

УДК 81'33

КАМО ГРЯДЕШИ, УКРАЇНЬСЬКА МОВО?

Левченко О. П., Кульчицький І. М., Ліхнякевич І. О.

У статті представлено попередні результати формування дослідницького корпусу публіцистики (на матеріалі електронного видання "Поступ"). Проаналізовано частотність уживання різних частин мови в цьому корпусі. Представлено також дані про найчастотніші слова публіцистики, проаналізовано їх склад.

Ключові слова: корпус, технологічні аспекти створення корпусу, публіцистичний стиль, частотність.

В статье представлены предварительные результаты формирования исследовательского корпуса публицистики (на материале электронного издания "Поступ"). Проанализированы частотность употребления различных частей речи в этом корпусе. Представлены также данные о наиболее частотных словах публицистики, проанализирован их состав.

Ключові слова: корпус, технологические аспекты создания корпуса, публицистический стиль, частотность.

The article presents the preliminary results of forming the publicistic corpus, based on the material of the "Postup" electronic publication. The frequency of use of different parts of speech in this corpus was analyzed. The most frequent words of the publicistic style were also presented and their composition was analyzed.

Key words: corpus, technological aspects of corpus creation, publicistic style, frequency.

Українська мова, як і мова будь-якого етносу, може існувати й розвиватися лише тоді, коли вона повноцінно функціонує в суспільстві. Основні середовища, у яких функціонує мова – це побут та наука. Побутове мовне середовище здебільшого формують засоби масової інформації (паперові та електронні), зовнішня реклама в населених пунктах та вздовж доріг, назви і вивіски соціальних закладів та різного роду оголошення, заклики тощо.

Сучасні дослідження публіцистичного дискурсу проводять у різних аспектах, однак щочастіше науковці звертають увагу на недостатність розвідок, присвячених впливові на мову сучасних засобів комунікації. Так, А. Дамбровська зазначає: "Необхідні подальші дослідження мови в "мережі", яка – ми є свідками цього явища – виразно впливає на сучасну польську мову: орфографію, морфологію, лексику, синтаксис і зв'язаність тексту. Польська мова в інтернеті є мовою усно-писемною. Але ми не знаємо, наскільки ці форми представлені, ми не знаємо ще, як мова сучасних ЗМІ впливає на польську мову. Це нові завдання, що постали перед дослідниками" [7]. Технічні реалії сьогодення, безперечно, впливають на мовлення в різних його проявах, зокрема на так звану комунікацію за допомогою комп'ютера (computer-mediated communication), яка, своєю чергою, накладає відбиток на інші види комунікації [8]. Дослідники сьогодні стверджують, що спілкування за допомогою "малих форм" електронної комунікації SMS-повідомлень, чатів та ін. вимагає від користувача максимально стискати текст, спрощувати його, використовувати короткі слова, заступати емоційно-експресивну лексику емотиконами, що, своєю чергою, позначається на якості інших текстів. Окрім того, симптоматичними є подані у [9] дані, де проаналізовано вживання близько 10 млн слів за період з 1800 до 2008 р., що дало змогу дійти висновку про те, що, незважаючи на численні неологізми, які поповнюють мову, за останні 10–20 років поява нових слів не компенсує зникнення старих. Зазначені науковці вважають, що частково таке явище зумовлене впливом автоматичних систем перевірки правопису, а також тенденцією використовувати вужчий словниковий запас під час електронного листування та зазначають зростання ролі англійської мови в академічному середовищі [9].

Однією з ознак електронних текстів дослідники вважають нетривалість їхнього існування, зауважуючи, що відбувається "девальвація письма і наявне вільніше

ставлення творців електронних текстів до правил, які керують творенням писаних текстів” [8, с. 75]. Нетривалість електронних текстів опосередковано теж зумовлює меншу увагу творця до їхньої якості.

Мета цієї праці – на матеріалі невеликого обсягу апробувати методику дослідження специфіки публіцистичного інтернет-дискурсу щодо різноманітності уживаних вербальних засобів та “якісного” складу найчастотніших слів.

Для аналізу обрано інтернет-видання “Поступ” [2]. Це щоденна львівська інтернет-газета, яка до червня 2006 р. була інтернет-версією газети “Поступ”. З вересня 2006 р. функціонує як самостійне інтернет-видання. Газета складається з низки рубрик, у яких традиційно вміщують відповідну інформацію: “Перша сторінка” – вагомні події політичного життя України; “Погляд” – вагомні події суспільного життя України; “Поступ у Львові” – політичні події у Львові; “Поступ з краю” – інші події життя в Україні; “Поступ у світі” – вагомні політичні події у світі; “Репортаж Поступу” – авторські матеріали на актуальні теми життя в Україні; “Точка зору” – аналітичні матеріали на актуальні теми життя в Україні; “Економічний Поступ” – проблеми української економіки; “Калейдоскоп Поступу” – матеріали туристично-відпочинкового характеру; “Спорт-Поступ” – спортивні події; “Post-Factum” – розмаїття фактів; “Кримінал” – кримінальні події; “КуПол: Новини” – новини львівського життя; “Сьогодні у світі” – останні світові новини; “postup.in.ua” – інформаційні огляди.

У дослідженні взяли участь автори цієї статті та магістри кафедри прикладної лінгвістики НУ “ЛП” випуску 2011 р., які в межах індивідуального завдання з курсу “Комп’ютерно-технологічні аспекти проектування сучасних лексикографічних систем” зібрали матеріал та зробили його первинну обробку. Кожний магістр опрацював один випуск інтернет-газети за 2010 р. У подальшому ми проаналізували виконані завдання і відкинули ті, якість виконання та достовірність результатів яких викликала сумніви. У такий спосіб було сформовано дослідницький корпус (ДК), що охопив матеріали публікацій 17-ти номерів інтернет-газети “Поступ” за період від червня до жовтня 2010 р. До аналізу залучено такі частини мови: іменник, дієслово (дієприкметник, дієприслівник), прикметник, прислівник, займенник. Аналіз лексики ідіолекту окремих авторів проведено не було з огляду на обсяг матеріалу. Загальні кількісні характеристики дослідницького корпусу подано у табл. 1.

Таблиця 1

Кількісні характеристики дослідницького корпусу

Загальна кількість слововживань	180370
Кількість слововживань вибраних частин мови	140164
Загальна кількість різних слів	16766

Для опрацювання матеріалу використано програми “Concordance” [6], “MS Excel” [4], інтегровану лексикографічну систему “Словники України” [1], частотний словник публіцистики (ЧСП) [5], за даними якого в масиві публіцистичних текстів довжиною 300000 слововживань зафіксовано по 20824 різних слова, та приблизні власні дані кількісного розподілу лексики за частинами мови в СУМ-11 [3].

Загальні статистичні дані опрацювання лексики дослідницького корпусу подані у таблицях 2, 3.

Загалом, як свідчать отримані дані, частка тієї чи іншої частини мови в ДК є типовою для публіцистики.

Симптоматичними є також результати проведеної кластеризації частот лексики дослідницького корпусу. З цією метою значення частот було розбито на окремі діапазони: 1–9; 10–99; 100–199; 200–2100. Перший діапазон було розбито на піддіапазони: 1; 2; 3; 4; 5; 6; 7; 8; 9. Наступним кроком для кожної, що досліджують, частини мови підраховано кількість слів, частота яких попала у відповідний діапазон. Результати подано у таблицях 4 та 5.

Таблиця 2

Розподіл слів дослідницького корпусу за частинами мови та вжитком

Частина мови	Кількість			
	Різних слів	Частотність вжитку		
		Абсолютна	Середня	Відносна
Іменник	7381	65054	9	36,10 %
Дієслово	4412	26732	6	14,80 %
Прикметник	3345	21810	7	12,10 %
Прислівник	1127	11691	10	6,50 %
Займенник	108	14877	138	8,20 %
Решта	393	40206	102	22,30 %
Разом:	16766	180370	11	

Таблиця 3

Співвіднесена частотність слів дослідницького корпусу та СУМу

Частина мови	Частотність слів				
	ДК		СУМ		ДК до СУМ
	Абсолютна	Відносна	Абсолютна	Відносна	
Іменник	7381	44,00 %	54900	41,40 %	13,40 %
Дієслово	4412	26,30 %	48200	36,40 %	9,20 %
Прикметник	3345	20,00 %	21900	16,50 %	15,30 %
Прислівник	1127	6,70 %	6600	5,00 %	17,10 %
Займенник	108	0,60 %	130	0,10 %	83,10 %
Решта	393	2,30 %	780	0,60 %	50,40 %
Разом:	16766		132510		12,70 %

За отриманими результатами, ядро найчастотніших слів становить: іменників – 3183, дієслів – 2225, прикметників – 1610, прислівників – 474, займенників – 25 (разом: 7492), що не свідчить про лексичну різноманітність публіцистичного дискурсу чи колективного ідіолекту авторів текстів (нагадаємо, що цей ідіолект складається з 16 766 одиниць).

Таблиця 4

Розподіл частот слів дослідницького корпусу за діапазонами

Частина мови	Частотність слів				
	ДК		СУМ		ДК до СУМ
	Абсолютна	Відносна	Абсолютна	Відносна	
Іменник	7381	44,00 %	54900	41,40 %	13,40 %
Дієслово	4412	26,30 %	48200	36,40 %	9,20 %
Прикметник	3345	20,00 %	21900	16,50 %	15,30 %
Прислівник	1127	6,70 %	6600	5,00 %	17,10 %
Займенник	108	0,60 %	130	0,10 %	83,10 %
Решта	393	2,30 %	780	0,60 %	50,40 %
Разом:	16766		132510		12,70 %

Таблиця 5

**Розподіл частот слів дослідницького корпусу за піддіапазонами
діапазону 1–9**

Частина мови	Діапазон								
	1	2	3	4	5	6	7	8	9
Іменник	3183	1107	596	384	291	214	146	135	113
Дієслово	2225	672	359	207	138	114	91	61	55
Прикметник	1610	540	254	169	106	80	57	57	37
Прислівник	474	141	92	55	43	31	37	15	19
Займенник	25	7	7	3	4	2	1	1	3

Далі проаналізуємо найчастотніші слова дослідницького корпусу. Свого часу А. Вержбицька дійшла висновку про те, що "стрижневими словами" є звичайні, а не маргінальні слова; вони можуть часто вживатися в семантичних доменах (домени емоцій, домени покарання); ці слова є центрами цілих фразеологічних кластерів, як, наприклад, російське слово душа; часто вживаються в прислів'ях, висловах, популярних піснях, назвах книг [10]. Такий підхід призвів до абсолютизації критерію частотності в низці досліджень. Не перебільшуючи важливість частотності щодо піднесення певного концепту до культурно значущого, зауважимо, що симптоматичним є власне сам перелік найчастотніших слів. У таблицях 6–8 подано найчастотніші слова.

Таблиця 6

Загальний список слів ДК, частота яких перевищує 500

Слово	Частина мови	Частота вжитку			
		ДК		ЧСП	
		Абс.	Відн.	Абс.	Відн.
Бути	дієслово	2345	1,30 %	5384	1,79 %
Який	займенник	1824	1,01 %	6698	2,23 %
Цей	займенник	1660	0,92 %	3979	1,33 %
Він	займенник	1359	0,75 %	8274	2,76 %
Україна	іменник	1271	0,70 %	27316	9,11 %
Той	займенник	1056	0,59 %	4476	1,49 %
Свій	займенник	916	0,51 %	18040	6,01 %
Вони	займенник	794	0,44 %	21586	7,20 %
Рік	іменник	786	0,44 %	24050	8,02 %
Вона	займенник	778	0,43 %	13988	4,66 %
Ми	займенник	767	0,43 %	19541	6,51 %
Такий	займенник	715	0,40 %	17959	5,99 %
Влада	іменник	640	0,35 %	9672	3,22 %
Я	займенник	617	0,34 %	17615	5,87 %
Могти	дієслово	558	0,31 %	8171	2,72 %
Мати	дієслово	530	0,29 %	14310	4,77 %
Український	прикметник	523	0,29 %	11152	3,72 %
Вибори	іменник	503	0,28 %	3177	1,06 %
Час	іменник	503	0,28 %	11509	3,84 %

Серед наведених прикладів звертає на себе увагу відсутність слів, які вербалізують "ціннісні" концепти. На нашу думку, належить вважати важливим чинником частоту вживання слова у прецедентних текстах, однак з огляду на роль засобів масової комунікації на суспільство, їм теж певною мірою властива ознака

прецедентності. Як засвідчує навіть попередній аналіз, така конфігурація найчастотніших слів, зокрема іменників та прикметників, переконує у тому, що відбувається нівелювання уваги до духовних цінностей.

Таблиця 7

Список іменників ДК, частота яких перевищує 200

Слово	Частота вжитку			
	ДК		ЧСП	
	Абсолютна	Відносна	Абсолютна	Відносна
Україна	1271	0,70 %	27316	9,11 %
Рік	786	0,44 %	24050	8,02 %
Влада	640	0,35 %	9672	3,22 %
Вибори	503	0,28 %	3177	1,06 %
Час	503	0,28 %	11509	3,84 %
Партія	464	0,26 %	4339	1,45 %
Президент	456	0,25 %	6181	2,06 %
Країна	366	0,20 %	10677	3,56 %
Людина	340	0,19 %	4919	1,64 %
Питання	337	0,19 %	6749	2,25 %
Закон	323	0,18 %	7224	2,41 %
Слово	314	0,17 %	3682	1,23 %
Голова	302	0,17 %	4030	1,34 %
Держава	302	0,17 %	9485	3,16 %
Рада	299	0,17 %	3030	1,01 %
Ієрархія	268	0,15 %	96	0,03 %
Проблема	257	0,14 %	5522	1,84 %
Мова	255	0,14 %	3194	1,06 %
Росія	237	0,13 %	3702	1,23 %
День	232	0,13 %	3555	1,19 %
Депутат	229	0,13 %	2831	0,94 %
Справа	225	0,12 %	5193	1,73 %
Міністр	219	0,12 %	2853	0,95 %

Звертає на себе увагу значна різниця частотності уживання слів *Україна* в електронному виданні "Поступ" (відносна частота 0,70 %), *український* (0,29 %) та даними, уміщеними у ЧСП (9,11 %) і відповідно (3,72 %). Текстові масиви ЧСП, сформовано на основі видань "Урядовий кур'єр", "Голос України", "Сільські вісті", "Культура і життя", "Україна молода", "Літературна Україна", "Молодь України", "Вісті з України", "Республіка", "Золоті ворота" за 1995 р.²

² На жаль, не вдалося використати Частотний словник українського публіцистичного стилю http://www.mova.info/freqcard2.aspx?l1=178&sl=pb4_plt, щоб простежити відмінності та динаміку слововживання.

Таблиця 8

Список прикметників ДК, частота яких перевищує 130

Слово	Частота вжитку			
	ДК		ЧСП	
	Абсолютна	Відносна	Абсолютна	Відносна
Український	523	0,29 %	11152	3,72 %
Політичний	462	0,26 %	7358	2,45 %
Місцевий	403	0,22 %	3727	1,24 %
Новий	401	0,22 %	8699	2,90 %
Російський	299	0,17 %	3860	1,29 %
Львівський	223	0,12 %	1048	0,35 %
Наївний	217	0,12 %	88	0,03 %
Невідомий	216	0,12 %	241	0,08 %
Державний	190	0,11 %	8292	2,76 %
Головний	182	0,10 %	3673	1,22 %
Великий	155	0,09 %	5145	1,72 %
Верховний	154	0,09 %	2709	0,90 %
Останній	139	0,08 %	4149	1,38 %
Національний	137	0,08 %	4839	1,61 %
Народний	133	0,07 %	2059	0,69 %
Виборчий	131	0,07 %	1564	0,52 %

Зіставлення здобутих результатів засвідчує факт суттєвого зниження частотності вживання слів *Україна, український* у публіцистичних текстах. Належить зауважити, що відмінності можуть бути пояснені локальністю видання "Поступ", однак тематика цього ЗМІ передбачає огляд світових та загальноукраїнських подій. У подальшому передбачено уточнення результатів за допомогою контент-аналізу.

Література

1. Інтегрована лексикографічна система "Словники України" / НАН України, Український мовно-інформаційний фонд, 2001–2004. – Версія 1.4. СУ
2. Поступ [Електронний ресурс]. – Режим доступу: <http://postup.brama.com/>. – Назва з екрана.
3. Словник української мови : в 11 т. / за заг. ред. І. К. Білодіда. – К. : Наук. думка, 1970–1980.
Т. 1–11. СУМ–11.
4. Уокенбах Дж. Microsoft Excel 2010. Библия пользователя / Дж. Уокенбах. – М.: ООО "И. Д. Вильямс", 2011. – 912 с.
5. Частотний словник публіцистики [Електронний ресурс]. – Режим доступу: <http://www.mova.info/cfq.aspx?fdid=publicist>. – Назва з екрана.
6. Concordance [Електронний ресурс]. – Режим доступу: <http://www.concordancesoftware.co.uk/index.htm>. – Назва з екрана.
7. Dąbrowska 2008 – Dąbrowska A. Współczesne problemy lingwistyki kulturowej [Електронний ресурс] / Anna Dąbrowska. – 2008. – Режим доступу: <http://filologia-polska.pwa.edu.pl/?page=publikacja&id=68>. – Назва з екрана.
8. Grzenia J. Komunikacja językowa w Internecie / Jan Grzenia. – Warszawa : PWN, 2008. – 214 s.
9. Statistical Laws Governing Fluctuations in Word Use from Word Birth to Word Death. Alexander M. Petersen, Joel Tenenbaum, Shlomo Havlin & H. Eugene Stanley. Scientific Reports 2, Article number: 313 [Електронний ресурс]. – Режим доступу: <http://www.nature.com/srep/2012/120315/srep00313/full/srep00313.html#f1>. – Назва з екрана.