

УДК 811.161.2'1'324'38

СУЧАСНІ МЕТОДИ ДОСЛІДЖЕННЯ МОВИ ПИСЬМЕННИКА У СЛОВ'ЯНОЗНАВСТВІ

Соломія БУК

Львівський національний університет імені Івана Франка
вул. Університетська, 1, Львів, 79000, e-mail: solomija@gmail.com
Кафедра загального мовознавства

У статті висвітлено сучасні методи дослідження мови письменника у слов'янознавстві, що набули особливого поширення із розвитком комп'ютерних технологій, а саме корпусний і статистичний. Здійснено аналіз “Словника Карела Чапека” та “Статистичного словника Достоевського”. Отримані висновки екстрапольовано на проєкт лінгвостатистичної параметризації творчості І.Франка.

Ключові слова: методи дослідження, мова письменника, ідіолект, частотний словник, корпус текстів.

Методологія та методи сучасного мовознавства становлять предмет посиленого інтересу в науковій сфері. Із такою тематикою відбуваються міжнародні конференції¹, таку назву мають монографії², розділи у колективних працях та збірниках, а також окремі предмети, які викладаються студентам у вищій школі.

Це пояснюється, зокрема, тим, що з розвитком комп'ютерних технологій з'явилися нові сучасні принципи та методи дослідження тексту, зокрема мови письменників, серед яких – статистичний та корпусний³, що привертають особливу увагу лінгвістів.

Для багатогранного дослідження авторських ідіолектів у письменницькій та статистичній лексикографії лінгвісти вдаються до кількісного опису доробку письменників, що дає точну інформацію про особливості функціонування лексики в їхній творчості. Просте-

¹ Міжнародна науково-практична конференція “Методи лінгвістичних досліджень”. м. Слов'янськ. Донецька обл. 17–20 листоп. 2010 р. <http://mova.interdon.net/17-20-11-2010-metodi-lingvistichnih-doslidzhen.html>, <http://www.slav.dpu.dn.ua/index.php/novyny/58-news/274-2011-03-15-12-02-51> [Цит. 08.12.2011]; Modern Methods in Linguistics held in honour of the anniversary of Prof. Gabriel L. Altmann, October 23rd and 24th, 2006, Budmerice Castle, Slovakia.

² Методи аналізу тексту: Збірник наукових праць = Methods of text analysis: Omnibus volume. Chernivtsi, 2009. Чернівці, 2009; Сухорольська С.М., Федоренко О.І. Методи лінгвістичних досліджень: Навч. посібник. Львів, 2009; Апресян Ю.Д. Иден и методы современной структурной лингвистики. Москва, 1966.

³ Фрумкина Р.М. Статистические методы изучения лексики. Москва, 1964; Mayenowa M. R. (red.) *Poetyka i matematyka*. Warszawa, 1965; Guiraud P. Zagadnienia i metody statystyki językoznawczej. Warszawa, 1966; Мустрик И. Математико-статистические методы в стилистике // Вопросы языкознания. 1967. № 3. С.42–52; Алексеев П. М. Методика квантитативной типологии текста. Ленинград, 1983; Тулова Ю. А. Проблемы и методы системно-квантитативных исследований лексики. Таллин, 1987; Partee B. H., Aliceter M., Wall R. E. Mathematical Methods in Linguistics. Dordrecht–Boston–London, 1990; Шайкевич А. Я. Количественные методы в языкознании // Лингвистический энциклопедический словарь. Москва, 1990. С.231–232; Математические методы в языкознании: (Обзор материалов конф. COLING-88) [Ю. Н. Марчук]. Москва, 1990; От Нестора до Фонвизина. Новые методы определения авторства. Москва, 1994; Перебийніс В. С. Статистичні методи для лінгвістів: Навчальний посібник. Вінниця, 2002; Matrix R. A statistical method and software tool for linguistic analysis through corpus comparison. Lancaster University PhD thesis, 2003; Левинський В. В. Лексична полісемія та квантитативні методи її дослідження // Мовознавство. 2003. №4. С.17–15; Köhler R., Altmann G. Aims and Methods of Quantitative Linguistics // Problems of Quantitative Linguistics. Chernivci, 2005. P. 12–42; Quantitative Linguistics. Berlin–New York, 2007. V. 62: Exact methods in the study of language and text: dedicated to Professor Gabriel Altmann on the occasion of his 75th birthday / Ed. by P. Grzybek and R. Köhler.

жується також чітка тенденція до використання корпусів тексту (КТ) для укладання словників мови автора та частотних словників (ЧС)⁴. За подібними принципами, хоч і з різним рівнем деталізації, описано творчість багатьох слов'янських авторів: у Чехії – К. Чапека⁵, Б.Грала⁶, О.Бжезіни⁷, у Польщі – К.Бачинського⁸, у Сербії – М.Павича⁹, С.Велмар-Янкович¹⁰, Д.Албахарія¹¹, М.Данойлича¹², у Болгарії – Н.Вапцарова¹³, в Україні – Т.Шевченка¹⁴, Г.Сковороди¹⁵, Лесі Українки¹⁶, І.Франка¹⁷, поетів другої половини ХХ ст. І.Драча, М.Вінграновського, Л.Костенко, І.Калинця та ін.¹⁸, у Росії – О.Пушкіна¹⁹, О.Грибоедова²⁰, Ф.Тютчева²¹, М.Лермонтова²², Ф.Достоевського²³, А.Чехова²⁴, Максима Горького²⁵, Л.Андрєєва²⁶, Б.Пастернака²⁷, О.Мандельштама²⁸, О.Твардовського²⁹, А.Вознесенського³⁰ та інших письменників.

⁴ Бук С. Слов'янський досвід укладання частотних словників мови письменника // Проблеми слов'янознавства. Вип. 60. Львів, 2011. С.217–224.

⁵ Slovník Karla Čapka. Praha, 2007. + CD

⁶ Cvrček V., Čermák F. Slovník Bohumila Hrabala. Praha, 2009. + CD

⁷ Holman P. Frekvenční slovník básnického díla Otokara Březiny. Köln am Rhein–Weimar–Wien, 1993.

⁸ Balowski M. Lista frekwencyjna poezji, prozy i dramatu Krzystofa Kamila Baczyńskiego. Prochowice, 1997.

⁹ Vasić S. Polazne osnove novije srpske proze. Knjiga II. Hazarski rečnik Milorada Pavića: frekvencijski rečnik. Beograd, 1998.

¹⁰ Vasić S. Polazne osnove novije srpske proze frekvencijski rečnik romana Bezdno Svetlane Velmar-Janković. Beograd, 2000.

¹¹ Vasić S. Polazne osnove novije srpske proze frekvencijski rečnik romana Mamac Davida Albaharija. Beograd, 2001.

¹² Vasić S. Polazne osnove novije srpske proze frekvencijski rečnik romana Oslobođioci i izdajnici Milorada Danojlića. Beograd, 2002.

¹³ Честотен речник на Вапцаровата поезия / Авт. кол.: Г. В. Крилова, А. А. Азарова, Е. А. Захаревич и др. Велико Търново, 1996.

¹⁴ Конкорданція поетичних творів Тараса Шевченка: У 4-х т. Торонто, 2001.

¹⁵ Online Concordance to the Complete Works of Hryhorii Skovoroda <<http://www.arts.ualberta.ca/~ukr/skovoroda/NEW/index.php>> [Цит. 08.12.2011].

¹⁶ Алексієнко Л. А., Дарчук Н. П. Принципи створення параметризованої бази даних за поетичними текстами Лесі Українки // Леся Українка і сучасність (До 130-річчя від дня народження Лесі Українки): Зб. наук. праць. Луцьк: 2004. С.344–352.

¹⁷ Бук С. Корпус текстів Івана Франка: спроба визначення основних параметрів // Прикладна лінгвістика та лінгвістичні технології: MegaLing-2006: Зб. наук. праць. Київ, 2007. С.72–82; Бук С. Квантитативна параметризація текстів Івана Франка: спроба проекту // Іван Франко: Студії та матеріали: Збірник наукових статей. Львів, 2010 (у друці).

¹⁸ Частотний словник сучасної поетичної української мови / Наук. кер. Н. П. Дарчук. Доступно з <<http://www.mova.info/article.aspx?l1=89&DID=77>> [Цит. 08.12.2011].

¹⁹ Баевский В. С. Справочные труды по поэзии Пушкина и его современников. Доступно з <<http://feb-web.ru/feb/pushkin/serial/v91/v91-065-.html?cmd=p.htm>> [Цит. 08.12.2011].

²⁰ Алфавитно-частотный и частотный словари языка комедии А.С. Грибоедова "Горе от ума" / Сост. А. В. Королькова. Смоленск, 1996; Словарь языка А. С. Грибоедова. Т. 1: А-З / Сост. А. Е. Поляков. Москва, 2008.

²¹ Егоров Б. Частотный словарь лексики Тютчева и его поэтическое мировоззрение // Тютчевский сборник. К двухсотлетию со дня рождения. Кракów, 2004. С.13–19.

²² Частотный словарь языка М. Ю. Лермонтова // Лермонтовская энциклопедия. Москва, 1981. С.717–774.

²³ Шайкевич А. Я., Андрущенко В. М., Ребецкая Н. А. Статистический словарь языка Достоевского. Москва, 2003.

²⁴ Гребенников А. О. Частотный словарь рассказов А. П. Чехова. Санкт-Петербург, 1999.

²⁵ Алексеев П. Частотный словарь автобиографической трилогии М. Горького. Санкт-Петербург, 1996.

²⁶ Гребенников А. Частотный словарь рассказов Л. Н. Андреева. Санкт-Петербург, 2003.

²⁷ Романова И. В. Частотный словарь "Стихотворений Юрия Живаго" Б. Л. Пастернака // Русская филология: Учен. зап. Смоленск, 1997. Вып. 3. С.28–55.

²⁸ Частотный словарь лирики О. Мандельштама: субъектная дифференциация словоформ. Ижевск, 2003.

Особливою ґрунтовністю розроблення теоретичної бази та її практичного втілення вирізняються проекти лінгвостатистичного аналізу мови К. Чапека та Ф. Достоевського, що дає підстави для ретельного дослідження їх макро- та мікроструктури, засад та принципів укладання.

Засвоєння й адаптація сучасних методів дослідження тексту, зокрема мови письменника, окремим об'єктом наукового дослідження у слов'янознавстві ще не були. Детальний аналіз Словника Карела Чапека і Статистичного словника Федора Достоевського в україністиці здійснено вперше. Вказане свідчить про актуальність та новизну статті.

Словник Карела Чапека³¹ започаткував у Чехії серію корпусних публікацій під назвою “Корпусна лексикографія”, мета яких – укладання частотних словників важливих періодів та основних авторів чеської культури.

У праці здійснено комплексний аналіз ідіолекту одного з найбільших чеських письменників і мислителів, описано фонологічні, морфологічні, словотворчі, лексичні особливості його мови, фразеологізми та ідіоми, авторські метафори та ін.

Книга складається з двох вступів, студій низки авторів “Словник Карела Чапека: його лексеми та номінації” і власне самого частотного словника усіх творів письменника. Завершують видання добірка Чапеківих “Думок, афоризмів, висловлювань”, бібліографія і додатки, що подають різносторонню статистичну структуру лексики К. Чапека, оригінальні ілюстрації та вичерпний погляд на деякі риси мови автора.

У загальному вступі викладено принципи та методи укладання Словника: його створив авторський колектив чеського народного корпусу філософського факультету Карлового університету в Празі на основі повного КТ К. Чапека обсягом 2 315 733 слововживань із використанням нових комп'ютерних технологій. Корпус має морфологічне маркування, що забезпечує автоматичну лематизацію слів (зведення словоформ до початкової, словникової форми). Розрізнення омонімії в КТ базувалося на стохастичних методах. Автоматичне опрацювання тексту доповнювалося ручною верифікацією, додатковим завданням якої було також зведення до одної форми різних варіантів одного слова.

Другий вступ – “Ресурси мови Карела Чапека” – автори визначили як “нелінгвістичні примітки”, оскільки він містить інформацію про біографічні факти письменника, які мали вплив на становлення його мовної особистості, на творення його ідіолекту.

Розділ “Словник Карела Чапека: його лексеми та номінації” є багатогранною колективною працею над лексиконом письменника. Він містить якісні фонологічні та морфологічні характеристики прямих та переносних, термінологічних та прагматичних номінацій автора, власних та загальних назв, а також фразем та ідіом, авторських метафор, колокацій.

Основну частину книги займає частотний словник Карела Чапека. Він має чотири частини: словник звичайних слів, гапаксів (слів, вжитих у тексті один раз – 19 585), власних назв (12 442) і скорочень (314), а доповнюють його розділ “Статистичні аспекти мови Карела Чапека. Особливості його лексикону” і додатки. Кожна лексема у словнику має ранг і кількісну характеристику у шести стилях: проза, драма, публіцистика, поезія, наукова література, кореспонденція (див. рис. 1).

²⁹ Шаповалов Б. С. Частотный словарь послевоенной лирики А.Т. Твардовского // Русская филология. Учен. зап. Смоленск. пед. ун-та. 1999. Смоленск, 1999. С.206–248.

³⁰ Баевский В. С. Частотная структура лексики А. Вознесенского (“Антимиры”, лирика) // Русская авторская лексикография XIX–XX веков / ИРЯ им. В.В. Виноградова РАН. Москва, 2003. С.111–112.

³¹ Slovník Karla Čapka / Hl. ed. F. Čermák. Praha, 2007. + CD

	celk. frekv.	proza	drama	publicistika	poezie	odborná	koresp.		celk. frekv.	proza	drama	publicistika	poezie	odborná	koresp.
afektový	2	0	0	1	0	1	0	airdalka	3	0	0	2	0	0	1
*aféra	93	40	0	40	0	0	13	ajej	2	2	0	0	0	0	0
• silvestrovská aféra, skandální aféra								aji	5	3	0	2	0	0	0
aférismus	2	0	0	1	0	0	1	ajimez	2	2	0	0	0	0	0
aférista	3	3	0	0	0	0	0	akácie	4	2	0	2	0	0	0
aforismus	4	2	0	1	0	0	1	akademický	4	0	0	4	0	0	0
aforistický	3	0	0	2	0	0	1	akademický	45	3	0	34	0	7	1
aforistický	5	0	0	5	0	0	0	akademie	131	8	3	91	0	0	29
africký	23	9	0	12	0	0	2	• akademie věd, česká akademie							
agar	2	0	0	2	0	0	0	akademik	3	1	0	2	0	0	0
<i>rosolovitá látka z mořských řas, užívaná v mikrobiologii jako živné prostředí pro bakterie (také agar-agar)</i>								akademismus	14	1	0	13	0	0	0
agenda	5	0	0	3	1	0	1	akát	12	8	0	4	0	0	0
agens	2	1	1	0	0	0	0	akátový	6	1	0	5	0	0	0
agent	59	25	1	24	0	0	9	akce	100	11	0	62	1	14	12
• americký agent, obchodní agent								• ošacovací akce							
agentura	31	6	0	19	0	0	6	akcent	17	0	0	17	0	0	0
• Agentura centrum, divadelní agentura								akcentovaný	3	0	0	3	0	0	0
agilní	3	2	0	1	0	0	0	akcentovat	5	0	0	2	0	1	2
agitace	9	1	1	7	0	0	0	akceptovat	7	1	0	6	0	0	0
agitáční	7	0	0	7	0	0	0	akcesorium	2	1	0	1	0	0	0
agitátor	4	1	1	2	0	0	0	<i>přídavný n. doplňkový objekt, fakt</i>							
agitovat	10	0	0	10	0	0	0	akcesorní	7	0	0	1	0	6	0
aglomerace	3	1	0	2	0	0	0	<i>vedlejší, nepodstatný, dodatekový</i>							
agnosticismus	3	0	0	2	0	1	0	akcie	22	16	0	6	0	0	0
agón (řec.)	2	0	0	2	0	0	0	akcionář	5	2	0	3	0	0	0
<i>závod, zápas</i>								akciový	15	2	1	12	0	0	0
agónie	17	11	1	5	0	0	0	akciz	2	0	0	0	0	0	2
agonizovat	2	1	0	1	0	0	0	<i>dřívejší nepřímá daň na některé spotřební předměty, zvl. potraviny; místo výběru této daně zvl. na hranicích města ap.</i>							
agonizující	4	1	0	3	0	0	0	akční	7	3	0	3	0	1	0
agora	4	4	0	0	0	0	0	aklimatizace	2	1	0	1	0	0	0
agrarisismus	2	0	0	2	0	0	0	aklimatizační	3	0	0	3	0	0	0
agrární	26	1	0	24	0	0	1	aklimatizovat (se)	2	2	0	0	0	0	0
agrárník	20	1	0	15	2	0	2	ako (slov.)	3	0	0	3	0	0	0
agreement (angl.)	2	2	0	0	0	0	0	akomodace	3	0	0	0	0	3	0
<i>doboda</i>								akorát	9	2	0	2	0	0	5
agregát	5	4	0	1	0	0	0	akord	29	6	0	19	0	3	1
agresivní	5	0	0	4	0	0	1	akr	8	8	0	0	0	0	0
agresivnost	2	0	0	2	0	0	0	akreditiv	5	3	0	1	0	0	1
aguardiente (špan.)	2	2	0	0	0	0	0	<i>příkaz bance k výplatě jisté sumy peněz určité pověřené osobě</i>							
<i>pálenka</i>								akribie	2	0	0	2	0	0	0
*ah	60	16	34	1	0	0	9	<i>přesnost, pečlivost, zvlášť ve vědecké práci</i>							
*aha	242	151	42	46	0	0	3	akrobat	3	0	0	2	0	1	0
ahah	6	6	0	0	0	0	0	akrostichon	2	0	0	2	0	0	0
ahaha	4	4	0	0	0	0	0	<i>báseň, v níž první nebo poslední básky (někdy slabiky, slova) versů tvoří slovo nebo větu (také akrostich)</i>							
ahahah	2	1	0	1	0	0	0	*akt	228	29	11	134	1	40	13
ahi	5	0	5	0	0	0	0	• poslední akt, první akt, příslušný akt, třetí akt, úřední akt, vyřizený akt							
*ach	493	215	87	90	38	5	58	akta	27	7	3	15	0	0	2
• ach bite, ach mordije								aktér	4	0	0	4	0	0	0
achajský	4	2	0	2	0	0	0	aktivismus	16	0	0	12	0	2	2
achtich	3	2	1	0	0	0	0	aktivista	4	0	0	4	0	0	0
achillovský	4	1	0	3	0	0	0	aktivita	34	1	0	19	0	14	0
acht (něm.)	2	1	0	0	0	0	1	• aktivita básnického díla, duchovní aktivita							
<i>klatba</i>								aktivně	11	1	0	7	0	3	0
airdale	11	0	0	11	0	0	0								

Рис. 1. Ранг і кількісна характеристика лексем у шести стилях (за: Slovník Karla Čapka. Praha, 2007)

Словник дає коротку дефініцію маловідомих слів та абreviатур, вказівку на джерело іншомовних слів.

Розділ “Статистичні аспекти мови Карела Чапека. Особливості його лексику” містить низку статистичних параметрів лексику письменника, обчислених на основі ЧС,

які неможливо було б отримати традиційним ручним способом опрацювання тексту, а саме: обсяг тексту у словоформах, обсяг словника у лексемах, багатство словника, кількість слів за частинами мови, розподіл слів за жанрами та роками тощо. Зокрема, обчислено, скільки слів К. Чапек написав кожного року з 1905 по 1908 (найбільше текстів створено у 1937, 1933 і 1934 рр.), а також кількість різних лексем на 1 000 різних слововживань для кожного року: найбагатшими виявилися 1907, 1911 та 1916 рр. (див. рис. 2) Найбільшими за обсягом є публіцистика (50 847 слововживань), далі – проза (36 638), кореспонденція (17 621), наукові тексти (9 537), драма (7 139) та поезія (3 601). Обчислений відсоток різних лексем на 1 000 слововживань уможливило спостерегти, що найбагатшою є лексика поезії (показник вищий, ніж 200), інші ж жанри суттєво їй поступаються (показник близький до 50).

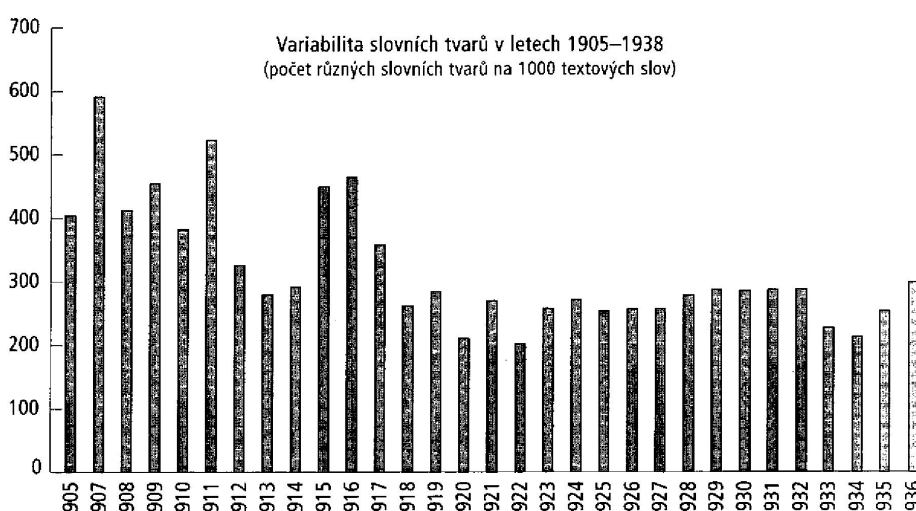


Рис. 2. Статистичні параметри лексики письменника (за: Slovník Karla Čapka. Praha, 2007)

Багатство словника К. Чапека, порівняно з іншими чеськими письменниками (Гашеком, Грабалом, Кундерою та ін.), розкрито у вигляді таблиці. Важливо, що автори застосували найновішу методіку квантитативної лінгвістики для визначення межі між самостійними (семантичними) та службовими (синсемантичними) частинами мови: знайшли для текстів точки *h* (*h*-point) і *k* (*k*-point) (див. працю І. Попеску та ін.³²).

Статистичну інформацію про тексти К. Чапека представлено також у зіставленні із КТ сучасної чеської мови (<http://ucna.ff.cuni.cz>), оскільки саме так можна виявити відмінності мови автора від сучасної: слова з вищою частотністю, ніж у загальномовному корпусі, у Словнику мають спеціальну позначку.

Список літератури, що охоплює близько 170 позицій корпусних, лінгвостатистичних, лексикографічних та інших типів праць, підтверджує глибину теоретичного й практичного опрацювання проблеми.

Додатки до ЧС мають різноаспектні розділи: “Неологізми в текстах Карела Чапека”, “Конкорданс лексеми *politický* / *політичний*”, “Колокаційна карта мови Карела Чапека”, “Мовні експерименти та каламбури із запереченням”, “Фраземи Карела Чапека”, що логічно завершують це комплексне дослідження. Диск з електронним корпусом усіх творів мислителя заохочує реципієнта до власних досліджень його мови та значно спрощує їх.

Ф. Достоєвському у сучасній російській лексикографії приділена особлива увага: 1998 р. вийшов диск “Ф. М. Достоевский. Тексты. Исследования. Инструментарий.

³² Popescu I.-I. et al. Word frequency studies. Berlin–New York, 2009.

(1998)”, підготовлений Інститутом російської мови РАН, що містить частотні словники до художніх творів письменника. Частотний словник його публіцистики доступний в Інтернеті³³. 2001 р. надруковано Словник його мови³⁴, а 2003 р. видано “Статистичний словник мови Достоевського”³⁵, укладений в Інституті російської мови РАН, який і став об’єктом нашого огляду. Як зазначають автори, його завдання – не тільки доповнити словник мови письменника³⁶, а й представити лексику Ф. Достоевського у кількісному вигляді.

Словник складається з двох частин: одностомної книги і компакт-диску. Він охоплює всю лексику трьох основних жанрів письменника: художніх, критичних і публіцистичних творів та епістолярію. Працю створено на КТ з 2,9 млн. слововживань, і вона, як зазначено в анотації, значно більша обсягом за будь-які інші частотні словники російської мови. За ступенем лінгвістичної диференціації Словник унікальний і в світовому масштабі. У таблицях Словника лексику Достоевського поділено за основними жанрами (див. рис. 3), за мікрожанрами (авторська мова, ремарки, пряма мова тощо), за окремими персонажами. Останній поділ дає змогу виявити й описати мовну особистість кожного персонажа³⁷. У практику лексикографії вводяться таблиці бінарних словосполучень й таблиці текстуальних зв’язків. Словник адресовано філологам і всім, хто любить творчість Достоевського.

Всего графических слов (в тыс.)	Всего о				Всего о	X	K	П
2889	183	52	53		2889	183	52	53
	5	4	1			5	4	1
...			
аббат	16	15	1	Австрия	74	1	71	2
аббатик	2	2		автобиография	6	3	2	1
абвахта	9	7	1	автограф	4			4
абerrация	4		4	автодафе	3	3		
абонироватьс я	2			автомат	2	2		
абордировать	3	3		автоматически й	2	1	1	
абрютировать	2	2		автономия	5		5	
абсентеизм	2		2	автономный	2		1	1
абсентеист	3		3	автор	230	34	16	30
							6	
абсолютно	3		3	авторитет	70	19	47	4
абсолютный	13	5	7	авторитетнее	2	1	1	
абструзный	2		2	авторитетно	2	1	1	
абсурд	25	2	18	авторитетный	5	1	3	1
абсурдный	3		1	авторский	9	3	2	4

X – художня проза, K – критика, П – публіцистика.

Рис. 3. Розподіл лексем за основними жанрами (за: Словарь языка Достоевского. Москва, 2001)

³³ Частотный словарь-конкорданс публицистики Ф. М. Достоевского. Доступно з <<http://dostoevskii.karelia.ru/principis.phtml>> [Цит. 08.12.2011].

³⁴ Словарь языка Достоевского. Лексический строй идиолекта. Москва, 2001.

³⁵ Шайкевич А.Я., Андрющенко В.М., Ребецкая Н.А. Статистический словарь языка Достоевского. Москва, 2003.

³⁶ Словарь языка Достоевского. Лексический строй идиолекта. Москва, 2001.

³⁷ Шайкевич А. Я. Персонажи романа “Идиот” в лингвостатистическом освещении // Достоевский и современность. Материалы IX Международных Старорусских чтений, 1994. Новгород, 1995. С.214–216; Шайкевич А. Я. Компоненты речевой характеристики персонажей (анализ индивидуальных статистических словарей) // Филологический сборник. Москва, 1991. С.393–401.

Щоби з'ясувати, для якого макрожанру характерні ті чи інші слова, автори вводять поняття “статистичної значущості слова” та її формулу, а самі слова називають “лексичними маркерами підкорпусів”. У цьому контексті підкорпусами можуть бути макрожанри, мікрожанри, періоди творчості, окремі тексти, окремі персонажі тощо. Таблиці мікрожанрів відкривають широкі можливості вивчення як самих цих типів текстів, так і стилістичної характеристики лексем. Так, з'ясовано, що слово “гроші” в листуванні має винятково високу статистичну значущість – 31, а це свідчить, що воно характерне власне для епістолярію письменника, а не для художніх чи публіцистичних творів. Формулу застосовують також для виявлення текстуальних зв'язків слів, властивих стилю саме Ф. Достоевського.

Після перших двох таблиць, що містять алфавітний реєстр лексем з абсолютними й відносними частотами за основними жанрами, розміщено таблиці, що подають ранговий список 100 найчастотніших лем сумарно, а також окремо за кожним із вказаних жанрів. Далі йдуть таблиці з інформацією про 40 найчастотніших іменників, дієслів, прикметників. Інформація про ранговий словник завершується чотирма таблицями: частотний спектр рангового словника лем всього КТ, а також белетристики, публіцистики, епістолярію.

Укладачі наголошують на необхідності зважати на той факт, що окремі слова, цілі класи слів мають різну розрізнявальну силу. Наприклад, у таблиці 7 Словника зафіксовано всього 10 іменників, спільних для всіх трьох списків (*человек, дело, время, раз, день, сердце, жизнь, случай, мысль, год*). П'ять слів (*минута, люди, лета, вопрос, душа*) об'єднують списки художньої літератури й критики; сім слів представлені в першому і третьому списках (*письмо, Бог, ночь, брат, вид, друг, час*). Лише два слова (*журнал и статья*) спільні у другому і третьому списках. Окрім того, таблиця 8 Словника містить 17 дієслів, спільних для всіх трьох списків (*быть, знать, говорить, мочь, сказать, хотеть, стать, видеть, думать, любить, пойти, взять, прийти, сделать, иметь, жить, дать*). Звідси випливає, що, порівняно з іменниками, дієслова мають меншу диференціальну здатність для жанрів.

У наступних трьох таблицях словника показано лексичні маркери основних жанрів (див. рис. 4).

Художественная литература		Критика и публицистика		Письма	
S		S		S	
44	он	74	всё	72	письмо
28	она	62	народ	70	писать
25	-с	50	мы	59	Достоевский
24	князь	50	наш	58	мой
22	вдруг	50	русский	53	роман
21	был	49	Европа	52	написать
20	я	37	народный	48	рубль
17	Алеша	36	Франция	47	Ф.
16	Лиза	35	Россия	46	получить
16	лицо	34	война	45	я
16	рука	31	европейский	42	ты
15	глаз	30	лишь	41	Аня
15	так	29	русский (сущ.)	38	Петербург
14	дверь	29	славянин	38	твой
14	комната	23	идея	37	будет
14	проговорить	23	литература	37	Паша
13	вскричать	23	новый	35	выслать
13	да (утверждение)	23	общество	34	прислать
13	Митя	23	политический	33	ваш
13	смотреть	23	статья	33	лист
13	что-то	23	турок	33	многоуважаемый
12	будто	22	восточный	32	Катков
12	весь	22	господин	31	год

Рис. 4. Найважливіші лексичні маркери основних жанрів (за: Словарь языка Достоевского. Москва, 2001)

Наступні таблиці словника містять квантитативну інформацію про граматичні класи лем та афікси. Як свідчать результати аналізу, художнім творам властива більша кількість дієслів, деад'єктивних прислівників, займенників 3-ї особи, вигуків, слів *так* і *ні* тощо. Для публіцистики, натомість, характерні власні назви, прикметники, суперлятиви, сполучники тощо. Листи здебільшого звернені у майбутнє, тому максимально посилюють умовність та невизначеність (*авось, будущий, в случае, дай Бог, ежели, кабы, мочь, на всякий случай, наверно*), публіцистика опирається здебільшого на форми теперішнього часу (серед її лексичних маркерів слово *есть*), а художня проза обирає минулий час (лексичний маркер *был*).

Завершують дослідження три великі таблиці, де для основних жанрів показано лексичні дані у зв'язку з окремими періодами життя й творчості Ф. Достоевського (1844, 1856, 1866 рр.), які дають цінний матеріал для вивчення еволюції його лексики, що досі фактично ще не була об'єктом лінгвістичного аналізу. За структурою ці таблиці такі ж, як і перші, іншою є лише змістова інтерпретація колонок. Так, виявлено, що частота вживання слів різна залежно від періоду творчості, наприклад, слово *который* трапляється з часом рідше, а використання багатьох сполучників (*а, но, однако, или, али, хоть, хотя, все-таки, если, чтобы*) зростає. Цікаво також відзначити суперництво синонімів у різний час: *отвечать* (93%–91%–47%) і *ответить* (7%–9%–53%), *нужно* (85%–15%–15%) і *надо* (15%–85%–85%), *отчего* (63%–32%–11%) і *почему* (37%–68%–89%) і т. д. Може змінюватися також повнота опрацювання певних фрагментів тексту: з часом автор все рідше вживає слова загальної позитивної оцінки (*добрый, хорошии, хорошио* 204–148–100); із початком другого періоду різко зменшується частота слів, що описують портрет персонажа: *глаза, голова, губы, рот, зубы, щеки, грудь, темный, черный* (загальна частота 533–299–297).

Інші таблиці, скажімо, “Частота деяких бінарних сполучень слова *ДО*” або “Текстуальні зв'язки слів”, дають багатий матеріал для вивчення комбінаторики слів у тексті. У вступі до Словника дано аналіз текстуальних зв'язків слова *русский* як прикметника та іменника і на цьому матеріалі зроблено висновок, що вони ведуть до “заповітних ідей письменника”, а текстуальні зв'язки слова *глаз* розкривають письменницьку техніку.

Порівнявши дані словника Достоевського з двома іншими частотними словниками російської мови (1977 і 1993 рр.), укладачі дійшли цікавого висновку, що тексти письменника містять менше іменників, прикметників і прийменників. Ще перспективніші дані про позитивні та негативні лексичні маркери Ф. Достоевського отримано внаслідок порівняння Словника з КТ його сучасників (Гоголя, Герцена, Тургенєва та ін). На цьому матеріалі здійснено низку статистичних досліджень лексики письменника³⁸.

Отже, хоча проекти лінгвостатистичного опису ідіолектів К. Чапека та Ф. Достоевського виникли і розвивалися незалежно один від одного, проте вони мають багато спільного: з метою різностороннього опису мови письменника укладено частотні словники творів з подібною макро- та мікроструктурою; для цього було використано КТ, що, у свою чергу, складаються з підкорпусів основних жанрів, в яких писав автор; КТ мають внутрішнє морфологічне маркування, що уможливило автоматичне отримання інформації про функціонування частин мови та граматичних категорій; в обидвох проєктах здійснено порівняння частотності лексем письменника з його сучасниками, а також із їх частотністю в загальномовному КТ, що дає підстави об'єктивно виявляти особливості лексики автора та його ключові слова.

³⁸ Див., зокрема, *Шайкевич А. Я.* Дифференциальные частотные словари и изучение языка Достоевского (на примере романа “Идиот”) // Слово Достоевского. Москва, 1996. С.195–253; *Шайкевич А. Я.* О статистическом словаре языка Достоевского // Русский язык в научном освещении. Москва, 2001. № 2. С.122–149.

Виявлені методики дослідження ідіолектів можуть бути використані й в україністиці. Комплексне застосування статистичних та корпусних методів для аналізу текстів українських письменників видається перспективним і цікавим. Подібним за задумом є проект квантитативної параметризації текстів Івана Франка³⁹, в основі якого лежить електронний КТ⁴⁰ із зовнішнім та внутрішнім маркуванням, на першому етапі якого здійснено статистичний опис великої прози письменника⁴¹.

Дослідження здійснено завдяки підтримці Kasy Mianowskiego.

MODERN TECHNIQUES FOR STUDYING WRITER'S LANGUAGE IN SLAVIC LINGUISTICS

Solomija BUK

*Ivan Franko National University of Lviv
1, Universytetska St., Lviv, 79000, e-mail: solomija@gmail.com
Department for General Linguistics*

In the article, modern techniques for studying writer's language in the Slavic linguistics are discussed, namely corpus and statistical. They have become popular with the development of computer technologies. The *Dictionary of Karel Capek* and *Statistical Dictionary of Fyodor Dostoyevsky* are analyzed. The obtained conclusions are extrapolated onto the project of linguostatistical parametrization of Ivan Franko works.

Key words: study techniques, writer's language, idiolect, frequency dictionary, text corpus.

³⁹ Бук С. Квантитативна параметризація текстів Івана Франка: спроба проекту // Іван Франко: Студії та матеріали: Збірник наукових статей. Львів, 2011 (у друці).

⁴⁰ Бук С. Корпус текстів Івана Франка: спроба визначення основних параметрів // Прикладна лінгвістика та лінгвістичні технології: MegaLing-2006: Зб. наук. праць. Київ, 2007. С.72–82.

⁴¹ Частотний словник повісті Івана Франка "Воа constrictor" (редакція 1884 р.) // Стежками Франкового тексту: комунікативні, лінгвсеміотичні, когнітивні, лінгвостатистичні та лексичні виміри прози. Львів, 2012 (у друці); Бук С. Статистична структура роману І. Франка "Борислав сміється" // Учен. зап. Таврич. нац. ун-та ім. В. И. Вернадського. Серія "Филологія". 2010. С.114–118; Бук С. Повість Івана Франка "Захар Беркут" у світлі статистичної лінгвістики // Науковий часопис Національного педагогічного університету імені М. П. Драгоманова. Серія 10. Проблеми граматики і лексикології української мови: Збірник наукових праць. Київ, 2010. Вип. 6. С. 172–176; Бук С. Лінгвостатистичний опис "Не спитавши броду" Івана Франка // Вісник ЛНУ. Серія "Франкознавство". Львів, 2012 (у друці); Бук С. Роман Івана Франка "Для домашнього огнища" крізь призму частотного словника // Мовознавство. 2011. № 4. С.56–66; Бук С. Статистичні характеристики роману Івана Франка "Основи суспільності" (на основі частотного словника твору) // Вісник: Проблеми української термінології. Львів, 2010. № 676. С.90–93; Бук С., Ровенчак А. Частотний словник повісті І. Франка "Перехресні стежки" // Стежками Франкового тексту (комунікативні, стилістичні та лексичні виміри роману "Перехресні стежки"). Львів, 2007. С. 138–369; Бук С. Інтерпретація лексики роману Івана Франка "Великий шум" у кількісно-статистичному аспекті // Studia Methodologica. Вип. 30. Тернопіль, 2010. С.145–153.

**СОВРЕМЕННЫЕ МЕТОДЫ ИССЛЕДОВАНИЯ
ЯЗЫКА ПИСАТЕЛЯ В СЛАВЯНОВЕДЕНИИ**

Соломія БУК

*Львовский национальный университет имени Ивана Франко
ул. Университетская, 1, Львов, 79000, e-mail: solomija@gmail.com.
Кафедра общего языкознания*

В статье анализируются современные методы исследования языка писателя в славяноведении, получившие особенное распространение с развитием компьютерных технологий, а именно корпусный и статистический. Проанализированы “Словарь Карела Чапека” и “Статистический словарь Достоевского”. Полученные выводы экстраполированы на проект лингвостатистической параметризации творчества И. Франко.

Ключевые слова: методы исследования, язык писателя, идиолект, частотный словарь, корпус текстов.

Стаття надійшла до редколегії 09.12.2011.
Прийнята до друку 23.12.2011.