що мають високу частоту в масиві документів, наприклад, у польській мові назви кольорів: *żółty* (жовтий), *biały* (білий), частин тіла: *ręka* (рука), *głowa* (голова), наукових зібрань: *kongres* (конгрес, з'їзд), *konferencja* (конференція) тощо.

2. Як уже було підкреслено вище, під час проектування мови ключових слів істотними проблемами, які слід оперативно розв'язувати, є багатозначність, синонімія та варіантність терміноодиниць, використовуваних у функції ключових слів системи iSybislaw. Це особливо важливо для різномовних користувачів та індексаторів, які потребують доступу до дієвого, ефективного інформаційно-пошукового інструменту. Творення класів еквівалентності відбувається у межах одиниць мови формального опису, а також мови предметного опису, який відображає зміст документа [18]. Зокрема, у формальній мові подано різні варіанти імені та прізвища автора, напр.: *Ожегов Сергей Иванович* = *Ožegov Sergej Ivanovič*, *Shevelov George Y.* = *Ševelev Jurij* = *Шевелёв Юрий Владимирович* = *Шевельов Юрій* (рис.1).

У мові фактографічній (мові предметного опису) подано синонімічні одиниці у межах однієї мови, напр., пол. *struktura tematyczno-rematyczna* = *funkcjonalna perspektywa zdania* = *aktualne rozczłonkowanie zdania*; і тотожні за значенням у багатьох мовах та алфавітах: чес. *aktuální větné členění* = *funkční větná perspektiva* = *větná perspektiva*; словен. *funkcionalna stavčna perspektiva* = *tematsko-rematska struktura*; рос. *актуальное членение предложения* = *коммуникативная перспектива высказывания* = *функциональная перспектива*

Ožegov	Sergej Ivanovič
Ожегов	Сергей Иванович
Karpilovs'ka	Ėvgenija Anatoliivna
Карпіловська	Євгенія Анатоліївна
Shevelov	George Y.
Ševelev	Jurij
Ševel'ov	Jurij
Szerech	Jurij
Szewelow	Jurij
Шевелёв	Юрий Владимирович
Шевельов	Юрій

Рис. 1. Фрагмент масиву ключових слів системи iSybislaw з варіантами власних імен

высказывания; болг. актуално членение на изречението = функционална изреченска перспектива = темо-ремна структура = тематично-рематична структура, укр. актуальне членування речення (рис. 2).

Усі вищезгадані одиниці утворюють один клас, і кожна одиниця відсилає до того самого масиву документів у системі iSybislaw. Завдяки цьому користувач, подаючи будь-який термін з певного класу, незалежно від мови й алфавіту, отримує той самий результат запиту, ту саму кінцеву інформацію, а також повноту інформації разом з економією часу, оскільки пошук ресурсів відбувається за допомогою ключових слів, створених на

struktura tematyczno-rematyczna	polски (pol)
aktualne rozczlonkowanie zdania	polски (pol)
funkcjonalna perspektywa zdania	polски (pol)
актуальное членение предложения	rosyjsки (rus)
коммуникативная перспектива высказывания	rosyjsки (rus)
функциональная перспектива высказывания	rosyjsки (rus)
актуално членение на изречението	bułgarsки (bul)
тематично-рематична структура	bułgarsки (bul)
темо-ремна структура	bułgarsки (bul)
функционална изреченска перспектива	bułgarsки (bul)
aktuální větné členění	czesки (cze)
funkční větná perspektiva	czesки (cze)
větná perspektiva	czesки (cze)
funkcjonalna stavčna perspektiva	słoweński (slv)
tematsko-rematska struktura	słoweński (slv)
актуальне членування речення	ukraiński (ukr)

Рис. 2. Клас еквівалентності термінів з дескриптором пол. *struktura tematyczno-rematyczna*

основі різних природних мов. Найголовнішою метою є уможливлення далі ефективного міжмовного пошуку інформації (МПІ), який «можна описати на абстрактному рівні як завдання, що полягає в пошуку документів у межах різних мов. У певному сенсі, МПІ становить граничний випадок проблеми незбігу лексики (англ. *vocabulary mismatch problem*), тобто явища, коли лексика запиту користувача і те, що міститься у релевантному документі, можуть істотно відрізнятись» [22: 1].

2.1. Перший етап конструювання класів еквівалентності полягає у збиранні термінологічних одиниць в одній мові, а потім їх упорядкування, передусім усунення багатозначності. Усунення багатозначності термінології, ужитої у функції ключових слів, має вирішальне значення, оскільки без її розв'язання користувач не має можливості доступу до релевантної інформації вже у межах однієї мови, а наслідки «перенесення» такої картини з однієї мови на інші стають непередбачуваними. Так, наприклад, багатозначному польському терміну *język* («система знаків» vs «орган мовлення») відповідають в українській та англійській мовах по два терміни, а саме: *мова, language*; *язик, tongue*. В англійській ситуацію ще ускладнює той факт, що термін *tongue*, хоча і досить рідко, уживають також у значенні «система знаків», пор. складений термін *mother tongue* «материнська мова». В українській мові термін *язик* також уживали в значенні «система знаків» у старих джерелах (староукраїнських і літературних творах новою українською літературною мовою). Проте його відсутність у складі певних аналітичних номінацій [пор. п. 2.3.] може слугувати підтвердженням того, що ця форма вже вийшла з ужитку.

Крім багатозначності цього типу, у термінології також трапляються явища більш регулярні, які можна виявити та описати в категоріях певних лексичних функцій, пор., наприклад, «процес→результат»: пол. *nominacja 1* (процес) → *nominacja 2* (результат), *nominalizacja 1* (процес) → *nominalizacja 2* (результат), *przekład 1* (процес) → *przekład 2* (результат), *zapożyczenie 1* (процес) → *zapożyczenie 2* (результат); укр. *номінація 1* (процес) → *номінація 2* (результат), *номіналізація 1* (процес) → *номіналізація 2* (результат), *переклад 1* (процес) → *переклад 2* (результат), *запозичення 1* (процес) → *запозичення 2* (результат). Хоча і тут мовці можуть уживати різні терміни для відображення різних значень, пор. пол. *zapożyczenie* (процес) — *zapożyczenie* (результат) та укр. *запозичування* (процес) — *запозичення* (результат). Отже, мовознавці намагаються розрізнявати такі значення, а запровадження та вживання різних термінів сприяє впорядкуванню цієї термінологічної системи. Багатозначні терміни на зразок: *номінація 1* (процес) → *номінація 2* (результат) слід враховувати з методологічного погляду. Проте їхня вага з погляду доступу до релевантного інформаційного масиву менша, ніж у пар типу пол. *język 1* vs *język 2*. Це відбувається завдяки тому, що такі терміни об'єднують семантична похідність, яка віддзеркалює близький зв'язок у позамовній і позадокументній реальності між вказаними процесами та їхніми результатами. Інформація, релевантна для перших, великою мірою є релевантною і для других.

Одним з цікавих прикладів може бути проектування класу термінів, пов'язаних з універбацією. У польській мовознавчій термінології усталився інтернаціональний термін *uniwerbizacja*, хоча вперше його було вжито в польській мові у формі *uniwerbalizacja* [21]. Однак варіантна форма *uniwerbalizacja* не прижилася в польському мовознавстві й узагалі її вживання можна визначити як одиничне. З огляду на це, а також на брак цього терміна у більшій кількості текстів, його не введено до масиву ключових слів системи iSybislaw. Натомість, у словенському мовознавстві функціонують два варіанти інтернаціонального терміна *univerbizacija* і *univerbacija* (так само й в інших слов'янських мовах, пор. укр. *універбація* і *універбізація*) і питомий термін *poenobesedenje*. Дефініційні обсяги питомого й інтернаціонального термінів різняться. Терміном *poenobesedenje* охоплено дуже різномірні словотворчі процеси. Як пише Й. Топоришич: «...*poenobesedenje* це постання одного слова з двох чи більше слів, наприклад, *seveda* «відомо, очевидно» з *se ve da* «відомо, що», а також заміна дво- чи кількаслівної назви одним словом за участі словотворчих механізмів, наприклад: *nedovršni glagol* «дієслово недоконаного виду» — *nedovršnik, cesta za avte* «дорога для автомобілів» — *avtocesta, drugi/nišji predsednik* «заступник голови — *podpredsednik*» [23: 187]. Отже, термін *poenobesedenje* можна потрактувати як родовий відносно інтернаціональних термінів *univerbizacija/univerbacija*. Проте у багатьох працях словенських мовознавців досить уживані обидва терміни. Уведення до переліку ключових слів двох цих одиниць як окремих класів викликало б інформаційну надлишковість і могло б вплинути на отримання неповної, недостатньої кількості документів, присвячених широко потрактованим завданням універбізації у словенських працях. З метою підвищення пошукової прагматики і створення системи, «дружньої» для користувача, усі три словенські одиниці: *poenobesedenje*, *univerbizacija* й *univerbacija* слід подати в одному класі внутрішньомовної еквівалентності (у межах однієї мови). Далі такий одномовний клас еквівалентності має бути пов'язаний з іншомовними класами, які відображають те саме семантичне поле «універбація» (рис. 3 і 4).

Таким чином, класи еквівалентності показують, хоч і спрощено, місце окремих термінів, ужитих у функції ключових слів, у взаємовідношеннях і створюють образ термінологічної системи певної мови (докладніше про це див. у працях: Є. Карпіловської [2], П. Ковальського [13], В. Кульпіної і В. Татарінова [3], І. Лучкув [14], О. Остапчук [16],

<i>poenobesedenje</i>	<i>sloweński (slv)</i>
<i>univerbacija</i>	<i>sloweński (slv)</i>
<i>univerbizacija</i>	<i>sloweński (slv)</i>

Рис. 3. Клас внутрішньомовної еквівалентності семантичного поля «універбація» (словенська мова)

uniwerbizacja	polski (pol)
универбація	rosyjski (rus)
univerbizace	czeski (cze)
univerbizácia	słowacki (slo)
poenobesedenje	słoweński (slv)
univerbacija	słoweński (slv)
univerbizacija	słoweński (slv)
універбація	ukraiński (ukr)
універбізація	ukraiński (ukr)

Рис. 4. Клас міжмовної еквівалентності семантичного поля «універбація»

Ю. Романюк [4], Є. Волкової [1]. Як бачимо, упорядкування термінології, вжитої у функції ключових слів, становить важливий етап створення обговорюваних класів.

2.2. Іншою процедурно-методологічною проблемою, яка постає під час праці над ІПМ, є нерівномірна розбудова масиву ключових слів в окремих мовах (рис. 5 і 6).

З огляду на те, що працю над системою iSybislaw зніціювало польське славістичне середовище, польська мовознавча термінологія становить найбільшу сукупність одиниць мови ключових слів. Понад 8 тис. ключових слів ґрунтовано на одиницях польської мови, з яких близько половини складають мовознавчі терміни.

Лексику польської мови ключових слів презентує виданий 1999 р. в паперовій версії і 2006 р. — у версії цифровій вищезгаданий «Словник ключових слів славістичного мовознавства» З.Е. Рудник-Карватової і



Рис. 5. Участь (у відсотках) слов'янських мов у масиві мови ключових слів системи iSybislaw

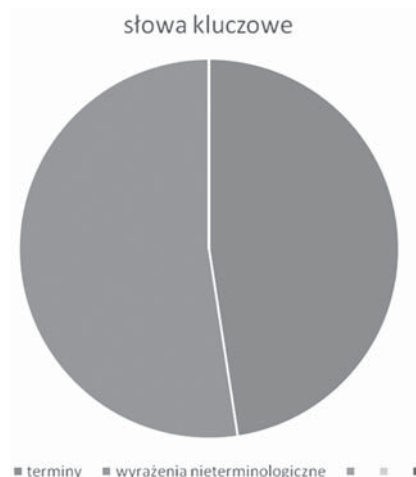


Рис. 6. Участь термінології у польському модулі мови ключових слів

Х. Карпінської, в якому зібрано близько 2500 одиниць, показових для польської мовознавчої славістики [18]. Для інших слов'янських мов та англійської мови, представлених у базі iSybislaw, поки що бракує таких словників, виданих у паперовій чи цифровій формі. Неповні словники ключових слів окремих слов'янських мов містяться в структурі системи iSybislaw в картотеках зразків (форматів) предметних описів документів. Отже, потрібно — згідно з прийнятою концепцією і графіком праці — постійно поповнювати й систематично розбудовувати такі сукупності для того, щоб забезпечити різномовним користувачам зручніше застосування мови ключових слів під час пошуку інформації.

2.2.1. У наявних працях під час розбудови термінологічних засобів, відображених у ключових словах певних слов'янських та англійської мов, функцію опису надано польській мові. Така ситуація мотивована передусім практичними міркуваннями, а також необхідністю вибору ключового слова у функції дескриптора, яке відповідало б терміну, доміантному у польській мові. Такий підхід був зручним з огляду на архітектуру інформаційно-пошукової системи iSybislaw. У Центрі славістичної наукової інформації Інституту славістики ПАН нині розпочато вивчення можливостей модернізації й удосконалення цієї системи, які більшою мірою унаслідок мовні преференції її користувачів. У першу чергу, заплановано уможливити одночасну зміну природної мови, на якій ґрунтовано ключові слова, висвічувані на рівні опису документа в його пошуковому образі, та зміну мови інтерфейсу користувача (англійської та російської). Зараз непольськомовні ключові слова можна знайти в повному обсязі, але побачити їх можна лише на рівні запису певного класу міжмовних еквівалентів. Такий спосіб є неоптимальним з погляду іншомовних користувачів бази. У процедурі опрацювання непольськомовної лексики необхідна також зміна, яка має полягати у виборі оптимального терміна (висвічуваного або дескриптора) у кожній мові.

2.3. Дуже важливою справою, що потребує осмислення, є також проблематика глибини індексування. Ідеться, між іншим, про питання, яке порушив Г. Фреге у своїй розвідці «Сенс і значення» [12]. Не можна не поцінувати великого внеску цього автора (як і багатьох інших логіків) у розвиток штучних мов. Розглядаючи справу в термінах Г. Фреге, можна твердити, що існують вислови, які мають різний сенс (екстенціонал), проте те саме значення (інтенціонал). Прекрасною ілюстрацією цього твердження є нетривіальне речення, яке окреслює тотожність виразів такого типу. Г. Фреге скористався з прикладу *Вечірня Зірка*, яка є *Вранішньою Зіркою*. Виявляється, як бачимо, що той самий екстенціонал може бути визначений у різний спосіб. На ґрунті мовознавчої термінології відмінності в сенсі термінів можуть бути істотними, пор., наприклад: пол. *narrativus/tryb narratywny*; болг. *преизказно наклонение* vs пол. *imperceptivus/tryb imperceptywny*; болг. *имперцептив* vs пол. *tryb nieświadka*; болг. *несвидетелско наклонение*. Протиставлені пари термінів різняться ознаками не факультативними, а облігаторними, суттєвими, які належало б відобразити в дефініції (*renarracja* vs *brak percepcji* vs *niebyscie świadkiem*). Відмінності цього типу можуть виявлятися також між різномовними термінами, пор., наприклад: пол. *język ojczysty* vs укр. *матірня мова/материна мова/материнська мова*; словен. *materni jezik/materinščina*. Як бачимо, проблеми, що постають на рівні природних мов у їх функції метамови (мови опису понять), мають становити точку відліку в аналізі інформаційно-пошукової мови паранатурального типу. Ідеться, власне, про те, щоб у якомога повніший спосіб спиратися на мовні знання користувача бази, однак водночас не переносити на рівень інформаційно-пошукової мови розбіжності, які виникли на ідіоматичному ґрунті.

Як видається, перенесення тонких розбіжностей такого роду на рівень інформаційно-пошукової мови може негативно позначитися не тільки на індексуванні документів, а й на повноті сукупності інформації, отриманої у відповідь на запит. Треба було б, очевидно, запровадити якусь додаткову систему відсилань, завдяки якій можна було б краще, хоча й трохи повільніше, індексувати, і ключові слова цього типу об'єднувати в класи еквівалентності.

3. Подані вище міркування стосовно побудови класів еквівалентності у масиві багатомовних ключових слів системи iSybislaw ведуть до кількох загальних висновків. Передусім встановлення внутрішньо- і міжмовних класів еквівалентності уможливорює користувачам доступ до знання через будь-яке слово, що входить до таких класів. Завдяки цьому зростає інформаційна цінність системи й відкриваються нові шляхи доступу до знання. Урахування в майбутньому, а також створення розбудованих класів еквівалентності, в яких будуть міститися одиниці з усіх слов'янських мов та з англійської мови, забезпечать ефективне використання бази iSybislaw для різномовних користувачів. Зокрема, істотним видається врахування англійської мови в структурі окремих класів еквівалентності, яке уможливить вільне послугування базою для тих користувачів, які не знають слов'янських мов.

БІБЛІОГРАФІЯ

1. Волкова Е. Проблемы белорусской лингвистической терминологии и вопрос соответствия ключевых слов (лингвистических терминов) в iSybislaw // *Slavia Orientalis*. — 2013. — № 62. — С. 285—292.
2. Карпіловська Є.А. Роль термінів-непрямих номінацій у тезаурусі інформаційно-пошукової системи славистичного мовознавства // *Studia z Filologii Polskiej i Słowiańskiej*. — 2014. — Т. 49. — С. 150—163. Доступ онлайн: <https://doi.org/10.11649/sfps.2014.017>
3. Кульпина В.Г., Татаринів В.А. Синонимия ключевых слов в системе славистической библиографии как эпистемологические рефлексы развития лингвистической терминологии. // *Studia z Filologii Polskiej i Słowiańskiej*. — 2014. — Т. 49. — С. 188—202. Доступ онлайн: <https://doi.org/10.11649/sfps.2014.017>
4. Романюк Ю.В. Українська авторська термінологія в системі iSybislaw (дієслово) // *Studia z Filologii Polskiej i Słowiańskiej*. — 2014. — Т. 49. — С. 245—258. Доступ онлайн: <https://doi.org/10.11649/sfps.2014.017>
5. Руднік-Карватова З.Е. Упорядкування термінології для потреб інформаційно-пошукової системи iSybislaw // *Українська мова*. — 2016. — № 3. — С. 105—117.
6. Artowicz E. Indeksy w semantyce leksykalnej Ju.D. Apresjana // *Zagadnienia Informatyki Naukowej*. — 2002. — № 1. — С. 61—78.
7. Babik W. Słowa kluczowe. — Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego, 2010.
8. Banasiak J. Synonymy and search synonymy in an IR system (on the basis of linguistic terminology and the iSybislaw system) // *Studia z Filologii Polskiej i Słowiańskiej*. — 2014. — Т. 49. — С. 176—187. Доступ онлайн: <https://doi.org/10.11649/sfps.2014.017>
9. Bojar B., Rudnik-Karwatowa Z. (1998), Koncepcja nowoczesnego systemu informacji slawistycznej // Dalewska-Greń H., Rusek J., Siatkowski J. (Ред.), *Językoznawstwo. Prace na XII Międzynarodowy Kongres Sławistów w Krakowie 1998*. — Warszawa: Energeia (Z Polskich Studiów Sławistycznych; seria 9), 1998. — С. 41—47.
10. Bojar B. Słownik encyklopedyczny informacji, języków i systemów informacyjno-wyszukiwawczych. — Warszawa: Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich. (Nauka, Dydaktyka, Praktyka; 56), 2002.
11. Bolter J.D., Grusin R. Remediation: understanding new media. — Cambridge: The MIT Press, 2000.
12. Frege G. Sens i znaczenie. // Frege G. *Pisma semantyczne*, przeł. B. Wólniewicz/ — Warszawa: Państwowe Wydawnictwo Naukowe, 1977. — С. 60—88.
13. Kowalski P. Synonimia w polskiej i słoweńskiej terminologii językoznawczej // *Toporišičeva obdobja*. — Ljubljana: Znanstvena založba Filozofske fakultete, 2016. — С. 99—106. Доступ онлайн: <http://centerslo.si/wp-content/uploads/2016/11/Kowalski.pdf>
14. Łuczaków I. Uwagi o ekwiwalencji słów kluczowych w systemie informacyjnym iSybislaw (na przykładzie wybranych terminów gramatycznych). // *Studia z Filologii Polskiej i Słowiańskiej*. — 2014. — Т. 49. — С. 203—218. Доступ онлайн: <https://doi.org/10.11649/sfps.2014.017>
15. Maryl M. *Życie literackie w sieci. Pisarze, instytucje i odbiorcy wobec przemian technologicznych*. — Warszawa: Instytut Badań Literackich, 2015.
16. Ostapczuk O. Język informacyjno-wyszukiwawczy i problem ekwiwalencji słów kluczowych językoznawstwa slawistycznego // *Zagadnienia Informatyki Naukowej*. — 2007. — № 2. — С. 54—63. Доступ онлайн: <http://bbc.uw.edu.pl/dlibra/doccontent?id=1795>
17. Przystek-Samokowa M. “Bibliografia językoznawstwa slawistycznego” i bibliograficzna baza danych światowego językoznawstwa slawistycznego iSybislaw jako przykład działań interdyscyplinarnych [online] // Kocójowa M. (Ред.), *Biblioteki, informacja, książka: badania interdyscyplinarne i praktyka w XXI wieku*. — Kraków: Instytut Informatyki Naukowej i Bibliotekoznawstwa Uniwersytetu Jagiellońskiego, 2010. — С. 447—452.
18. Rudnik-Karwatowa Z., Banasiak Z., Mikos Z. (2013), Językowe problemy optymalizacji wyszukiwania informacji w systemie iSybislaw // *Slavia Orientalis*. — 2013. — Т. 62. — № 4. — С. 631—646.
19. Rudnik-Karwatowa Z., Karpińska H. *Słownik słów kluczowych językoznawstwa slawistycznego*. — Warszawa: Sławistyczny Ośrodek Wydawniczy, 1999.

20. Rudnik-Karwatowa Z., Mikos Z., Bojar J. (2007), Nowoczesny system informacji slawistycznej: zadania, dotychczasowe wyniki i perspektywy // Zagadnienia Informacji Naukowej. — 2007. — № 2. — С. 19—40.
21. Siatkowska E. Syntetyczne i analityczne nazwy w języku czeskim i polskim // Prace Filologiczne. — 1964. — T. 18. — № 2. — С. 219—231.
22. Sorg P., Cimiano P. Cross-lingual Information Retrieval with Explicit Semantic Analysis // Working Notes for the CLEF 200 Workshop, 2008. Доступ онлайн: <http://www.aifb.kit.edu/web/Inproceedings1837>
23. Toporišič J. Enciklopedija Slovenskega Jezika. — Ljubljana : Cankarjeva založba, 1992.
24. Woźniak-Kasperek J. Języki informacyjne: między tradycją a nadzieją na cyfrową przyszłość // Przegląd biblioteczny. — 2010. — № 1. — С. 5—18.

Статтю отримано 30.10.2017

Paweł Kowalski, Jakub Banasiak

Institute of Slavic Studies of the PAS, Warsaw

KEYWORDS AND CLASSES OF EQUIVALENCE IN THE ISYBISLAW SYSTEM AS A TOOL AND SUBJECT OF STUDY

The paper discusses the main stages of creating (establishing) classes of equivalence within a multi-lingual keyword set in the bibliography of world Slavic linguistics iSybislaw which is available at www.isybislaw.ispan.waw.pl. This database is an information and retrieval system that was made available to users on the Internet in 2007. Its idea was first presented to the wider scientific community at the 12th Slavic International Congress in Cracow in 1998. Its main information retrieval tool is the keyword language that has two basic functions in iSybislaw: metainformational and retrieval function. The keyword metadata is based on mapping of the content of the documents in the database, while the retrieval feature of the language allows users to derive (obtain) relevant information from all documents. We present the most important theoretical and practical problems of creating such classes and methodological solutions used by the designers of the iSybislaw system. The establishment of classes of intra-linguistic and inter-lingual equivalence allows users to gain knowledge of any expression that constitutes a given class. The information value of the system is enriched and new access points to the knowledge presented in the system are created. In the future terminological units from all Slavic languages as well as English that are going to be included in the equivalence classes will allow efficient use of the system for multilingual users. It is particularly important to include English in the structure of individual equivalence classes, which will allow non-Slavic speakers to use the database easily.

Keywords: *information-retrieval systems, information-retrieval languages, keyword, linguistic terminology, iSybislaw.*