

А. Загнітко, д. філол. н., Г. Ситар, к. філол. н.,  
 І. Данилюк, І. Щукіна  
 Донецький національний університет (Донецьк)  
 УДК 811.161.2'344

**ПРИНЦИПИ І ЗАКОНОМІРНОСТІ УКЛАДАННЯ БАЗИ "УКРАЇНСЬКІ  
 ПРИЙМЕННИКИ ТА ЇХНІ ЕКВІВАЛЕНТИ"**

*The method of Ukrainian prepositions electronic processing, ways of differentiating between the proper-prepositions, the words, which are used in the value of prepositions, the words, which appear in the function of prepositions, the words which come only by analogy like prepositions, and the words which are used occasionally as prepositions, are reflected. The procedure of queries creation in a base is represented.*

У рамках міжнародних проектів "Слов'янські прийменники в синхронії і діахронії: морфологія і синтаксис" і "Граматика категорії прийменника в російській та українській мовах. Загальнолітературна мова і діалекти" проблемною групою лінгвістів Донецького національного університету опрацьована концепція створення бази українських прийменників, що максимально охоплює їхній сучасний склад з урахуванням різних напрямів еволюції, динаміки й статичності.

База даних прийменників української мови реалізована за допомогою програмних продуктів компанії Microsoft – Microsoft Word і Microsoft Access. У її сьогоdnішньому вигляді база даних становить таблицю, що складається з 32 стовпців (полів), кожний з яких містить певну інформацію про прийменник, і 2028 рядків (записів), що відповідає кількості виявлених на сьогоdnішній день прийменників та їхніх еквівалентів в українській мові. Для зручнішої роботи з базою була сконструйована так звана форма, що складається з текстових полів, полів зі звичайним списком і таким, що розкривається кнопками, іншими елементами управління і деяких вбудованих процедур автоматичної обробки даних.

Реєстр прийменникових одиниць різного типу формувався в кілька етапів. На першому був вивчений реєстр прийменників, зафіксованих в 11-томному Словнику української мови, у зіставленні з реєстром прийменників Словаря української мови за редакцією Б. Грінченка і реєстром прийменників Великого тлумачного словника української мови. Кістяк бази даних прийменників української мови побудований на основі трьох словників: "Словарь української мови" (за ред. Б. Грінченка, 4 т., 1907 р.; перевид.: К., 1996–1997), "Словник української мови" (11 т., 1970–1980 р.), "Великий тлумачний словник сучасної української мови" (2001). Насамперед були відібрані одиниці, позначені в словниках як прийменники. Потім були виділені лексеми, позначені як інші частини мови, але такі, які функціонують як прийменники, що простежується з наведених наприкінці словникової статті прикладів. На останньому етапі враховувалися одиниці неприйменникових частин мови, які потенційно можуть вживатися зі значенням або у функції прийменника. Пізніше ця гіпотетична здатність підтверджувалася відібраними прикладами такого вживання в спеціально створених корпусах художніх, наукових та офіційно-ділових текстів ХХ ст. (близько 1, 5 млн. слів у кожному корпусі). Крім того, максимально був врахований матеріал списків прийменників російської і білоруської мов, люб'язно надані проф. М. В. Всеволодовою (Москва) і проф. М. Й. Конюшкевич (Гродно).

Отже, кожен запис бази даних, що відповідає певному прийменнику, складається з таких полів (див. мал. 1.):

1. "Прийменник" – тут наводиться сам прийменник або прийменникова конструкція.

2. "Керування". У цьому полі відображається інформація про те, з якими відмінковими формами іменних частин мови або ж з інфінітивом поєднується прийменник. Якщо прийменник поєднується з декількома відмінковими формами (напр.,  $O$  зі знахідним і місцевим відмінками  $= O + N_4/N_6$ ), то кожній такій конструкції присвячений окремий запис (у базі це  $O_1$  і  $O_2$ ). Потреба створення окремого запису зумовлена нетотожними семантико-синтаксичними відношеннями, різною синонімією, антонімією, частотністю вживання і под.

3. "№" привласнюється значення 1, 2 або 3. Якщо приєменник не має розгалуженої сполучуваності, то поле одержує значення 0;

4. "Тип". У цьому полі визначається тип приєменника – первинний (*без, в, з-під, між, навпроти, під* і под.) або вторинний (*попереду, коло, край, у ролі* та ін.).

5. "Статус". Відповідно до класифікації можливих статусів приєменника, запропонованої проф. М. Й. Конюшкевич (Гродно), поле може набувати одного з п'яти значень.

5.1. "Приєменник". До власне-приєменників належать усі традиційно кваліфіковані приєменники типу *у, за, про* тощо.

5.2. "Зі значенням приєменника". Коло слів, вжитих зі значенням приєменник, утворюють такі, які мають відповідники (дублети) з-поміж власне-приєменників, але містять додаткові (з узагальненою конкретною семантикою, без образності) елементи: *у напрямку до = до, у процесі слухання = протягом, у стані облоги = в облозі, під облогою, під час засідання = на засіданні*. Характерною особливістю таких приєменників постає рівнорядність постпозитивної іменниково-відмінкової форми: *під час засідання кафедри = на засіданні кафедри* та ін.

5.3. "У функції приєменника" – це такі елементи, які можуть бути опущені (редуковані) без відповідної заміни, але значення виразу при цьому не змінюється: *розмовляв з дівчиною на ім'я Настя = розмовляв з дівчиною Настею; перебувають на грані зникнення = зникають; виступив у ролі миротворця = виступив миротворцем* і под. Сюди ж слід віднести конструкції, в яких є можливим додавання узгодженого означення типу *такий (а, е, і), цей (я, е, і): у результаті обрання мене депутатом – у такому результаті, у ролі миротворця – у такій ролі, у разі вибуху – у цьому разі, у випадку розбіжностей – у такому випадку* тощо.

5.4. "Вживається за аналогією". Слова, що тільки за аналогією виступають приєменниками, – це досить обмежена група слів, основний масив якої складають структури з кількісним значенням тих чи інших вимірів (довжина, глибина, розмір, товщина, ширина, об'єм та ін.): *завдовжки понад, завдовжки від; завтовшки понад, завтовшки від; завширшки понад, завширшки від, з відстані (не) більш(е); з відстані (не) більш(е), (а)ніж; з відстані (не) більш(е), (а)ніж у/в; з відстані (не) менш(е); з відстані (не) менш(е), (а)ніж; з відстані (не) менш(е), (а)ніж у/в; з відстані від; з відстані від ... до; з відстані до; з відстані понад; з відстані у/в та ін.*

5.5. "Вживається оказіонально" – утворення, в яких супровідною постає певна образність і які використовуються в ролі приєменника в одиничних випадках: *на світанку людської історії, на горизонті сучасної науки* і под.

6. "Позиція". У цьому полі вказується, якою може бути прислівна позиція приєменникової одиниці: препозиція (*у сфері науки*); постпозиція (*рік тому, тиждень по тому*); препозиція / постпозиція (*заради слави / слави заради*); інтерпозиція (*один до одного*); рамка (*далеко я живу від університету*); препозиція / рамка (*заввишки у три метри / у три метри заввишки*).

У наступних полях репрезентована інформація про відображення приєменника в трьох зазначених словниках, у корпусі текстів і узагальнені відомості, повний обсяг яких буде отриманий на фінальному етапі створення бази. Зокрема, вказується, чи наявна приєменникова одиниця в словнику або тексті взагалі (відповідне поле може набувати значення "так" або "ні", тобто ставиться галочка), якщо так, то з якими значеннями, варіантами і синонімами, до якої з частин мови зараховують аналізовану приєменникову одиницю укладачі словника. Наприклад, у словнику Б. Грінченка слово *навпаки* визначається як прислівник, хоча з прикладу видно, що це безсумнівний приєменник: *Покотила паляницю навпаки сонця* (Г. Квітка-Основ'яненко) (Словарь, II: 472). Кожне значення ілюструється прикладом. Список значень наводиться в спеціальному полі, яке знаходиться над кнопкою "Додати". На сьогодні виділено 16 основних значень, що уявляється достатнім для реалізації завдання створення відносно вичерпного реєстру українських приєменників та переліку їхніх значень:

а) об'єктне (*без, в, в боротьбі з, в боротьбу за, до, з, під керівництвом, про* та ін.): *Скіфське плем'я арімастів вступило з грифами в боротьбу за золото й коштовні камені, то були безстрашні варвари* (П. Загребельний, Диво);

б) адресатне (*до, для, на ім'я* тощо): *Що це важить супроти приємності бути поголеним! Хоч процедура часом і бувала болючою (особливо для тих, хто мав цупку*

рослинисть), зате закінчення було радісним, а найбільш приємним був результат (І. Багрянний. Сад Гетсиманський);

в) інструментальне (в, за, на і под.): *Дай заграю я на дудку, а то давно вже грав* (М. Драгоманов);

г) локативне (біля, близько, ближче, в, в бік, в глиб, від, до, за, коло, край, на, над, під, після, при, у межах, у місці, у напрямі, через та ін.): *Рано-вранці біля сільської Ради юртувалося десятків два підвід* (Г. Тютюнник. Вир);

ґ) міри і ступеня (більше, завшишки, коло, мало не, менше, на, на... більше, навбільши, навглибки, об, по, у міру тощо): *Сушені груші, сливи та яблука свої. Господи Боже – мало не два карбованці готовими грішми видалось, аж страх подумати – таке нечуване марнотратство* (У. Самчук. Волинь);

д) причини (від, внаслідок, з, з огляду на, за, завдяки, зважаючи на, під впливом, у зв'язку з, у нагороду за, у надії на, у силу, через і под.): *Таж, якщо матуся витримала стільки горя, що звалилося на її слабенькі плечі, то тільки завдяки своїй вірі і завдяки моральній підтримці* (І. Багрянний. Сад Гетсиманський);

е) способу дії (без, з, без допомоги, до, за допомогою, на, у розрізі, у супроводі, у такт, у тон, через та ін.): *Найвідважніші з гуляк, роззойдуючись, злітали аж під покрівлю і, вигинаючись, без допомоги рук, намагалися відкусити від граната, калача або від баранини* (П. Загребельний. Роксолана);

ж) наслідку (внаслідок, під дією, під владою, під впливом, у результаті, тощо): *Маленька піщанка завжди залишається піщанкою, тоді як навіть могутні гори руйнуються під дією стихій, і що вищі вони, то тяжчі й страхітливіші руйнування на їхньому велетенському тілі* (П. Загребельний. Роксолана);

з) порівняльне (подібно до, порівняно з, порівняно до, проти, у порівнянні з і под.): *І вже вона не здається важкою в порівнянні з «Красою», навпаки, здається, що вони обидві мають щось надзвичайно спільне й рідне між собою* (В. Винниченко. Сонячна машина);

і) темпоральне (в еру, в епоху, до, з ...до, за, напередодні, одночасно з, під час, під час перебування, пізніше, після, по, при, протягом, у добу, у процесі, у ході тощо): *Мистецтво як форму духовно-релігійного діалогу людини з Богом (на основі концепції активності мистецтва, які сформувалися в епоху середньовіччя в лоні християнського віровчення) розглядає М. Бровко* (Квасюк Л. В. Іконопис і словесність у культурі України XVI–XVIII століть: естетичний аспект: Автореф. дис. ... канд. філос. наук. – К., 2000);

к) умови (за умови, при, при нагоді, у випадку, у разі та ін.): *Не діждавшись нічого від своїх в'язнів, комендант так само байдуже дав час для роздумів до обіду, пригрозивши, що в разі мовчанки він розстріляє кожного десятого* (П. Загребельний. Диво);

л) допустове (без огляду на, всупереч, наперекір, на противагу, незважаючи на, незалежно від, попри і под.): *Він одривається від ґрат і ходить по темній камері, не звертаючи уваги ані на те, що камера, всупереч тюремним законам, зовсім не освітлена, ані на тяжке, насичене духом людського поту, цвіллю й смородом блощиць повітря* (І. Багрянний. Сад Гетсиманський);

м) мети (в ім'я, в інтересах, для, для блага, з метою, за, задля, заради, на, на честь, по, у пошуках, у справі тощо): *Всякий акт, що робиться в інтересах цієї грандіозної мети, є акт святий* (В. Винниченко. Сонячна машина);

н) атрибутивне (у/в, від під, після та ін. – за умови вживання в присубстантивній позиції): *Маруся Чурай (історичний роман у віршах)* (Л. Костенко);

о) кількісно-атрибутивне (від ... до, о, об, об'ємом, об'ємом (не) більш(е), об'ємом (не) менш(е), об'ємом (не) менш(е) за, об'ємом (не) менш(е) (а)ніж, об'ємом (не) менш(е) як, об'ємом від, об'ємом від ... до, об'ємом до, об'ємом з/із/зі, об'ємом понад, об'ємом у/в, обсягом, обсягом (не) більш(е), обсягом від, обсягом від ... до, обсягом до, обсягом з/із/зі, обсягом понад, обсягом у/в): *Порушення цієї норми тягне за собою відповідальність у вигляді штрафу від 100 до 1000 неоподатковуваних мінімумів та компенсацію оператору ТМ понесених збитків та недоотриманого прибутку* (Про кабельне, ефірно-кабельне телебачення та телеінформаційні мережі. Закон України);

п) суб'єктне значення, яке у межах приєднаних одиниць є вторинним, оскільки відповідні конструкції є трансформами ядерних речень: *Прелюдія до скандалу пролунала у виконанні Адама Прайса, члена британського парламенту від націоналістичної партії Уельсу Plaid Cymru, – він роздобув копію листа Тоні Блера й запідозрив його в лобюванні інтересів сталєвого магната* (Ю. Загоруйко. «Стальгейт» чи «сміттегейт»? // Дзеркало

тижня. – 2002. – 2–8 березня (№ 8 (383)) (сп. дії+суб.) ← *Адам Прайс виконавав прелюдію...*

Окреме місце посідають предикативна й структурна функції. Щодо першої, то йдеться про регулярність вживання прийменника в присудковій позиції: *в ізоляції від, в опозиції до, в опозиції з/із/зі, у гонитві за, у залежності від, у зв'язок з/із/зі, у полоні, у претензії на, у ранзі*. Структурна функція є асемантичною, її роль зведена до поєднання лексем. Вживання прийменникових одиниць у структурній функції є поодиноким і складає периферію: *По матері вона черкешенка, тому намірилася розбудити в синові мужчину, знайшовши для нього гідну жону з свого племені* (П. Загребельний. Роксолана).

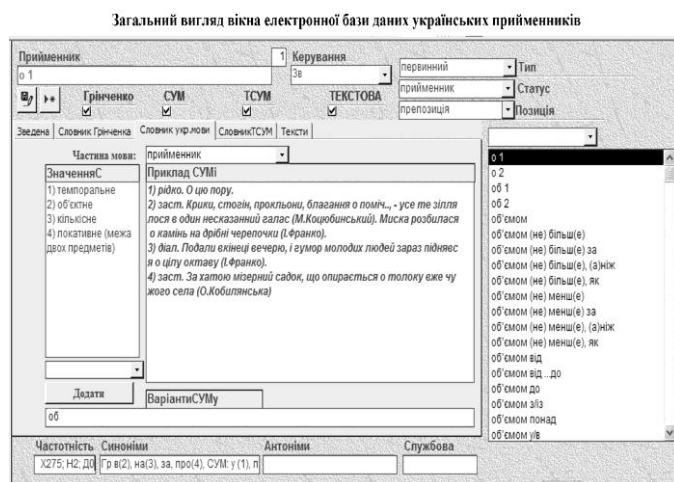
Стосовно поля "Тексти", то тут відображені значення і приклади вживання тих прийменників, які не були виявлені в словниках, але знайдені в корпусах текстів трьох функціональних стилів української мови, а також у мережі Інтернет.

У полі "Варіанти" зазначена варіативність прийменникових одиниць у такому вигляді, як її подають словники: напр., прийменник *VI* (з Р.в.) має тільки один варіант – *В*, а *B2* (зі З.в.) – *В, VI, VO, UB*.

Поля "Синоніми" та "Антоніми" заповнювалися не тільки з урахуванням лексикографічних відомостей, але й на підставі спостережень проблемної групи над функціонуванням прийменників та їхніх еквівалентів у текстах різних функціональних стилів української мови: прийменникова сполука *аналогічно до* має синоніми *аналогічно (з/із/зі), відповідно (до / з/із/зі), за аналогією до, подібно (до)* та антоніми *на відміну від, на протилежність (до), на протилежність*.

Запис типу «X5; N13; D17» у відповідному полі відображає абсолютну частотність у корпусах художніх текстів (твори класиків української літератури XIX і XX століть, а також сучасних письменників), наукових текстів (переважно автореферати дисертацій з різних галузей знань), а також текстів сучасного офіційно-ділового стилю (кодекси, закони, постанови, накази і под.) загальним обсягом біля 4500000. У випадку відсутності ілюстративного матеріалу в текстовій базі (тоді абсолютна частотність дорівнює нулю) були використані приклади з інших джерел, зокрема, сучасної української періодики, а також ресурси Інтернету (як джерело вказується повна адреса відповідної веб-сторінки).

Рисунок 1.



Стосовно поля "Тексти", то тут відображені значення і приклади вживання тих прийменників, які не були виявлені в словниках, але знайдені в корпусах текстів трьох функціональних стилів української мови, а також у мережі Інтернет.

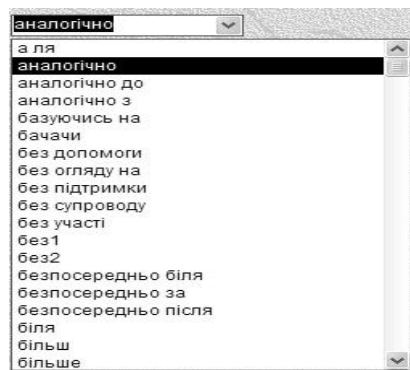
У полі "Варіанти" зазначена варіативність прийменникових одиниць у такому вигляді, як її подають словники: напр., прийменник *VI* (з Р.в.) має тільки один варіант – *B*, а *B2* (зі З.в.) – *B, VI, VO, UB*.

Поля "Синоніми" та "Антоніми" заповнювалися не тільки з урахуванням лексикографічних відомостей, але й на підставі спостережень проблемної групи над функціонуванням прийменників та їхніх еквівалентів у текстах різних функціональних стилів української мови: прийменникова сполука *аналогічно до* має синоніми *аналогічно (з/із/зі), відповідно (до / з/із/зі), за аналогією до, подібно (до)* та антоніми *на відміну від, на протилежність*.

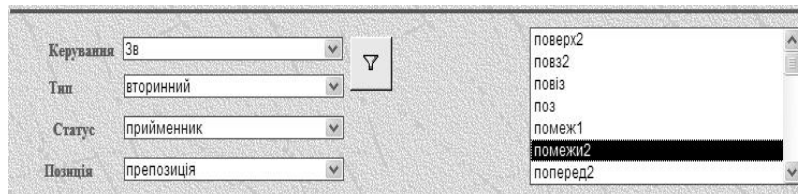
Запис типу «X5; H13; D17» у відповідному полі відображає абсолютну частотність у корпусах художніх текстів (твори класиків української літератури XIX і XX століть, а також сучасних письменників), наукових текстів (переважно автореферати дисертацій з різних галузей знань), а також текстів сучасного офіційно-ділового стилю (кодекси, закони, постанови, накази і под.) загальним обсягом біля 4500000. У випадку відсутності ілюстративного матеріалу в текстовій базі (тоді абсолютна частотність дорівнює нулю) були використані приклади з інших джерел, зокрема, сучасної української періодики, а також ресурси Інтернету (як джерело вказується повна адреса відповідної веб-сторінки).

База супроводжується певним набором запитів, що дозволяють одержувати узагальнену інформацію, а також елементи навігації у вигляді активного списку (див. рис. 2 і 3).

*Рисунок 2.* Відсортований за абеткою активний навігаційний список для пошуку прийменникової одиниці у базі



*Рисунок 3.* Частина макету форми для роботи із запитами. Праворуч – відсортований за абеткою активний навігаційний список із результатами запиту.



При укладанні бази "Українські прийменники та їхні еквіваленти" визначальними є структурна типологія, врахування парадигматичних ресурсів прийменника і простеження його синтагматики. Урахування цього дає підстави для цілісного відображення прийменника зі встановленням відповідних функціональних виявів. Особливу вагу має

встановлення граничності прийменника, умов його внутрішнього членування і впливу на семантику як внутрішніх, так і зовнішніх показників.

#### Література

1. Берко А., Верес О. Організація баз даних: практичний курс. – Львів: Національний університет «Львівська політехніка», 2004. – 152 с.
2. Вихованець І. Р. Прийменникова система української мови. – К.: Наукова думка, 1980. – 288 с.
3. Всеволодова М. В., Клобуков Е. В., Кукушкіна О. В., Поликарпов А. А. К основам функционально-коммуникативной грамматики русского предлога // Вестник Московского университета. – Серия 9. Филология. – М.: Изд-во МГУ. – 2003. – № 2. – С. 17–59.
4. Карпіловська Є. А. Вступ до комп'ютерної лінгвістики. – Донецьк: ТОВ «Юго-Восток, Лтд», 2003. – 184 с.
5. Українські прийменники: Синхронія і діахронія: пробний зошит / За ред. проф. А. П. Загнітка. – Донецьк: ДонНУ, 2003. – 162 с.

*А. Карчова, М. Шимкова\**

Інститут мовознавства імені Л. Штура Словацька академія наук  
(Братислава, Словацька республіка)  
УДК 81'322.3

#### МОРФОЛОГІЧНА АНОТАЦІЯ ТЕКСТІВ СЛОВАЦЬКОГО НАЦІОНАЛЬНОГО КОРПУСУ

*Abstract. From 2004 to 2005, four different kind of texts have been manually disambiguated and morphologically annotated: Orwell's 1984 novel, selected Slovak Wikipedia texts, Internet magazine InZine and SME daily newspaper articles. Empirically estimated precision of manual annotation is presented, together with description of the annotation process. The result is high quality manually annotated corpus of Slovak language texts, containing over 300 000 words.*

Одночасно із першими кроками побудови Словацького національного корпусу (СНК) в 2002 – 2003 рр. розпочалася підготовка морфологічної анотації текстів. Паралельно зі збиранням текстів та їх первісним опрацюванням формувалася внутрішня структура корпусу, йдеться про технічний аспект, та здійснювалася стильово-жанрове опрацювання текстів. Важливою справою було придбання придатного корпусного програмного засобу чи менеджера. Не менш важливим було також запитання: розпочинати морфологічну анотацію чи ні, і якщо так, то яку саме?

На момент початку робіт над створенням СНК у Словаччині комп'ютерна або корпусна лінгвістика як галузь науки не були достатньо розбудованими і ще не існували відповідні до морфологічної анотації програмні забезпечення, не кажучи вже про лінгвістичні описи, які можна було б використати для комп'ютерного опрацювання словацької мови. У той же час у світі ті, хто створювали та використовували електронні корпуси, щораз більше орієнтувалися на лінгвістичні анотації корпусів, на відміну від минулого, коли корпуси будували на вірець комп'ютерних баз даних слів чи текстів із майже виключно лексикографічним призначенням, а лінгвістичні інформації, які вносилися в корпуси, вважалися зайвими через змінність граматичних категорій. Лінгвістичне анотування корпусів стимулювало появу знаряддя для морфологічної анотації флективних мов. Таке ж знаряддя безпосередньо для словацької мови було створено на Фізико-математичному факультеті Карлового університету в Празі, і виникло як другорядний продукт знаряддя на лематизацію, аналіз та зняття неоднозначності для чеської мови, яке застосовували в Чеському національному корпусі.

На першому етапі побудови Словацького національного корпусу – етапі підготовки морфологічної анотації текстів, – було здійснено об'єднання вже наявних підходів до морфологічних анотацій корпусних текстів [3], проведено тестове перевіряння чеського знаряддя, яке для нашого відділу СНК віддала на підставі договору вже згадувана інституція, а також вирішено, яким чином здійснюватиметься анотація текстів СНК. На підставі тестового аналізу празького програмного забезпечення було виявлено, що:

а) програмне забезпечення, створене на початку 90-х років у межах проекту Чеського національного корпусу SYN 2000, було доволі часто критиковане за надто строгий технічний підхід; за відсутність мітки для категорії виду у дієслова, яка є важливою специфічною ознакою більшості слов'янських дієслів; під критику підпав також

\* © А. Карчова, М. Шимкова, 2006