

УДК 811.161.2'33:82'09

DOI: 10.18523/lcmp2522-9281.2019.5.158-169

Ілля Данилюк

<https://orcid.org/0000-0001-8681-9321>

Анатолій Загнітко

<https://orcid.org/0000-0001-7398-6091>

Ганна Ситар

<https://orcid.org/0000-0001-8806-8322>

КОРПУС ТЕКСТІВ ЮРІЯ ШЕВЕЛЬОВА: СТРУКТУРА, ФУНКЦІЇ, НАВІГАЦІЯ

Статтю присвячено проблемі створення корпусу текстів Юрія Шевельова. Визначено основні структурні компоненти корпусу текстів, виділено функції, схарактеризовано особливості його наповнення, пояснено екстралінгвістичне й лінгвістичне розмічення.

Ключові слова: конкорданс, корпус текстів, лінгвоперсона, лінгвоперсонологія, розмітка корпусу, частотний аналіз, Юрій Шевельов.

Вступ

Лінгвоперсонологія ґрунтована на філософському розумінні персоналізму (Б. П. Боун, Р. Г. Лотце, Дж. Х. Хауісон, У. Джеймс, Дж. Ройс, М. У. Калкінс, У. Е. Гокінг, Г. У. Опорт, Е. Муньє, Т. О. Бофорд та ін.), одним із виявів якого постає теза про духовне оновлення особистості, тяглість її самовдосконалення¹. У сучасній лінгвістиці актуалізовані

¹ Дослідження виконано в межах фундаментального наукового проекту «Об'єктивна і суб'єктивна мовносоціумна граматики: комунікативно-когнітивний та прагматико-лінгвокомп'ютерний виміри» (0118U003137).

питання мовного портрета особистості, встановлення структурних і функційних компонентів останнього. Заявлені та низка інших питань постають істотними для сучасної прикладної лінгвістики, почасти для психолінгвістики, лінгвопсихіатрії, нейролінгвістичного програмування, сугестивної лінгвістики та ін. У лінгвоперсоналогії як науці про мовну особистість – одноосібну (монолінгвоперсона) чи колективну (полінгвоперсона) – лінгвоперсона досліджують у функційно-комунікативних, вербально-семантичних, лінгвопсихологічних, мовленнєвоповедінкових, дискурсивно-ситуативних, мотиваційно-сугестивних та інших вимірах. Актуальними є питання формування мовленнєвих потреб персони, визначення їх тяглості, діагностування видозміни мовних і мовленнєвих потреб у життєвому просторі лінгвоперсони з внутрішнім диференціюванням статусу та функційного навантаження мовних і мовленнєвих переходів, кодування й перекодування, формування множин лінгвоіндивідуалізацій, умотивованих територіальними, соціальними, віковими, естетичними, етичними, емоційно-ситуативними та іншими чинниками. Особливо важливим постає дослідження знаних лінгвоперсон, до яких належить Юрій Шевельов – видатний лінгвіст, історик літератури, критик. У його мовленнєво-дискурсивних практиках співвіднесені різні мовні стихії, нерівнорядні дискурсивні практики, активізовані територіальні, соціальні, вікові, естетичні, емотивні та інші мотиватори індивідуальних видозмін. Мотивованим є наукове завдання встановлення лінгвокорпусної структурованості дискурсивних практик Юрія Шевельова. Заявлене корелює з практичним завданням – уведенням в активний обіг не лише науково-творчої спадщини видатного науковця, а й використання напрацьованих Юрієм Шевельовим мовних моделей і конструкцій, лінгвоіндивідуалізацій у сучасній мовно-кодіфікованій практиці.

Теоретичне підґрунтя. Уперше поняття *лінгвоперсони* на початку ХХ ст. використав М. Трубецької із опертям на ідеї персоналізму (Е. Мунье та ін.). У другій половині ХХ ст. у зв'язку з активізацією студіювань мовної особистості, питань активної, асоціативної граматики (Караулов, 1987) поняття *лінгвоперсони* набуло іншого виміру й статусу. Активізацію лінгвоперсоналогічних студіювань спостережувано в кінці ХХ – на початку ХХІ ст., коли істотними постали теоретичні й практичні напрацювання комунікативних стратегій і тактик, дискурсології, лінгвістики тексту, спроби дослідження мовленнєвоповедінкових стереотипів,

розширення студій із прикладної лінгвістики, окреслення теоретичних засад об'єктивної граматики та ін. Не менш значущими були спроби мовного портретування особистості, створення авторських корпусів текстів.

Мовну особистість Юрія Шевельова розглядав Р. Трифонов, з'ясовуючи лінгвокультурні складники індивідуального мовлення науковця на матеріалі есе та листів, а М. Мозер розкриває основні етапи динаміки творчої особистості Ю. Шевельова на дорозі відрадянщини (Трифонов, 2009; Трифонов, 2015; Мозер, 2014). Не менш актуальними є студіювання А. Даниленка, М. Коцюбинської, Л. Масенко, Л. Тарновецької та ін. У мережі інтернет сьогодні розміщено більшість праць Ю. Шевельова, що загалом не вирішує питання цілісного бачення його творчої лінгвоперсона, що можливе за умови цілісної корпусациї його спадщини, напрацювання лінгвокомп'ютерних технологій вільної навігації в межах гіпертексту. Тому метою студіювання є експериментально-дослідницька модель текстової корпусациї лінгвоперсона Юрія Шевельова з визначенням основних компонентів корпусу, закономірностей вільної лінгвотекстової навігації.

Методологічне підґрунтя дослідження

Цілісна корпусация лінгвоперсона Юрія Шевельова, а також розгляд особливостей мовного портрета науковця потребує застосування відповідних методів, з-поміж яких значущими постають метод дискурсивного аналізу з використанням методики корпусного контент-аналізу для встановлення усіх можливих фіксацій досліджуваного слова з фіксацією його граматичних змін, вузьких і розширених контекстів, а також – для вияву невідповідних сполук слів – колокацій, зіставлення останніх із регулярними та квантитативно домінувальними. Використання корпусного контент-аналізу може слугувати опертям створення скетча слова з внутрішньокорпусним диференціюванням узуальних та оказіонально-авторських (індивідуалізованих) сполук слів, визначення їх навантаження в межах комунікативних реєстрів, вияву комунікативних інтенцій.

Індивідуально-авторські експериментальні науково-дослідні корпуси: функційне навантаження в дослідженні лінгвоперсона

Викінчене мовне портретування можливе за умови охоплення усієї сукупності породжених особистістю текстів із діагностуванням терито-

ріальних, соціальних, генераційних та інших мотивацій змін у дискурсивних практиках. Сукупність наголошених питань постає однією із площин сучасної лінгвоперсонології, окремі напрями якої активно студійовані на кафедрі загального та прикладного мовознавства і слов'янської філології Донецького національного університету імені Василя Стуса (А. Загнітко, І. Данилюк, Ж. Краснобаєва-Чорна, Г. Ситар та ін.). Опрацювання рівнів, категорій та аспектів лінгвоперсонології уможливить створення об'єктивної та суб'єктивної мовносоціумної граматики, в центрі якої перебуває моно- і/або полілінгвоперсона в неозначеній кількості її і/або їхніх дискурсивних практик. Діагностування видозмін у 1) структурі мовної особистості, 2) рівнях її реалізації (вербально-семантичного, лінгвокогнітивного, мотиваційного (за Ю. Карауловим), функційного (В. Конецька), ситуативно-психологічного), 3) співвідношенні компонентів (формування уміння комунікувати, потреби у спілкуванні й належної компетентності, формування мовної свідомості та мовної і/чи мовленнєвої поведінки) з опертям на врахування: а) органічної потреби спілкування рідною мовою; б) комунікативної необхідності спілкуватися чужою мовою; в) дискурсивно-мотивованої практики (педагогічна, лікарська та ін.) спілкування нерідною мовою, можливе за умови створення цілісного корпусу текстів мовної особистості.

Індивідуально-авторські експериментальні науково-дослідні корпуси дають змогу встановити рівні й аспекти лінгвоперсона з опертям на її персонотексти. Через призму сумарної величини текстового корпусу можна визначити певні тематично-ключові елементи розкриття тяглості розвитку лінгвоперсона, формування лакун, мотивованих територіальними, соціальними та іншими чинниками, а також відповідним мовним ландшафтом функціонування лінгвоперсона, реконструювання її творчого потенціалу та механізмів його реалізації. Прикметно, що реконструйованими можуть поставати результати навчання, розпросторення й самореалізація лінгвоперсона та ін. Текстовий корпус уможлиблює простеження закономірностей мовносоціальної зумовленості лінгвоперсона та статусу лінгвоіндивідуалізацій як індивідуального самовияву: потяг, бажання, схильність, світогляд, переконання. Лінгвоперсона виявляє себе через стосунки з іншими, де відносини з іншими індивідами – це переживання й усвідомлення, переконання ж окреслюють спрямування особистості,

її мовні цінності. Згодом постають значущими знання, уміння, навички та звички користуватися мовою, диференціювати її в різних корпоративних групах, уміти використовувати різні мовні коди залежно від ситуацій, що підтверджує мовножиттєвий простір Юрія Шевельова в різні періоди його життя – в дитинстві, студентські роки, харківський період, роки мюнхенської еміграції, викладання в Лундському університеті, а згодом – у Гарвардському й Колумбійському університетах.

Послідовне використання корпусного контент-аналізу забезпечує визначення емоційного тла дискурсивних практик, вияву в них відчуттів і почуттів, рівнів пам'яті та ін. У цілісності авторський експериментальний науково-дослідницький корпус текстів забезпечує з'ясування властивостей темпераменту лінгвоперсони, простеження зміни її мовленнєвоповедінкових реалізацій, встановлення типологічних властивостей лінгвоперсони. Загалом дискурс лінгвоперсони – її життєвий простір із різними функційно навантаженими сукупностями самореалізацій у мову та в мові.

Індивідуально-авторський експериментальний науково-дослідний корпус лінгвоперсони Юрія Шевельова: структура, технології

Юрій Шевельов – видатна особистість, яка реалізована в численних текстах, різних дискурсивних практиках, нерівнорядних лінгволандшафтних просторах.

Для дослідження особливостей мовлення Юрія Шевельова було вирішено укласти дослідницький корпус текстів. Цей корпус текстів спроектували й реалізували викладачі кафедри загального та прикладного мовознавства і слов'янської філології Донецького національного університету імені Василя Стуса А. Загнітко, І. Данилюк і Г. Ситар. У межах проходження навчальної практики в підготовці текстів для корпусу активну участь взяли студенти 2 курсу спеціальності «Прикладна лінгвістика».

Корпус текстів Юрія Шевельова створений за допомогою вільного корпусного менеджера *NoSketch Engine* (<https://www.sketchengine.eu/>), розробленого в університеті Масарика (Брно, Чехія) (Rychlý & Smrž, 2004). Цей корпус разом з іншими доступний на сервері кафедри загального та прикладного мовознавства і слов'янської філології ДонНУ імені Василя Стуса за адресою corpora.donnu.edu.ua.

За задумом авторів, корпус містить усі видані на сьогодні тексти Юрія Шевельова, зокрема:

Шевельов Юрій. З історії незакінченої війни (упорядники Оксана Забужко, Лариса Масенко). Київ: Видавничий дім «Києво-Могилянська академія», 2009. 471 с. ISBN 978-966-518-519-2; Шевельов Юрій. З історії незакінченої війни (упорядники Оксана Забужко, Лариса Масенко). Київ: Видавничий дім «Києво-Могилянська академія», 2009. 471 с. ISBN 978-966-518-519-2; Шерех Юрій. Поза книжками і з книжок. Київ: Вид-во «Час», 1998. 456 с. ISBN 966-95238-3-4; Шерех Юрій. Друга черга. «Бібліотека Прологу і Сучасности Ч. 130», 1978. 393 с.; Шерех Юрій. Не для дітей. Нью-Йорк: Вид-во «ПРОЛОГ», 1964. 416 с.; Шевельов Юрій. Вибрані праці: у 2 кн. Книга II. Літературознавство. Київ: Видавничий дім «Києво-Могилянська академія», 2009. 1151 с. ISBN 978-966-518-496-6; Шевельов Юрій. Вибрані праці: у 2 кн. Книга I. Мовознавство. Київ: Видавничий дім «Києво-Могилянська академія», 2009. 583 с. ISBN 978-966-518-494-2; Шерех Юрій. Третя сторожа. Балтимор–Торонто: Смолоскип, 1991. 454 с.; Шевельов Юрій. «Я, мені, мене... (і довкруги)». Спогади. У двох томах. Видавництво часопису «Березіль», Видавництво М. П. Коць. Харків – Нью-Йорк, 2001; Забужко О., Шевельов Ю. Вибране листування на тлі доби: 1992–2002: з додатками, творами, коментарями, причинками до біографій та іншими документами. Київ: Вища Поліція, ВД Факт, 2011.

Українськомовна частина корпусу на сьогодні охоплює 104 документи українською мовою, містить 1 346 424 токени.

У створеному корпусі використано екстралінгвістичну й лінгвістичну розмітку. **Екстралінгвістична розмітка** корпусу об'єднує:

а) *метатекстові дані*. До них належать:

- поле *area* (*сфера*) з можливими значеннями *літературознавство*, *мовознавство*, *загальні праці*;
- *author* (*автор*) – у цьому корпусі тільки *Юрій Шевельов*, створено для можливості об'єднання з іншими корпусами;
- *genre* (*жанр*): есе, стаття, монографія, інтерв'ю, передмова, виступ, доповідь, спогади, стаття, виступ, стаття, доповідь, лист, вступне слово;
- *name* (*назва твору*);
- *source* (*джерело*);
- *style* (*стиль*): публіцистичний, науковий та епістолярний;

- *type (tin)*: мовою оригіналу, переклад з англійської, переклад з німецької, переклад з французької;
 - *year (рік)*;
- б) *структурну розмітку*. Корпус містить дані:
- про межі тексту в тегах <doc> – </doc>;
 - про межі абзацу в тегах <p> – </p>;
 - межі речень у тегах <s> – </s>;
 - спеціальний тег <g/> позначає розділові знаки, які не відокремлюються пробілом від попереднього токена.

Лінгвістична розмітка аналізованого корпусу текстів Юрія Шевельова сьогодні є результатом автоматичного морфологічного аналізу й лематизації, здійснених за допомогою авторських інструментів. Структура стандартного тегу до кожного токена є такою: на першій позиції стоїть позначка граматичного класу слова, далі – позначки підкласів, усі позначки – односимвольні латиницею або цифрами, за кожним підкласом закріплена позиція, яка не змінюється для різних класів.

Наприклад, для слова *конференцією* тег має вигляд – Izzooin1m (рис. 1).

іменник		жіночий рід		орудний відмінок		не-істота		м'яка група
I	z	z	o	o	i	n	l	m
	загальна назва		однина		іменниковий тип відмінювання		перша відміна	

Рис. 1. Розшифрування тегу для словоформи «конференцією»

Таблиця 1

Класи слів у корпусі текстів Юрія Шевельова

іменник	I.*
дієслово	D.*
дієприкметник	Dk.*
дієприслівник	Ds.*
прикметник	K.*
займенник	Z.*
прислівник	S.*
прийменник	J.*
сполучник	P.*
числівник	C.*
частка	T.*
вигук	W.*
аббревіатура	A.*
решта	R.*

Таблиця 2

Позначки граем роду в корпусі текстів Юрія Шевельова

ч. р.	..c.*
ж. р.	..z.*
с. р.	..s.*

значками в тегах у табл. 2.

Кількісні параметри створеного корпусу є такими: у 104 документах зібрано 1 346 424 tokenів, з яких 1 037 949 – це слова у 66 039

Докладно систему позначок для тегів викладено в дослідженні (Загнітко & Данилюк, 2013). Зокрема, класифікацію граматичних класів подано в табл. 1. Категорію роду описано позначками в тегах у табл. 2.

реченнях. Загальний лексикон охоплює 138 187 різних слівформ та розділових знаків, описаних 157 оригінальними тегами, і 32 172 лем.

У корпусному менеджері доступні типові функції, як-от *побудова конкордансу* на підставі простого пошуку, пошуку в лемах, пошуку фрази, слівформи, символу або певного шаблону, побудованого за допомогою регулярного виразу.

Рис. 2. Інтерфейс для вибору типу запиту в корпусі

Запит може спиратися на додатковий пошук у контексті з фільтруванням потрібних лем чи слівформ на відстані до 15 токенів управо чи ліво від головного слова.

Рис. 3. Інтерфейс для вибору контексту запиту в корпусі

Нарешті, пошук можна обмежити різними типами текстів, передбаченими екстралінгвістичною розміткою.

У побудованому конкордансі можливі різні типи сортування та фільтрування даних, частотний аналіз морфологічних позначок чи словоформ для леми тощо.

Іншою типовою функцією корпусного менеджера є частотний аналіз із можливістю вибору мінімальної чи максимальної частоти, частиномовних фільтрів, N-грамів.

Параметри словника частот

Корпус:

Підкорпус: [відомості створити новий](#)

Шукати атрибут:

вживати N-грами. Значення N: від до

сквати/вкласти під-N-грами

Параметри фільтрування:

Фільтрувати словник за: Регулярний вираз:

Мінімальна частота:

Максимальна частота: (0 = без обмежень зверху)

Білий список: Файл не вибрано

Чорний список: Файл не вибрано

Занести не-слова

Параметри виводу:

Показувані частоти: Кількість знайдень Кількість документів ARF

Тип виводу: Простий

Ключові слова

Опорний (під)корпус: (цілий корпус)

Давати перевагу: рідким словам частим словам

Змінити показувані атрибут(и)

Можна вибрати один або декілька показуваних атрибутів. Ця опція може бути часовитратною.

Рис. 4. Інтерфейс для побудови частотного словника в корпусі

Інструмент частотного аналізу дає змогу, крім того, вибрати всі леми та всі словоформи.

Титульна Пошук Словник частот Про корпус Серверні задачі Інструкція 	Словник частот Корпус: Корпус текстів Юрія Шевельова (Шереха) Total number of items: 22,218 Сторінка <input type="text" value="1"/> Перейти Наступна >																																									
	<table border="1"> <thead> <tr> <th>word</th> <th>частота</th> </tr> </thead> <tbody> <tr><td>і</td><td>26.971</td></tr> <tr><td>в</td><td>22.811</td></tr> <tr><td>не</td><td>20.855</td></tr> <tr><td>з</td><td>14.831</td></tr> <tr><td>що</td><td>13.703</td></tr> <tr><td>на</td><td>13.464</td></tr> <tr><td>й</td><td>10.224</td></tr> <tr><td>до</td><td>9.251</td></tr> <tr><td>у</td><td>8.249</td></tr> <tr><td>а</td><td>7.040</td></tr> <tr><td>як</td><td>6.548</td></tr> <tr><td>це</td><td>5.777</td></tr> <tr><td>його</td><td>5.114</td></tr> <tr><td>про</td><td>5.075</td></tr> <tr><td>він</td><td>4.950</td></tr> <tr><td>але</td><td>4.659</td></tr> <tr><td>я</td><td>4.074</td></tr> <tr><td>від</td><td>3.920</td></tr> <tr><td>було</td><td>3.840</td></tr> <tr><td>тільки</td><td>3.444</td></tr> </tbody> </table>	word	частота	і	26.971	в	22.811	не	20.855	з	14.831	що	13.703	на	13.464	й	10.224	до	9.251	у	8.249	а	7.040	як	6.548	це	5.777	його	5.114	про	5.075	він	4.950	але	4.659	я	4.074	від	3.920	було	3.840	тільки
word	частота																																									
і	26.971																																									
в	22.811																																									
не	20.855																																									
з	14.831																																									
що	13.703																																									
на	13.464																																									
й	10.224																																									
до	9.251																																									
у	8.249																																									
а	7.040																																									
як	6.548																																									
це	5.777																																									
його	5.114																																									
про	5.075																																									
він	4.950																																									
але	4.659																																									
я	4.074																																									
від	3.920																																									
було	3.840																																									
тільки	3.444																																									
Стягнути Змінити параметри																																										

Рис. 5. Фрагмент словника частот словоформ із корпусу Юрія Шевельова

Висновки

Отже, створений корпус Юрія Шевельова є дослідницьким, повнотекстовим і динамічним, має екстралінгвістичну та лінгвістичну розмітку. З-поміж важливих функцій варто виокремити: а) побудову конкордансу на підставі простого пошуку, пошуку в лемах, пошуку фрази, словоформи, символу або певного шаблону, створеного за допомогою регулярного виразу; б) частотний аналіз для словоформ, лем та тегів.

Нині колектив авторів працює над усуненням помилок автоматичного морфологічного аналізу й наповненням англійськомовної частини корпусу. Перспективою дослідження є створення вичерпного корпусу текстів Юрія Шевельова, що охоплюватиме документи українською та англійською мовою та стане надійним підґрунтям для вивчення особливостей мовлення видатного вченого.

Список використаної літератури

- Загнітко, А. П., Данилюк, І. Г. (2013). Корпус текстів граматичної службовості. В *Прикладна лінгвістика та лінгвістичні технології: MegaLing-2012* (с. 102–112). Київ: УМІФ.
- Караулов, Юрій. (1987). *Русский язык и языковая личность*. Москва: УРСС.

- Мозер, Міхаель. (2014). *Юрій Шевельов на дорозі відрядництва*. Взято з <http://www.historians.in.ua/index.php/en/ukrayinska-mova/1066-mikhael-mozer-yurii-shevelov-na-dorozi-vidradianshchennia>.
- Трифонов, Р. А. (2009). Метамовні фрагменти спогадів Юрія Шевельова – репрезентанти індивідуальної картини світу лінгвіста. *Вісник Харківського національного університету ім. В. Н. Каразіна*, 843 (55), 19–26.
- Трифонов, Роман. (2015). Лінгвокультурні складники індивідуального мовлення Юрія Шевельова (на матеріалі есе та листів). Взято з <http://www.historians.in.ua/index.php/en/ukrayinska-mova/1697-roman-tryfonov-linhvokulturni-skladnyky-indyvidualnoho-movlennya-yuriya-shevelova-na-materiali-ese-ta-lystiv>.
- Rychlý, Pavel, & Smrž, Pavel. (2004). Manatee, Bonito and Word Sketches for Czech. In *Proceedings of the Second International Conference on Corpus Linguistics* (pp. 124–132). Saint-Petersburg: Saint-Petersburg State University Press.

References

- Karauov, Yuri. (1987). *Russkyi yazyk i yazykovaia lychnost [Russian language and language personality]*. Moskva: URSS [in Russian].
- Mozer, Mikhael. (2014). *Yurii Shevelov na dorozhi vidradianshchennia [George Shevelov on the Soviet-away-road]*. Retrieved from <http://www.historians.in.ua/index.php/en/ukrayinska-mova/1066-mikhael-mozer-yurii-shevelov-na-dorozi-vidradianshchennia> [in Ukrainian].
- Rychlý, Pavel, & Smrž, Pavel. (2004). *Manatee, Bonito and Word Sketches for Czech*. In *Proceedings of the Second International Conference on Corpus Linguistics* (pp. 124–132). Saint-Petersburg: Saint-Petersburg State University Press.
- Tryfonov, R. A. (2009). Metamovni frahmenty spohadiv Yuriiia Shevelova – reprezentanty indyvidualnoi kartyny svitu linhvista [Metamorphic Fragments of George Y. Shevelov's Memoirs – Representatives of the Linguist's Individual Picture of the World]. *Visnyk Kharkivskoho natsionalnoho universytetu im. V. N. Karazina [Bulletin of Kharkiv National University named by V. N. Karazin]*, 843 (55), 19–26 [in Ukrainian].
- Tryfonov, Roman. (2015). *Linhvokulturni skladnyky indyvidualnoho movlennia Yuriiia Shevelova (na materialii ese ta lystiv) [The linguistic and cultural components of George Y. Shevelov's individual speech (based on essays and letters)]*. Retrieved from <http://www.historians.in.ua/index.php/en/ukrayinska-mova/1697-roman-tryfonov-linhvokulturni-skladnyky-indyvidualnoho-movlennya-yuriya-shevelova-na-materiali-ese-ta-lystiv> [in Ukrainian].
- Zahnitko, A. P., & Danyliuk, I. H. (2013). Korpus tekstiv hramatychnoi sluzhbovosti [Text corpus for grammatical auxiliary]. In *Prykladna linhvistka ta linhvistychni tekhnologii [Applied linguistics and linguistic technologies]: MegaLing-2012* (pp. 102–112). Kyiv: UMIF [in Ukrainian].

Abstract

Illya Danyliuk, Anatoliy Zahnitko, Hanna Sytar

GEORGE Y. SHEVELOV'S TEXT CORPUS: STRUCTURE, FUNCTIONS, NAVIGATION

Background. In modern linguistics, the questions of the linguistic portrait of the person, the establishment of its structural and functional components are actualized. The study of eminent linguistic persons is very important, and

George Y. Shevelov is one of them – as an eminent linguist, historian of literature, and critic. The scientific task to establish the linguistic corpus structure of George Y. Shevelov's discursive practices is highly motivated.

Purpose. The **purpose** of the article is to establish and implement an experimental and research model of text corpus of George Shevelov's linguistic person with the determination of the main corpus components, patterns of free linguistic text navigation.

Methods. The method of discursive analysis, body content analysis are used.

Results. Text corpus contains all the Ukrainian-language published texts by George Y. Shevelov, it was created on the free NoSketch Engine corpus manager and is available on the server of the Department of General and Applied Linguistics and Slavic Philology in Vasyl' Stus Donetsk National University at corpora.donnu.edu.ua.

The quantitative parameters of the created corpus are as follows: 104 documents contain 1 346 424 tokens, and 1 037 949 are words in 66 039 sentences. The total lexicon includes 138 187 different word forms and punctuation marks, 157 original tags, and 32 172 lemmas.

Conclusions and Prospects. The created corpus of George Y. Shevelov is research type, full-text and dynamic, has extralinguistic and linguistic markings. Among the important features: a) building a concordance based on simple search, search in lemmas, search for a phrase, wordform, symbol or a certain pattern created using the regular expression; b) frequency analysis for word forms, lemmas, and tags.

The prospect of the research is the creation of a comprehensive George Y. Shevelov's Texts Corpus, which will cover documents in Ukrainian and English and will be a reliable basis for studying the features of the outstanding scientist's speech.

Keywords: concordance, text corpus, linguistic person, linguistic personality, corpus marking, frequency analysis, George Y. Shevelov.

Матеріал надійшов 26.07.2019