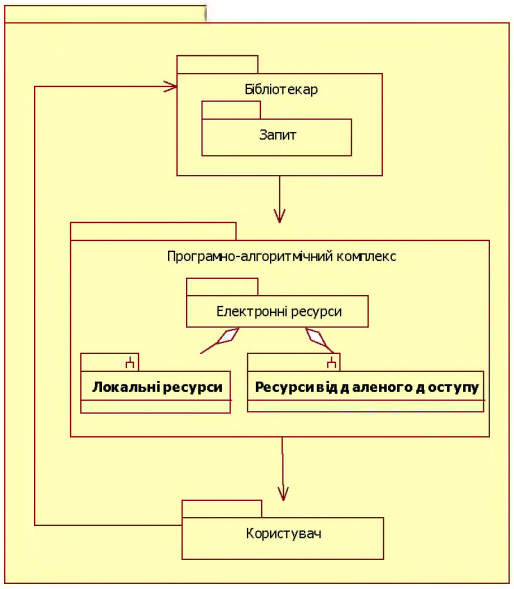

Рис. 1. Модель інформаційного забезпечення віртуального наукового колективу

Використання такої технології забезпечує співпрацю науковців різних країн для досягнення поставленої мети без фізичного переміщення у один населений пункт.

Література

1. Cummings J.N. Collaborative research across disciplinary and institutional boundaries / Cummings J. N., Kiesler S. // Social Studies of Science. – 2005. – Vol. 35, № 5. – P. 703–722.
2. Petersen S.A. The role of enterprise modeling in virtual enterprises / Sobah Abbas Petersen // Collaborative Networks and Their Breeding Environments: IFIP TC5 WG 5.5 Sixth IFIP Working Conference on VIRTUAL ENTERPRISES (26-28 September, 2005, Valencia, Spain) / Edited by Luis M. Camarinha-Matos, Hamideh Afsarmanesh and Angel Ortiz. – Valencia , 2005. – P. 109–117.
3. Towards a framework for creation of dynamic virtual organizations / Luis M. Camarinha-Matos, Ivan Silveri, Hamideh Afsarmanesh, Ana Ines Oliveira // Collaborative Networks and Their Breeding Environments: IFIP TC5 WG 5.5 Sixth IFIP Working Conference on VIRTUAL ENTERPRISES (26-28 September, 2005, Valencia, Spain) / Edited by Luis M. Camarinha-Matos, Hamideh Afsarmanesh and Angel Ortiz. – Valencia, 2005. – P. 69–81.
4. Лобузіна К. Технології організації знаннєвих ресурсів у бібліотечно-інформаційній діяльності : монографія / Катерина Лобузіна ; відп. ред. О. С. Онищенко ; НАН України, Нац. б-ка України ім. В. І. Вернадського. – Київ, 2012. – 252 с.

II. КОРПУСНА ЛІНГВІСТИКА

Синтагматична параметризація еквівалентів слова у парадигмі корпусної лінгвістики

Алла Лучик

д. філол. н., професор, професор кафедри загального і слов'янського мовознавства, Національний університет «Києво-Могилянська академія», Україна, E-mail: allal@meta.ua

Ірина Остапова

к. т. н., старший науковий співробітник Українського мовно-інформаційного фонду НАН України, Україна, E-mail: irinaostapova@gmail.com

This work offers possible ways to analyze syntagmatic properties of word equivalents, formed by a "preposition + noun" model. Linear sets in which word equivalents are functioning were discovered on the basis of texts from the Ukrainian National Linguistic Corpus. Grammar parameters of structural elements adjacent to studied units, which make up the syntagmas of 2–7 words, were described according to linguistics methodology. Integrated solution of assigned tasks should be preceded by syntagmatic splitting of Ukrainian National Linguistic Corpus texts, which is one of the aspects of its further development.

Ключові слова — Український національний лінгвістичний корпус, еквівалент слова, синтагма, синтагматичні відношення.

Сьогодні лінгвістичні корпуси та лексичні бази в Інтернеті — найбільш популярні ресурси як для лінгвістів-дослідників, так і для фахівців у сфері інформаційних технологій.

Розвинені корпусні технології покликані перетворити лінгвістику в експериментальну науку у сучасному розумінні, тим самим розширивши її

теоретичні горизонти. Сьогодні все частіше виявляється усвідомлення лінгвістами необхідності фронтального обстеження мови і у феноменальному, і у концептуальному аспектах.

У загальному розумінні лінгвістичний корпус – це представлений у цифровому форматі, великий за обсягом, уніфікований, структурований, розмічений і філологічно компетентний масив текстів природною мовою, доповнений системою керування — універсальними програмними засобами для пошуку та опрацювання різноманітної лінгвістичної інформації.

Лінгвістичні корпуси на початку історії свого розвитку були покликані замінити традиційні лексичні картотеки, які споконвіку створювалися та використовувалися лінгвістами при укладанні словників та виконанні мовознавчих досліджень. Із розвитком комп'ютерних технологій завдання лексичних картотек було вирішено новими методами: програмні засоби забезпечують граматичну ідентифікацію та дозволяють забезпечити прямий доступ до всіх контекстів будь-якої досліджуваної мовної одиниці. Але для цього комп'ютерна система повинна уміти виконувати граматичну ідентифікацію усіх цих досліджуваних одиниць та забезпечувати прями доступ до їхніх контекстів.

Обсяги мовного матеріалу, який залучається до мовознавчого дослідження, комплексність, оперативність опрацювання зазначеного матеріалу та можливість прямого доступу до великої кількості лінгвістичних фактів — це ті переваги, які надає лінгвістичний корпус досліднику. Для забезпечення цих переваг лінгвістичний корпус повинен мати відповідну будову, спрямовану на представлення, маркування та експлікацію необхідної інформації у системі. Отже, загальна структура будь-якого лінгвістичного корпусу передбачає наявність як мінімум трьох підсистем, що взаємодіють між собою: а) репрезентації (представлення) даних; б) маркування («тегування», «анотування»; розмітки) і в) експлікації даних (у форматах як для людини-дослідника, так і для програмних агентів).

В Українському мовно-інформаційному фонді НАН України з метою дослідження української мови та укладання сучасних словників створено Український національний лінгвістичний корпус (УНЛК) [5, с. 145–162].

Основні напрямки УНЛК:

– надання текстової інформації за певними критеріями;

– створення вхідних потоків лінгвістичної інформації для різноманітних дослідницьких систем;

– інтеграція різнопланових лінгвістично-програмних засобів обробки текстів у єдиному середовищі.

У системі УНЛК виділяються дві основні функційні підсистеми: бібліографічна та повнотекстова.

Бібліографічна частина являє собою електронну бібліотеку — колекцію цифрових ресурсів, яка є основою для розробки будь-якого корпусу. Тому бібліографічна підсистема служить інструментом для збирання, збереження, моделювання й використання природномовної інформації в цифровому вигляді. Обсяг ресурсів не обмежений, тобто корпус постійно поповнюється новими джерелами.

Друга можливість — це повнотекстовий пошук. Користувач уводить пошукову фразу, задає максимально бажану кількість слів між пошуковими та обирає додаткові параметри повнотекстового пошуку, а саме:

– урахування порядку слів;

– пошук у певній підмножині об'єктів;

– використання процедури лематизації;

– використання синонімічної лексикографічної бази даних;

– використання певних синонімічних рядів;

– вибір граматичних параметрів для кожного слова, що входить у пошуковий фрагмент.

Результатом повнотекстового пошуку є список бібліографічних описів. Але на відміну від пошуку за бібліографією, користувач отримує прямий доступ до кожної локалізації пошукового фрагмента в тексті, тобто до всіх контекстів, які містять пошуковий фрагмент. Обравши джерело, користувач може переглядати контексти (для зручності пошуковий фрагмент виділено червоним кольором). Розмір (довжину) контексту можна змінювати. Кожне слово з контексту з'єднане зі словниками: граматичним, тлумачним і словником синонімів.

Ще одним засобом для створення користувацьких наборів даних у системі є так званий «кошик» (тимчасове сховище). Користувач може відібрати об'єкти збереження, які його цікавлять, з різних пошукових запитів і зберегти образ даних для подальшої роботи в інших сеансах. Користуючись таким інструментом, дослідник може відібрати джерела певного автора, стилю чи жанру і працювати лише з цією частиною корпусу; така процедура дає можливість виділити із загального корпусу свій власний підкорпус, орієнтований на розв'язання особистих завдань. Такий підхід дозволив відмовитися від занадто суб'єктивною ідеї створення «еталонного корпусу» і надає можливість досліджувати засобами корпусу «нееталонних», аномальних мовних явищ. Отже, реалізовані в УНЛК функції кошика дозволяють породжувати в режимі реального часу та в рамках єдиного лінгвістичного корпусу віртуальні «підкорпуси», які моделюють ті або інші лінгвістичні ефекти із забезпеченням кожному дослідникові можливостей реалізації його власних уявлень про еталонність, виваженість і збалансованість. Це змінює парадигму корпусних досліджень: ми досліджуємо не наперед заданий набір текстів, а дискурси, незалежно від того, якими текстами вони підтримуються. Система лінгвістичного корпусу є багатоплановою і може мати чимало різноманітних застосувань. В Українському мовно-інформаційному фонді вона, насамперед, використовується як джерельна база лінгвістичної

інформації для створення фундаментальної академічної багатотомної лексикографічної системи «Словник української мови».

Серед інших застосувань слід відзначити здійснення лінгвістичних досліджень з метою виявлення нових мовних явищ та формалізації наявних.

Зрозуміло, що технологія УНЛК надає засоби для граматичного та семантичного маркування текстів. Лінгвістичний корпус може бути і зручним середовищем для статистичного опрацювання текстової інформації.

Оскільки основним напрямком УНЛК є надання текстової інформації за певними критеріями, ми звернулись до лінгвістичного корпусу для встановлення синтагматичних властивостей еквівалентів слова, що дало нам можливість вирішувати поставлені задачі на максимально великому за обсягом матеріалі.

Відзначимо, що у сучасному мовознавстві вже детально проаналізовані парадигматичні властивості еквівалентів слова, зокрема прислівникового, прийменникового, сполучникового статусу існування. Висвітлені семантичні, стилістичні їхні параметри, а також шляхи набування цими елементами статусу лексикографічних одиниць [1]. Нагадаємо, що вслід за Р. П. Рогожниковою під еквівалентами слова нами розуміються одиниці, що становлять собою «зв'язані сполучення, які характеризуються стійкістю, єдністю значення, здебільшого постійною, незмінною формою. У мовленнєвому потоці еквіваленти слова становлять цілісну одиницю і стосовно наголосу, звичайно вони мають один словесний наголос» [3, с. 4]. До цієї дефініції належить додати те, що творення еквівалентів слова відбувається за чітко встановленими у мові моделями.

Проте і досі у зазначених нарізнооформлених одиниць не встановлені синтагматичні характеристики, які, безумовно, впливають на набуття ними цілісності, відсутність якої не дозволяє кваліфікувати їх як лексеми. Максимально ж повне пізнання мовної одиниці виявляється можливим лише при обстеженні її як парадигматичних, так і синтагматичних властивостей, на чому наполягав свого часу видатний Ф. де Соссюр, висвітлюючи свої погляди на досить яскравих прикладах. «З обох поглядів мовну одиницю можна порівняти з частиною будинку, наприклад з колоною: з одного боку, — відзначав вчений, — колона перебуває у певному відношенні з підтримуваним нею архітравом, — це розташування двох одиниць, однаково присутніх у просторі, нагадує синтагматичне відношення; з іншого боку, якщо це колона доричного ордера, вона викликає в думці порівняння з іншими ордерами (іонічним, коринфським тощо), які є елементами, відсутніми у даному просторі, — це асоціативне відношення»[4, с. 157].

Як бачимо, обстежуючи дії мовного механізму, Ф. де Соссюр важливе місце приділяв дослідженню як парадигматичних, так і синтагматичних відношень мовних одиниць. З приводу останніх учений висловлювався так: «Перше, що нас вражає в цій організації, — це синтагматичні єдності: майже всі одиниці мови залежать або від того, що їх оточує у мовленнєвому ланцюжку, або від частин, з яких вони складаються самі» [4, с. 162].

В українському мовознавстві обстеження лінійних властивостей мовного знака за допомогою формалізованого аналізу першою здійснювала М. М. Пещак. Цим шляхом дослідниця розв'язувала надскладні питання лінгвістичної інтерпретації ділових документів XIV ст. М. М. Пещак переконувала, що для того, «щоб обсяг об'єкта лінгвістичного дослідження міг бути використаним у вигляді формальної ознаки, потрібно відшукати для нього одиницю виміру, яка була б і окресленою, і в даному випадку максимально однозначно інтерпретованою» [2, с. 9]. Традиційно в дослідженнях зазначеного типу одиницями виміру бувають літери, морфеми, слова, словосполучення, речення. Проте мова реалізується в тексті, де структурні одиниці вишиковуються в особливий ряд, вступаючи між собою у відношення, зумовлені лінійним характером мовленнєвого потоку [2, с. 80]. Вивчення особливостей функціонування мовних одиниць зобов'язує дослідника до обстеження закономірностей побудови зазначеного ряду, який для цього може бути поділений на низку послідовно пов'язаних між собою синтагм. У лінгвістичних студіях усталилося таке розуміння синтагми: «Ритмико-інтонаційна єдність, мовленнєвий такт, що складається з одного чи кількох слів, об'єднаних у смисловому та інтонаційному відношенні» [6, с. 613].

У пізнанні ж особливостей формування еквівалентів слова за допомогою обстеження закономірностей їхнього функціонування у синтагмі враховуємо думку Ф. де Соссюра про те, що «член синтагми отримує мовну вартість лише внаслідок свого протиставлення до попереднього члену або наступного, чи до того й іншого разом» [4, с. 156].

Отже, для встановлення синтагматичних характеристик еквівалентів слова нами залучені тексти УНЛК, в яких були встановлені фрагменти уживання еквівалентів слова, утворених за моделями ***на*** + *іменник у знахідному відмінку*, ***в*** + *іменник чоловічого роду у знахідному відмінку,* ***в*** + *іменник у місцевому відмінку,* ***без*** + *іменник у родовому відмінку*: *на взір, в акурат, в гурті, без пам'яті.* Зазначені фрагменти були поділені на синтагми. Аналізовані тексти засвідчили наявність синтагм з одним — сімома лексичними елементами.

Опис структури синтагматичних одиниць почнемо з двослівних синтагм. На першому етапі цей аналіз базуватиметься на дослідженні компонентного складу синтагм, встановленні частиномовної належності сусідніх до еквівалентів слова одиниць. У подальших дослідженнях це коло завдань буде розширене вивченням засобів вираження відношень між компонентами синтагми. Зазначений підхід не новий у лінгвістиці і має вагомі результати. Так,

М. М. Пещак ще у 70-ті роки застосування подібної методики дослідження дозволило здійснити спробу авторизації неавторизованих документів на прикладі грамот XIV ст. [2]. Запропонований підхід, по суті, є домашинним аналізом, який дозволить у подальшому автоматизувати ту частину спілкування з машиною, яку ускладнюють семантично неподільні, але нарізнооформлені одиниці мови.

Отже, найтиповішими синтагмами, до складу яких входять еквіваленти слова, є двослівні структури. При цьому у понад 75 % синтагм еквіваленти слова виявляються початковими елементами синтагми, закривають же такі синтагми майже завжди дієслова: ***без вагань*** *припав,* ***без запинки*** *зіграю,* ***без пам'яті*** *закохався,* ***в акурат*** *вистачило,* ***в гурті*** *засміялись.* Лише в одному із зафіксованих прикладів лівим розгортанням прислівникового еквівалента двокомпонентної синтагми є власне прислівник: *в акурат посередині.* Проте якщо еквівалент слова закриває синтагму, її початковий компонент може бути практично будь-якою частиною мови – дієсловом, іменником, прикметником, дієприкметником, займенником, сполучником, часткою: *арештували* ***без вагання****, банальності* ***в ходу****, негарний* ***на взір****, вони* ***без копійки****, або* ***в гостину****, оце* ***в акурат****.*

У трислівних синтагмах, яких дещо менше, ніж двокомпонентних, еквіваленти слова можуть започатковувати лінійні ряди, закривати їх, а також перебувати у центрі виокремлених конструкцій: ***без запинки*** *відповів Андрій, будуть продаватися* ***на вагу****, дає* ***на вибір*** *Одіссеєві.* Основною позицією еквівалентів слова у трикомпонентних синтагмах, на відміну від двокомпонентних, є кінцева, яка фіксується у понад 50 % цих структур. У ролі початкового і центрального компонентів прислівникові еквіваленти тут виступають рівною мірою. Спільною ознакою із попереднім типом синтагм у зазначених виявляються поширене сусідство аналізованих одиниць із дієсловами, хоча тут це і не має такого ступеня регулярності. Так, у кінцевій позиції конструкцій, прислівникові еквіваленти лише у 15 % синтагм лівим «сусідом» виявляється дієслово, проте такими можуть практично рівною мірою бути іменники, займенники, частки, сполучники, прислівники: *до* ***попа в гостину****, для всіх* ***без винятку****, варнякаючи щось* ***на бігу****, бути завжди* ***в гостях****, хоч би* ***в гості****, жерли ж* ***без пам'яті.***

Ця тенденція спостерігається і у синтагмах, у яких аналізовані одиниці займають центральну позицію. Тут, хоча і переважають синтагми із початковим дієслівним елементом, на цій позиції можуть перебувати іменники, займенники, сполучники, частки: *дає* ***на вибір*** *Одіссеєвіі, ноги* ***в головах*** *ставиш, всі* ***без винятку*** *фракції , і* ***без кінця*** *обдурюю, лише* ***на вигляд*** *теоретик.* Закривають же зазначені лінійні ряди переважно також дієслова або іменники: *всіх* ***без винятку*** *переоформити ловив* ***на бігу*** *Скорика.* Трапляються випадки, коли прислівниковий еквівалент перебуває у дієслівному оточенні: *обіцяю* ***без затримки*** *висадити.*

Чотирислівні синтагми у близько 70 % закриваються прислівниковим еквівалентом, що виступає правим поширювачем дієслова, рідше – іменника, займенника, прикметника, які здатні лівобічно об'єднуватися із займенником, прийменником, сполучником, числівником: *і кожен скаже* ***без вагань****, їдемо до гетьмана* ***в гостину****, була від нього* ***в захваті****, до своєї матері* ***в гостину****, ясна і ніжна* ***без кінця****.*

Практично в усіх чотирислівних синтагмах центральна позиція прислівникового еквівалента є наступною за початковим компонентом, який може бути дієсловом, прикметником, прийменником, сполучником, часткою: *розговіється* ***в гостях*** *за піснею, гарний* ***на взір*** *козак Чміль, від усіх* ***без винятку*** *підприємств, що* ***в гурті*** *каша істся, вже* ***в деякій мірі*** *примирила Ліду.*

Як початковий у чотирислівних синтагмах зафіксовані лише вживання прислівникового еквівалента *в акурат*: ***в акурат*** *на одну могилу,* ***в акурат*** *під Новий рік,* ***в акурат*** *десять років тому.* Цей самий прислівниковий еквівалент здатен започатковувати й п'ятислівні синтагми: ***в акурат*** *хоч тепер баба вгамується,* ***в акурат*** *із закінченням контрольного матчу.* Проте у цього типу синтагм, на відміну від попередніх, початковими можуть бути й інші прислівникові еквіваленти: ***без вагання*** *я помінявся б усім,* ***на бігу*** *щось там норовили вхопити,* ***на вигляд*** *не такий уже моцак.* Як можна побачити, розгортання таких синтагм здійснюється переважно займенниками та службовими словами. Об'єднання із такими мовними одиницями у зазначеному типі синтагм властиве і тим прислівниковим еквівалентам, які займають кінцеву позицію, тобто їхніми лівобічними сусідами теж здатні виявляютися займенники і службові слова, хоча такими можуть бути і дієслова та прислівники: *не запрошувала до себе* ***в гості****, й поставив її теж* ***в головах****, не дозволить тобі жити* ***в боргу****, і одчайдушний він зараз* ***на вигляд****.*

Подібну мережу синтагматичних відношень виявляють і ті прислівникові еквіваленти, що займають центральну позицію у зазначених лінійних рядах, хоча вони і переважно поєднуються з дієсловами, що перебувають у препозиції чи постпозиції. Так, у п'ятислівних синтагмах у препозиції до прислівникових еквівалентів можуть перебувати дієслова, іменники, сполучники, а у постпозиції — дієслова і прийменники: *не ходити* ***в гостину*** *до Скоропадських, до гетьмана* ***в гостину*** *зволив прибути, а* ***в деякій мірі*** *на пораду артистам, і* ***в гостин****у до тебе став.* У єдиній зафіксованій шестислівній синтагмі, прислівниковий еквівалент, що займає центральну позицію, теж взаємодіє у препозиції з дієсловом, а у постпозиції – з прийменником: *який проліг «****в акурат****» по діагоналі кварталу.*

У найскладнішій за кількістю компонентів семислівній синтагмі прислівниковий еквівалент виступає початковим, правобічні поширювачі його упорядковуються за схемою: частка + числівник + займенник + іменник + частка + дієслово: ***без клопоту*** *ні одна вам сім'я не обійдеться.*

Фіксуються також однослівні синтагми, що є самодостатніми у смисловому відношенні, а це, безумовно, і сприяє трансформації прийменниково-іменникових сполучень, які у нашому випадку входять у структурні моделі, побудовані за схемами ***в*** + *іменник чоловічого роду однини у місцевому відмінку*, ***без*** + *іменник множини у родовому відмінку*, ***без*** + *іменник чоловічого роду однини у родовому відмінку*, ***до*** + *іменник чоловічого роду однини у родовому відмінку*, у прислівниковий клас одиниць: ***в захваті, без жартів, без ладу, до вподоби***: *— О, яка великодушність, графе! Який широкий жест! Я, далебі,* ***в захваті...*** (Аліков Ю., Капустян В.); *Тому що Львів — магічне місто. Окей,* ***без жартів*** (В. Винниченко); *Біля одної праворуч, спертий на лікті, напів сидить, напів лежить Довбуш. Посередині на муравї розтаборилась як-небудь,* ***без ладу,*** *ватага опришків* (Б. Антонич); *Але їм це, здається,* ***до вподоби*** (А. Азімов).

Відзначимо, що нові смисли у прийменниково-іменникових конструкцій виникають у тих випадках, коли вони здатні утворювати окремі синтагми, адже смислове навантаження, зокрема у двослівних синтагмах, переважно падає на другу частину, якою є повнозначне слово, семантика котрого і набуває певних зрушень. Звичайно, такі значеннєви зсуви не відбуваються раптово, потрібен певний час. Це, наприклад, засвідчують дані дослідження, здійснені М. М. Пещак, яка зафіксувала, що двослівні прийменниково-іменникові синтагми ще з XIV ст. об'єднували в собі ті сполучення прийменника й іменника, з яких у сучасній українській мові утворилися прислівникові еквіваленти, зокрема прийменника ***безо*** та іменника в родовому відмінку, прийменників ***на, въ (во)*** та іменників у знахідному відмінку тощо [2, с. 89].

Проте, щоб встановити, як впливає на формування прислівникового статусу синтагматичне оточення або позиція прислівникового еквівалента у синтагмі, необхідним у подальших дослідженнях буде з'ясування й особливостей взаємодії усіх видів зв'язку її елементів, якими є граматичні, структурні, інтонаційні та смислові. Цьому, беззаперечно, сприяла б синтагматична розмітка текстів корпусу, яка б дозволила не вручну, а автоматично виявити усі наявні синтагми з еквівалентами слова. На сьогодні, на жаль, синтаксична розмітка корпусів є поки що екзотикою. Проте робота у цьому напрямі буде серйозним проривом у корпусних технологіях, а також встановленні й оформленні нових лінгвістичних фактів. Оскільки досвід показує, що найефективнішим способом представлення лінгвістичних знань є лексикографічні системи у вигляді словників, то при просуванні у цьому напрямі, зокрема і створенні словника синтагм, варто очікувати значного прориву у галузі комп'ютерної лексикографії.

Література

1. Лучик А. А. Прислівникові еквіваленти слова в українській мові / А. А. Лучик. — Katowice: Wydwo US, 2009. — 169 с.; Лучик А. А. Еквіваленти слова як предмет мовознавчих досліджень / А. А. Лучик. // Вісник Донецького університету. Серія Б.: Гуманітарні науки. — Донецьк: Донеччина, 2001. — С. 36–42.
2. Пещак М. М. Стиль ділових документів XIУ ст. (структура тексту) / М. М. Пещак. — К.: Наукова думка, 1979. — 268 с.
3. Рогожникова Р. П. Словарь эквивалентов слова: наречные, служебные, модальные единства / Р. П. Рогожникова. — М.: Рус яз., 1991. — 254 с.
4. Соссюр Ф. де. Курс загальної лінгвістики / Ф. де Соссюр. — К.: «Основи», 1998. — 324 с..
5. Широков В. А. Комп'ютерна лексикографія / В. А. Широков. — К.: Наукова думка, 2011. — 352 с.
6. Українська мова. Енциклопедія / Редкол.: Русанівський В. М., Тараненко О. О., Зяблюк М. П. та ін. — К.: Вид-во «Укр. енцикл.» ім. М. П. Бажана, 2007. — 856 с.

Стратегії й методи вдосконалення автоматичного морфологічного анотування Корпусу української мови

Маргарита Лангенбах

к. філол. н., асистент кафедри української мови та прикладної лінгвістики, Інститут філології Київського національного університету імені Тараса Шевченка, Україна, E-mail:labacompli@gmail.com

The article reviews the typical problems of the dictionary-based part-of- speech tagging. The main attention is focused on the non-recognized words. The experiment is based on the textual samples derived from the Corpus of the Ukrainian Language. The examples were classified by the specific features. The article suggests the strategy of increasing the efficiency of the dictionary-based part-of- speech tagger.

Ключові слова — автоматичний морфологічний аналіз, машинна морфологія, АГАТ, графемний аналіз, словниковий морфологічний аналіз.

Використання автоматичних морфологічних аналізаторів природних мов вже досить поширене у світовій практиці, їх розробка спирається на серйозне теоретичне і практичне підґрунтя, проте для жодної мови світу досі не вдалося укласти цілком досконалу систему граматичного кодування тексту. Як зазначає К.Меннінг, попри всі успіхи в цій сфері, межу точності 97–98% поки що не подолано. Та й ці цифри, за його словами, є до певної міри ідеалізованими [11].