

УДК 81'324.3676 Твори Франка

DOI <https://doi.org/10.32447/2663-340X-2022-11.5>

ВІЗУАЛІЗАЦІЯ АКТИВНОСТІ ОСНОВНИХ МОРФОЛОГІЧНИХ ХАРАКТЕРИСТИК КОРПУСУ ТЕКСТІВ ІВАНА ФРАНКА

Бук Соломія Несторівна

доктор філологічних наук,

доцент кафедри загального мовознавства

Львівського національного університету імені Івана Франка

вул. Університетська, 1, Львів, Україна

orcid.org/0000-0001-8026-3289

www.scopus.com/authid/detail.uri?authorId=26631678100

Бабелюк Оксана Андріївна

доктор філологічних наук,

професор кафедри іноземних мов та перекладознавства

Львівського державного університету безпеки життєдіяльності

вул. Клепарівська, 35, Львів, Україна

orcid.org/0000-0003-4837-1225

Наукову розвідку присвячено візуалізації активності частин мови у творах І. Франка. Квантитативне співвідношення морфологічних категорій у тексті – важлива характеристика як функційних стилів, так й ідіолектів окремих письменників. Морфологічне маркування електронного корпусу текстів великої прози І. Франка, який містить понад 500 000 слововживань, уможливило детальний кількісний аналіз частин мови у двох площинах: у словнику (реєстрі) і тексті автора. Основою корпусу стали усі українськомовні тексти великої прози І. Франка: «Борислав сміється» (1880–1881), «Захар Беркут» (1883), «Не спитавши броду» (1885–1886), «Для домашнього огнища» (1892), «Основи суспільності» (1894–1895), «Перехресні стежки» (1900), «*Voas constrictor*» (1905–1907), «Великий шум» (1907), «Петрії й Довбушуки» (1909–1912). Кожній частині мови (іменнику, прикметнику, займеннику, числівнику, дієслову, прислівнику і службовим) присвячено окрему діаграму, що показує її кількісне співвідношення у кожному із зазначених творів. Графічне зображення особливостей поведінки цих класів слів унаочнило, що найбільшу різноманітність у реєстрі показують іменники, прикметники та дієслова. У тексті ж іменники демонструють поведінку, протилежну до дієслів, прикметників і прислівників – хронологічно їх кількість зростає: чим пізніший твір, тим більше у ньому іменників. Ані у реєстрі, ані у тексті не простежено кореляції між обсягом твору та особливістю функціонування тієї чи іншої частини мови. Результати, зокрема графічні, проливають нове світло на особливості ідіолекту І. Франка, його творчої манери письма, добору морфологічних категорій для якнайточнішого висловлення думки. Вони мають потенціал використання франкознавстві (нпр., Енциклопедія І. Франка), у курсах морфології, стилістики, статистичної, комп'ютерної і прикладної лінгвістики, методики викладання української мови у середній та вищій школі тощо.

Ключові слова: частини мови, частотність, корпус текстів, лінгвостатистика, прикладна лінгвістика, українська мова, І. Франко.

Постановка проблеми в загальному вигляді та обґрунтування її актуальності. Творчість І. Франка – глибоко самобутнє явище, яке до сьогодні викликає різноаспектний інтерес у дослідників. Морфологічна параметризація його текстів – важлива складова моделювання його ідіостилу (Бук, 2021). Трендовим підходом до цього завдання у сучасній лінгвістиці є укладення електронного корпусу текстів письменника, а необхідною вимогою цього лінгвістичного конструкту є морфологічне мар-

кування. Саме завдяки йому стало можливим автоматичне виявлення і статистичний аналіз основних морфологічних категорій в електронному корпусі текстів великої прози І. Франка (КТ ВПФ). Продемонструємо це на прикладі частин мови.

Квантитативне співвідношення цих морфологічних класів отримало свій опис в англійському (Barrett et al., 2007; Leech et al., 2001), польському (Kamińska-Szmaj, 1989; Zgólkowa, 1987), японському (Masamitsu, 2005) мовознав-

стві тощо. Порівнянню щільності частиномовних тегів (PoS tag densities) у семи англійських документах чотирьох різних сфер (медичній, фінансовій, військовій та наративній) присвячено частину праці (Barrett et al., 2007). Якісне та відсоткове порівняння різних частин мови у дитячій літературі у зіставленні з усною літературою (фольклорна пісня, діалектна розповідь), писемністю (преса, сучасна поезія), усною мовою (школа, діти) досліджує у своїй роботі С. Кшиська: автор довів, що іменники вимагають у тексті інших частин мови, а саме прикметників (Krzyńska, 2008).

Багато праць про ідіолект письменника отримали детальну інформацію про частини мови, оскільки вони відіграють велику роль у стилістиці та визначенні авторського стилю: К. Чапека (Čermák, 2007), М. Павича (Vasić 1998), Ю. Пшибося (Zembaty-Michalakowa, 1982), Ц. Норвіда (Puzynina, Korpysz, 2008–2012) і частково І. Франка (Бук, 2011). У пропонованій статті останнє дослідження доповнено як фактичним, так й інтерпретаційним матеріалом. Активність частин мови у словнику і тексті ВПФ, а також їхня графічна візуалізація стали окремим предметом аналізу такому ракурсі вперше. Ще сам І. Франко ретельно ставився до добору в тексті тієї чи іншої частини мови, наголошуючи на їх квантитативних характеристиках. У статті про переклад «Каменярів» відображено його міркування з цього приводу: «Коли слова перших трьох категорій (іменники, дієслова, прикметники творять головну основу словесного твору, надаючи йому зміст і акцію, то слова другої категорії (займенники, прислівники, злучники) – се наче тінювання в малярстві, що оживляє і упластичнює картину» [Франко 1976–1986. Т. 39: 20].

У багатьох галузях гуманітаристики (філології, педагогіці, соціології, економіці) та й у науці загалом можна спостерегти чітку тенденцію до гейміфікації та візуалізації, причому і метафоричної, і графічної (див., наприклад (Vabelyuk et al., 2001)). Їх мета – полегшити сприйняття складного (зокрема, цифрового) матеріалу. Саме тому до аналізу матеріалу у цій статті застосовано графічне розрізнення, візуалізоване у діаграмах.

Мета наукової розвідки – виявити й візуалізувати активність функціонування основних морфологічних характеристик у дев'яти українськомовних творах великої прози І. Франка: «Борислав сміється» (1880–1881), «Захар Беркут» (1883), «Не спитавши броду» (1885–1886), «Для домашнього огнища» (1892), «Основи суспільності» (1894–1895), «Перехресні стежки»

(1900), «Voа constrictor» (1905–1907), «Великий шум» (1907), «Петрії й Довбушуки» (1909–1912). **Завдання** – обчислити частотність частин мови у кожному з цих текстів; зобразити їх кількісні характеристики візуально; порівняти їх між собою та між творами.

Основна частина. Обсяги творів ВПФ різні (від найкоротшого «Voа constrictor» (34 215) до найдовшого «Перехресні стежки» (93 888)), тому доцільно оперувати тільки відносними частотами їх одиниць. Притримуємося класичного підходу до поділу слів на частини мови, до службових зараховуємо прийменник, сполучник, частку та, умовно, вигук.

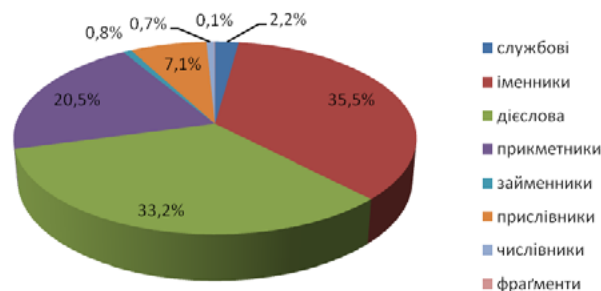


Рис. 1. Діаграма частин мови у словнику ВПФ

Діаграма на рис. 1 демонструє, що найбільшу частину словника ВПФ займають іменники (35,5%), дієслова (33,2%) та прикметники (20,5%), таким чином залишаючи для інших морфологічних класів лише 10%: прислівник (7%), службові (2,2), числівник (0,7%). При цьому, найбільший розхил між творами демонструють іменники (6,1%): від «Борислава ...» (30,8%) до «Великого шуму» (36,9%), трохи менший розхил – дієслова (4,6%): знову ж від «Борислава ...» (36,9%) до «Великого шуму» (32,3%) та прикметники (3,2%): від «Захара ...» (20,8%) до «Великого шуму» (17,6%). Очікувано стабільну приблизно однакову частку словника мають закриті частини мови: займенники, числівники, службові.

Зіставляючи кількісні характеристики частин мови у словнику (реєстрі) й тексті ВПФ, доходимо висновку, що результат не повністю збігається з твердженням [Перебийніс та ін. 1985: 186]: «хоч процентне співвідношення їх [частин мови] у тексті і реєстрі є різне: кількість істотних розходжень між процентними показниками частин мови у тексті більша, ніж у реєстрі, особливо у таких повнозначних частин мови, як дієслово, іменник, прислівник». Так, у ВПФ дієслова, навпаки, у тексті мають меншу різницю (2%), а у словнику – більшу (4%). Отже, у тексті ВПФ дієслова функціонують однорідніше, ніж в українській художній прозі загалом.

Співвідношення частин мови у текстах творів неоднакове, тому далі зіставлено частотність кожного морфологічного класу за конкретним твором. **Службові частини** кількісно стоять на першому місці у тексті усіх дев'яти обстежених романів І. Франка: «Борислав ...» (тут вони займають найбільшу частку тексту – 27,4%), «Великий шум» (26,9%), «Не спитавши ...» (26,6%), «Основи ...» (25,9%), «Перехресні ...» (25,7%), «Для домашнього ...» (24,5%), див. рис. 2. Найменшу частку вони займають у творах «Захар ...» (24,2%) і «Петрії ...» (24,5%) – тут найчастотнішою частиною мови є іменник, що займає 26,3% та 24,5% відповідно.

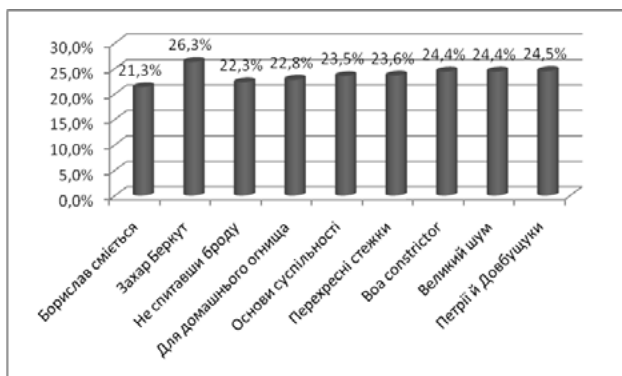


Рис. 2. Діаграма службових частин мови у тексті творів ВПФ

Іменники у всіх творах, окрім двох згаданих останніх, стоять на другому місці: «Воа constrictor» та «Великий шум» (24,4%), «Перехресні ...» (23,6%), «Основи ...» (23,5%), найменшу відносну кількість покриваючи у «Для домашнього ...» (22,8%), «Не спитавши ...» (22,29%), «Борислав ...» (21,3%), рис. 3.

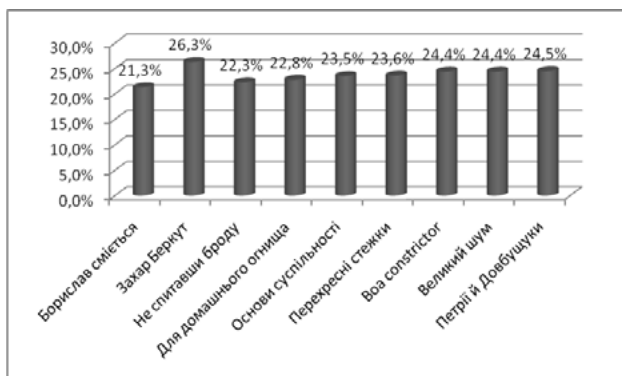


Рис. 3. Діаграма іменників у тексті творів ВПФ

Далі ранг частин мови у всіх романах збігається: дієслово, займенник, прикметник, числівник. Отже, на третьому місці стоять **дієслова**, яких майже однаково у «Для домашнього ...» (18,8%), «Перехресних ...» (18,6%), «Борис-

лав ...» (18,4%), «Не спитавши ...» (18,32%), «Основи ...» (18,1%), «Петрії ...» (18%), «Воа constrictor» (17,8%) і найменше у «Великому шумі» (17,4%) та «Захарі ...» (16,8%), рис. 4.

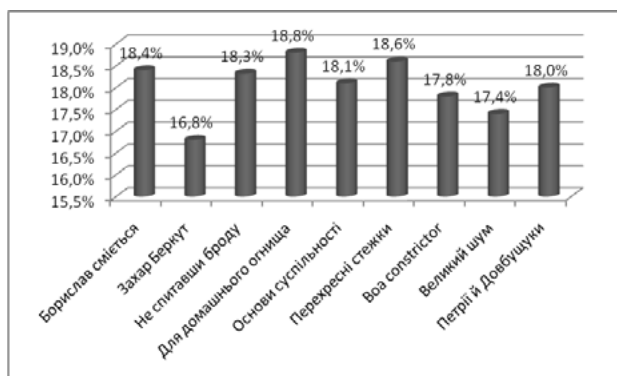


Рис. 4. Діаграма дієслів у тексті творів ВПФ

На четвертому місці – **займенники**, яких також найбільше у «Для домашнього ...» (16,5%), «Петрії ...» (16,1%), «Перехресні ...» (16%) і також найменше у «Захарі ...» (14,3%), рис. 5.

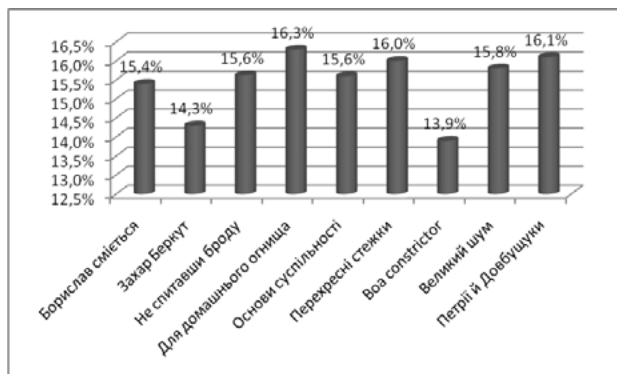


Рис. 5. Діаграма займенників у тексті творів ВПФ

На п'ятому місці – **прислівники**, яких також найбільше у «Петріїх ...» (9,6%), «Бориславі ...» (9,1%), «Основах ...» (8,9%), «Для домашнього ...» (8,8%) а найменше у «Захарі ...» (7,6%) і «Великому шумі» (7,0%), рис. 6.

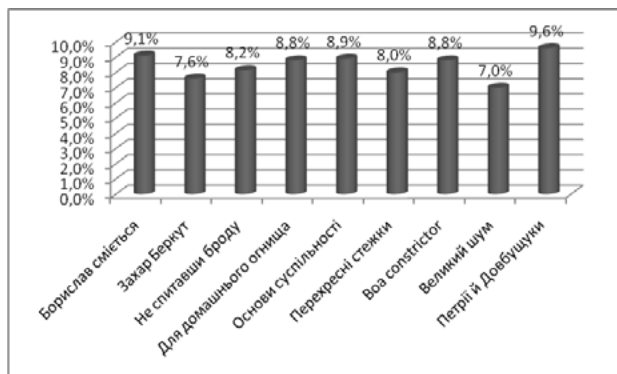


Рис. 6. Діаграма прислівників у тексті творів ВПФ

На другому з кінця місці за частотністю у тексті перебувають **прикметники**, що, домінують у «Захарі ...» (9,8%), «Петріях ...» (8,9%) і «Воа constrictor» (8,2%), а найменшу кількість мають в «Перехресних ...» (6,9%) та «Бориславі ...» (7,0%), «Основах ...» (7,2%), «Великому шумі» (7,5%), рис. 7.

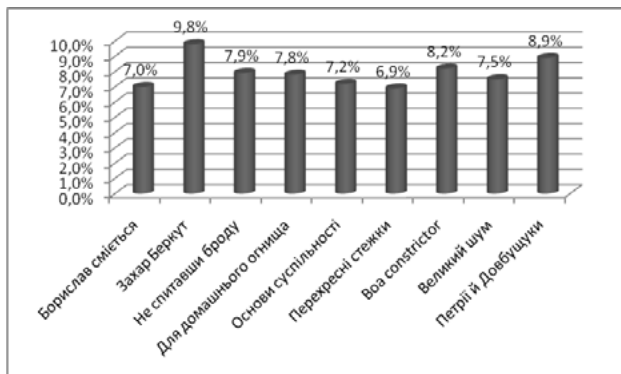


Рис. 7. Діаграма прикметників у тексті творів ВПФ

На останній позиції у тексті – **числівники**, вони переважають у «Воа constrictor» (1,4%), «Бориславі ...» і «Петріях ...» (1,3%), «Перехресних ...» (1,2%), що зумовлено тематикою творів: у «Воа constrictor», «Бориславі ...» важливу роль відіграє обчислення коштів, необхідних для повстання, і сума зібраних грошей; у «Петріях ...» та «Перехресних ...». Відсотково найменше числівників у тексті «Осн...» (0,8%), «Захара ...», «Для домашнього ...», «Великого шуму» (1%), рис. 8.

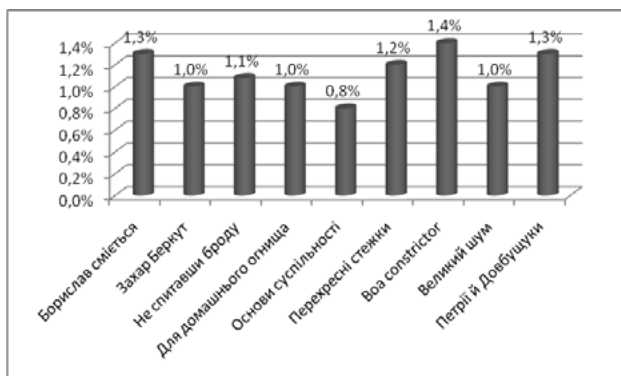


Рис. 8. Діаграма числівників у тексті творів ВПФ

Із візуалізованого аналізу частотності частин мови у вигляді діаграм видно, що у текстах ВПФ найбільш рівномірно (із розхилом 2%) функціонують дієслова: від 16,8% («Захар ...») до 18,8% («Для домашнього ...»), а також прислівники: від 7,6% («Захар ...») до 9,6% («Петрії ...») – рис. 4 і 6.

Цікаво, що серед закритих морфологічних класів великий розхил у тексті мають

як службові частини мови (5,8%): від 21,6% («Петрії ...») до 27,4% («Борислав ...»), так і займенники (2,4%): від 13,9% («Воа constrictor») до 16,3% («Для домашнього ...»). Числівники очікувано демонструють стабільність з розхилом 0,6%: від 0,8% («Основи ...») до 1,4% («Воа constrictor»). Цікавим фактом є те, що у зіставленні зі сучасною українською художньою прозою найбільшу різницю (4,5%, тобто фактично третину!) мають займенники (15,5% : 11,0%). Різниця ж усіх інших морфологічних класів коливається близько 1%.

У словнику (**реєстрі**) відповідних творів картина інша. Словник **іменників** відносно найбагатший у «Великому шумі» (36,90%) та «Перехресних ...» (36,6%), а відносно найбідніший у «Бориславі ...» (30,79%) і «Захарі ...» (31%), див. рис. 9. Можна зауважити, що кількість різних іменників зростає за хронологією творів: раніші мають менше цієї частини мови, а пізніші – більше.

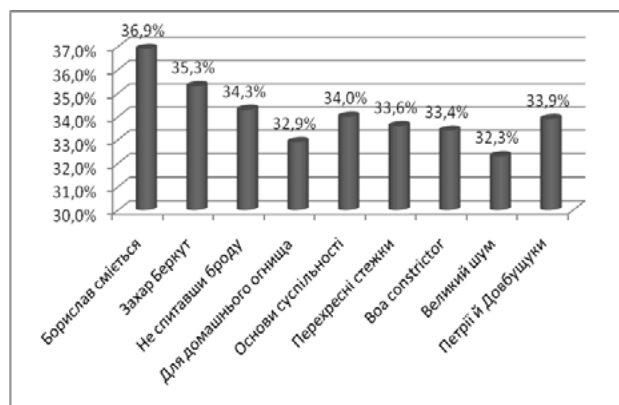


Рис. 9. Діаграма іменників у реєстрі творів ВПФ

Реєстр **дієслів**, навпаки, найбільший у «Бориславі ...» (36,91%), а найменший у «Великому шумі» (32,34%) та «Для домашнього ...» (32,9%), демонструючи протилежну тенденцію, обернену до хронології написання творів, рис. 10.

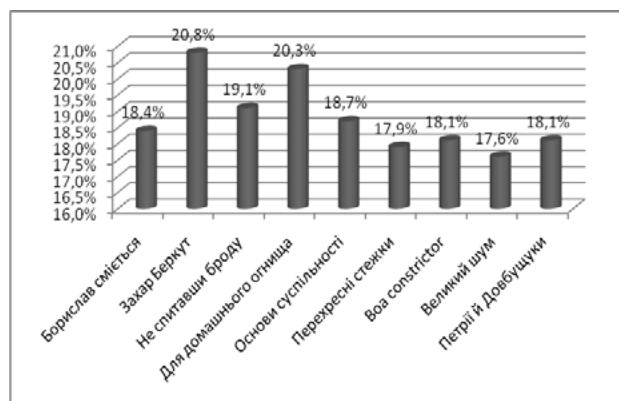


Рис. 10. Діаграма дієслів у реєстрі творів ВПФ

Різних **прикметників** найбільше у «Захарі ...» (20,77%) та «Для домашнього ...» (20,3%), а найменше – в пізніших романах «Перехресних ...» (17,95%) і «Великий шум» (17,56%), рис. 11.

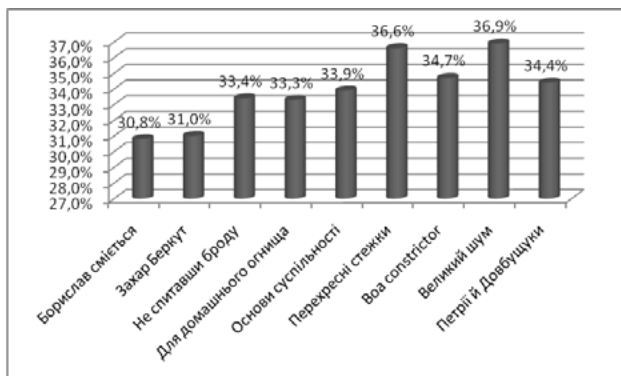


Рис. 11. Діаграма прикметників у реєстрі творів ВПФ

Прислівники мають подібну тенденцію: кількісно переважають у раніших романах «Борислав ...» (8,84%), «Захар ...» (8,80%), «Для домашнього ...» (8,7%), «Не спитавши ...» (8,6%) і зменшуються у словнику пізніших творів «Перехресні ...» (7,39%), «Великий шум» (7,64%), рис. 12.

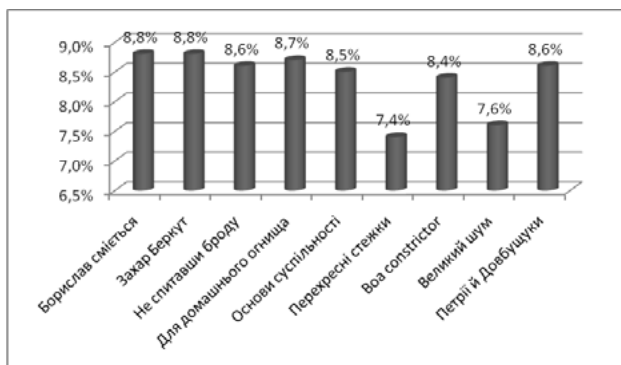


Рис. 12. Діаграма прислівників у реєстрі творів ВПФ

Частка **займенників та числівників** у словниках романів відносно стала (рис. 13–14).

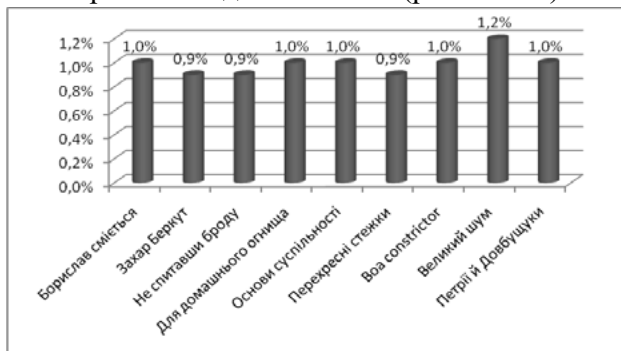


Рис. 13. Діаграма займенників у реєстрі творів ВПФ

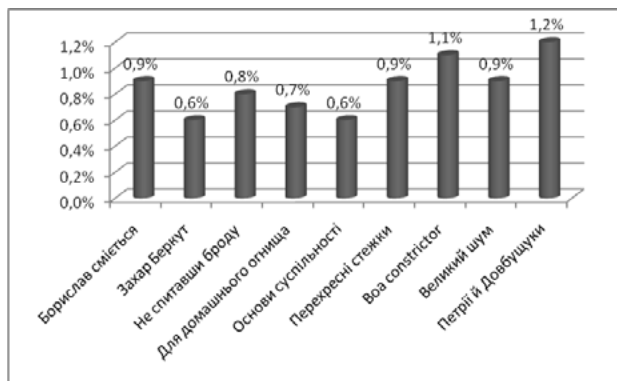


Рис. 14. Діаграма числівників у реєстрі творів ВПФ

Службові частини мови найбільшу частку реєстру займають у «Великому шумі» (3,45%) та «Воа constrictor» (3,3%), а найменшу – у «Перехресних ...» (2,64%), «Захарі ...» (2,7%) та «Петріях ...» (2,8%), рис. 15.

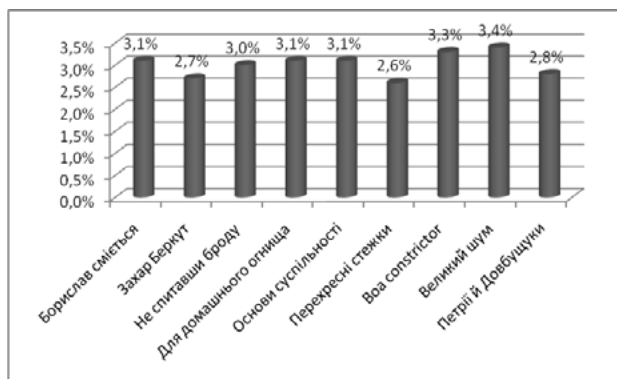


Рис. 15. Діаграма службових частин мови у реєстрі творів ВПФ

Кореляції між обсягом твору та кількістю вживання тієї чи іншої частини мови не простежено ані у реєстрі, ані у тексті. Припущення, на перший погляд, логічне, про ріст використання займенників у тексті зі збільшенням обсягу тексту не знайшло підтвердження. Можна констатувати відносну сталість співвідношення кількості морфологічних класів у словнику, крім описаної тенденції зміни відносної кількості прикметників та прислівників, залежно від часу написання твору.

Підсумки та перспективи дослідження. Підсумовуючи викладене, можна стверджувати, що кожен твір ВПФ має свій індивідуальний лінгвостатистичний портрет морфологічних категорій. На основі даних про розподіл частин мови у ВПФ обсягом понад 500 000 слововживань, отриманих завдяки морфологічній анотації корпусу його текстів, виявлено особливості поведінки цих класів слів у словнику і тексті творів письменника, зокрема, те, що най-

більшу різноманітність у реєстрі демонструють іменники, прикметники та дієслова.

Іменники у тексті демонструють поведінку, протилежну до дієслів, прикметників і прислівників: хронологічно їх кількість зростає: чим пізніший твір, тим більше у ньому іменників (див. рис. 9).

Ані у реєстрі, ані у тексті не простежено кореляції між обсягом твору та особливістю функціонування тієї чи іншої частини мови.

Результати, зокрема графічні, проливають нове світло на особливості ідіолекту І. Франка, його творчої манери письма, добору морфологічних категорій для якнайточнішого висловлення думки. Вони мають потенціал використання франкознавстві (нпр., Енциклопедія І. Франка), у курсах морфології, стилістики, а також статистичної, корпусної, комп'ютерної і прикладної лінгвістики, методики викладання української мови у середній та вищій школі тощо.

ЛІТЕРАТУРА

1. Бук С. Велика проза Івана Франка: електронний корпус, частотні словники та інші міждисциплінарні контексти : монографія. Львів : ЛНУ імені Івана Франка, 2021. 424 с.
2. Бук С. Частини мови у словнику і тексті Івана Франка (на матеріалі великої прози). *Лінгвістичні студії*. 2011. Вип. 22. С. 62–66.
3. Перебийніс В. С., Муравицька М. П., Дарчук Н. П. Частотні словники та їх використання. Київ : Наукова думка, 1985. 204 с.
4. Франко І. Зібрання творів: у 50 т. / АН УРСР, Ін-т л-ри ім. Т. Г. Шевченка. Київ : Наукова думка, 1976–1986. Т. 1–50.
5. Babelyuk O.A., Koliasa O.V., Matsevko-Bekerska L., Matuzkova O., Pavlenko N. The Interaction of Possible Worlds through the Prism of Cognitive Narratology. *Arab World English Journal (AWEJ) Volume 12. Number 2 June 2021*, URL: <https://ssrn.com/abstract=3892970> or <http://dx.doi.org/10.2139/ssrn.3892970>
6. Barrett L., Greenberg B., Schwartz M. A Syntactic Feature Counting Methods of Selecting Machine Translation Training Corpora. *Language and Computers: Studies in Practical Linguistics 2007*. 60(1). P. 1–19.
7. Kamińska-Szmaj I. Słownictwo tekstów popularnonaukowych w ujęciu statystycznym. *Rozprawy Komisji Językowej. Wrocławskie Towarzystwo Naukowe*. 1989. XVI. S. 69–87.
8. Krzyśka, Sławomir Statystyka języka wierszy dla dzieci // Katarzyna Krasoń, Beata Mazepa-Domagala (red.), *Wyrazić i odnaleźć siebie, czyli o sztuce, ekspresji, edukacji i arteterapii*, Wydawnictwo Górnośląskiej Wyższej Szkoły Pedagogicznej im. Kardynała Augusta Hlonda. Mysłówice, 2008. S. 271–279. URL: http://www.staff.amu.edu.pl/~slawek/statystyka_wierszy.pdf
9. Leech G., Rayson P., Wilson A. *Word Frequency in Written and Spoken English. Based on the British National Corpus*. Pearson Educational Ltd. 2001. 304 p.
10. Masamitsu I. Quantitative linguistics in Japan. // Köhler R., Altmann G., Piotrowski R. (eds.), *Quantitative Linguistik – Quantitative Linguistics. Ein internationales Handbuch – An International Handbook*. Berlin–New York : de Gruyter, 2005. S. 82–95.
11. Puzynina J., Korpysz T. Internetowy słownik języka Cypriana Norwida, Pracownia Słownika Języka Norwida, Wydział Polonistyki, Uniwersytet Warszawski. Warszawa, 2008–2012. URL: <https://sownikjezykanorwida.uw.edu.pl/titulowa>.
12. Slovník Karla Čapka / Hl. ed. František Čermák. Praha : Nakladatelství í Lidové noviny ; Ústav Českého národního korpusu, 2007. 715 s. + CD
13. Zembaty-Michalakowa M. Poezja Juliana Przybosa w świetle badań statystyczno-językowych na tle porównawczym. Wrocław, 1982. 312 s.
14. Zgólkowa H. Ilościowa charakterystyka słownictwa współczesnej polszczyzny. *Wybrane zagadnienia*. Poznań : UAM, 1987. 145 s.
15. Vasić S. Polazne osnove novije srpske proze. Knjiga II. Hazarski rečnik Milorada Pavića: frekvencijski rečnik. Beograd : Kultura, 1998. 840 s.

REFERENCES

1. Buk S. (2021) Velyka proza Ivana Franka: elektronnyi korpus, chastotni slovnyky ta inshi mizhdystsyplinarni konteksty [Long prose fiction by Ivan Franko: electronic corpus, frequency dictionaries, and other interdisciplinary contexts]: monograph. Lviv: Ivan Franko Lviv National University [in Ukrainian].
2. Buk S. (2011) Parts of speech in the vocabulary and text by Ivan Franko (based on long prose fiction) [Chastyny movy u slovnyku i teksti Ivana Franka (na materialy velykoi prozy)] // *Linguistic studies [Linhvistychni studii]*. Vyp. 22. P. 62–66 [in Ukrainian].

3. Perebyinis V. S., Muravytska M. P., Darchuk N. P. (1985) Frequency dictionaries and their use [Chastotni slovnyky ta yikh vykorystannia]. Kyiv : Naukova dumka, 1985. 204 p. [in Ukrainian].

4. Franko I. (1976–1986) Collection of works [Zibrannia tvoriv]: in 50 vol. / AN URSSR, In-t l-ry im. T. H. Shevchenka. Kyiv : Naukova dumka. Vol. 1–50. [in Ukrainian].

AUTHOR'S POSTMODERN IRONY IN ENGLISH LITERARY TEXT

Buk Solomija Nestorivna

dr. hab.,

associate professor of the Department of General Linguistics

Ivan Franko National University of Lviv

1, Universytetska St., Lviv, Ukraine

orcid.org/0000-0001-8026-3289

www.scopus.com/authid/detail.uri?authorId=26631678100

Babelyuk Oksana Andriivna

dr. hab., professor,

professor of the Department of Foreign Languages and Translation Studies

Lviv State University of Life Safety

Kleparivska Street, 35, Lviv, Ukraine

orcid.org/0000-0003-4837-1225

The article is focused on the visualization of the activity of parts of speech in the works of Ivan Franko. The quantitative ratio of morphological categories in the text is an important characteristic of both functional styles and idiolects of individual writers. The morphological markup of the electronic corpus of long prose fiction by I. Franko, which contains more than 500,000 tokens, allowed a detailed quantitative analysis of parts of speech in two planes: in the vocabulary and text. The corpus covers all Ukrainian novels by I. Franko: Boryslav Laughs [Boryslav smiïetsia] (1880–1881); Zakhar Berkut (1883); Without Asking a Wade [Ne spytavshy brodu] (1885–1886); For the Hearth [Dlia domashnoho ohnyshcha] (1892); Pillars of Society [Osnovy suspilnosti] (1894–1895); The Cross-paths [Perekhresni stezhky] (1900); Boa constrictor, (1905–1907); The Great Noise [Velykyi shum] (1907); Petriji j Dovbuščuky (1909–1912).

A separate chart is dedicated to each part of the speech (noun, adjective, pronoun, number, verb, adverb, and auxiliary words) showing its quantitative ratio in every novel.

Graphic representation of the behavior of these classes of words showed that in the vocabulary, nouns, adjectives, and verbs show the greatest diversity. In the text, nouns show the opposite behavior to verbs, adjectives and adverbs - chronologically their number increases: the later the work, the more nouns are used in it. No correlation was found between the size of the work and the peculiarity of the functioning of a particular part of speech neither in the vocabulary nor in the text.

The results, in particular graphic, shed new light on the features of the idiolect of Ivan Franko, his creative style of writing, the selection of morphological categories for the most accurate expression of thought. They have the potential to be used in Franko's studies (for example, I. Franko's Encyclopedia), in courses in morphology, stylistics, as well as in statistical, computational and applied linguistics, methods of teaching Ukrainian as a second language, etc.

Key words: *parts of speech, frequency, text corpus, linguostatistics, applied linguistics, Ukrainian language, Ivan Franko.*