

		сучасної української художньої прози <sup>2</sup>		русько-го языка <sup>3</sup>	української художньої прози <sup>4</sup>	
сповіщати	5 (38%) сообщать	3 (12%)	1 (2%)	6	3	1
сповістити	9 (69%) сообщить	8 (32%)	3 (5%)	27	12	3
повідомляти	–	4 (16%)	1 (2%)	–	4	1
повідомити	–	11 (44%)	7 (12%)	–	19	8
інформувати	– информировать	–	1 (2%)	–	–	1
сприймати	1 (8%) воспринимать	10 (40%)	3 (5%)	1	13	3
сприйняти	– воспринять	10 (40%)	2 (3%)	–	15	2
довідуватися	7 (54%) узнавать	–	1 (2%)	17	–	1
довідатися	13 (100%) узнать	15 (60%)	11 (18%)	131	36	13
таїти	4 (31%) тайть	3 (12%)	1 (2%)	6	3	1
пам'ятати	11 (85%) помнить	24 (96%)	27 (45%)	70	97	44
знати	13 (100%) знать	25 (100%)	60 (100%)	568	1323	567

Таким чином, наша вибірка майже повністю підтверджує дані розглянутих частотних словників, пропонуючи нові можливості дослідження частотності вживання слів.

#### Література

1. Демська-Кульчицька О. Репрезентативність як ознака текстового корпусу // УМ. – 2005. – № 3.
2. Демська-Кульчицька О., Балог В. Національний корпус української мови: підкорпус художньої літератури // Лексикографічний бюлетень. – Київ, 2005. – Вип. 12.
3. Словник української мови. – Т. 3. – К.: Наук. думка, 1972. – С. 642–644.
4. Частотний словник сучасної української художньої прози / У 2-х томах. – Т.1. – К.: Наук. думка, 1981. – 864 с.
5. Частотний словник сучасної української художньої прози / У 2-х томах. – Т.2. – К.: Наук. думка, 1981. – 856 с.
6. Частотный словарь русского языка / Ред. Л. Н. Засорина. – М.: Русский язык, 1977. – 936 с.

**Т. Монахова\***

Інститут української мови НАН України (Київ)  
УДК 161.2.81'322

#### КОРПУС ТВОРІВ ВАЛЕРІЯ ШЕВЧУКА: ФОРМАТ РОЗМІТКИ БАЗОВИХ КОНЦЕПТІВ

У статті розглянуто проблему корпусного підходу до вивчення концептів. Пропонуються принципи анотування семантичних реалізацій концепту в художньому тексті. *Ключові слова:* концепт, семантична анотація.

Корпусна лінгвістика сьогодні пропонує свої методи для дослідження різних аспектів мови: її морфологічної будови, синтаксичного устрою, лексичних значень тощо. Ми здійснюємо спробу застосування корпусних методів до вивчення базових концептів мови творів Валерія Шевчука. Наше дослідження передбачатиме створення авторського корпусу мови Валерія Шевчука із анотуванням формально-сміслових параметрів базових концептів. Планований корпус творів Валерія Шевчука згідно з усталеними класифікаціями текстових корпусів [2: 25] буде ілюстративним, а отже створюватиметься після визначення й детального вивчення досліджуваних концептів.

**Концепт** — це складне абстрактне явище духовної культури, людської психіки, моральних принципів тощо, яке базується на осмисленні його тлумачень в усі періоди

\* © Т. Монахова, 2006

історії людства відповідно до національної конотації, індивідуального досвіду людини. Це свідоме оперування культурною метайнформацією через використання загальнолюдських, національних і особистісних смислів. Концепт може презентуватися символами, образами, реалізується у фразеологічних одиницях, закріплюється в культурних традиціях. Його лінгвальна реалізація здійснюється через вживання назв концептів (імен концептів) у прямих і переносних значеннях.

Виходячи з такого розуміння концептів, їх аналіз є передовсім аналізом смислів, які нашаровуються в структурі самого концепту, а точніше їх реалізації в конкретному художньому тексті (корпусі художніх текстів). Процедура семантичного аналізу смислів концепту можна продемонструвати на прикладі концепту *шлях* у творчості Валерія Шевчука, зокрема в романі «Дім на горі».

Формат розмітки становить чи не основну проблему корпусного анотування. Якщо схеми морфологічного маркування текстових даних розроблені досить чітко [1; 2], то принципи анотації концептів у лінгвоукраїністиці ще потребують з'ясування. Отже, спробуємо запропонувати певні підходи до реалізації розмітки формально-смислових параметрів базових концептів у корпусі текстів Валерія Шевчука на прикладі концепту *шлях*.

Для здійснення так званої концептної анотації необхідно зробити таке:

1. У текстах В. Шевчука визначити всі лінгвальні реалізації концепту, наприклад: «Йому здалося, що ця вулиця — не просто шлях, який так собі кладеться, весь простір наповнювався тими золотистими хвилями, і чи від блимливих калюж серед дороги, чи від повітря видалося йому, що шлях, яким ішов, дивно запрозорів».

2. Визначені контексти згрупувати за прямим і переносним значенням слова-імені концепту (*шлях, дім, дорога, стежка тощо*), наприклад: а) «Марія подумки проміряла той шлях, що пройшли його разом, тихий смуток заповнив її радість» — переносне значення (шлях = подружнє життя); б) «Вряди-годи озиралися вони на синю дорогу, по якій ішов Микола, приставляли долоні дашком і вдивлялися» — пряме значення.

3. Слова-імена концептів з прямим значенням сортувати в такий спосіб: другорядні прямі значення — основні прямі значення. **Основне** (центральне) **пряме значення** — це таке пряме значення, яке організовує у тексті простір, є частиною хронотопу або конкретним образом твору. **Приклад**: «Лізніше він пройде по цій дорозі ще раз». **Другорядне пряме значення** — це таке вживання слова в прямому значенні, яке не впливає на розвиток подій у творі; це супровідне, обставинне пряме значення. **Приклад**: «Ген віддалік на далекій курній дорозі їхав віз, тягнучи за собою довгий білий шлейф».

4. Слова-імена концептів з переносним значенням розподілити таким чином: пасивне переносне значення — активне переносне значення. **Пасивне переносне значення** — вживання слова-імені концепту фігурально з метою створення художнього ефекту. **Приклад**: «Саме туди вела синя, мерехтлива дорога, по якій ішов» — метафора й епітет зі словом «дорога». **Активне переносне значення** — фігуральне слововживання, яке має програмний, лейтмотивний смисл. Активні переносні значення реалізують смисли концепту. **Приклад**: «Він таки був тут, біля їхніх вікон, а зараз уже повільно відходить по синій дорозі» — дорога в значенні «потойбічне життя, смерть»).

5. Окрему категорію становлять подвійні смисли, тобто коли у вислові пряме основне значення накладається на активне переносне. Таке значення можна назвати символічним. Власне, **символічні значення** становлять смисли концепту. **Приклад**: «Серед неба світив місяць, і був він у ту ніч такий яскравий, що дорога, по якій вони йшли, засвітилася синім вогнем. Микола обережно взяв її під руку, і вона довірливо подалася до нього: мали пройти по тій дорозі десять років. Через рік до них прилучиться на їхньому шляху ще один супутник, і наприкінці десятого року стане їх семеро» — а) конкретна дорога, на якій зустрілися і покохалися парубок і дівчина; б) подружнє життя, народження дітей.

6. Символічні значення далі характеризувати за такими ознаками:

— іконічні (або візуальні) [3: 98], тобто всі зовнішні характеристики, виражені прикметниками, прислівниками тощо. **Приклади**: *мерехтлива дорога, синя дорога тощо*.

— функціональні ознаки [3: 98] — дійові характеристики, виражені найчастіше дієсловами. **Приклади**: *шлях кладеться, вела дорога, дорога засвітилася вогнем тощо*.

— етнічні ознаки — специфічні національні характеристики: *долати важку дорогу (за асоціацією з чумаками, бурлаками); шлях полотном стелиться тощо*.

Процедуру семантичного аналізу концептів можна подати схематично:

Таблиця 1.

ІМ'Я КОНЦЕПТУ			
Пряме значення		Переносне значення	
Другорядне значення	пряме	Основне значення	Активне переносне значення
		СИМВОЛІЧНЕ ЗНАЧЕННЯ	
		Пасивне переносне значення	

Для кодування смислів за пропонованими параметрами послуговуємося латинськими символами:

1. C — concept — концепт, ім'я концепту;
2. L — literal meaning — пряме значення;
3. F — figurative meaning — переносне значення;
4. M — minor — другорядне;
5. B — basic — основне;
6. A — active — активне;
7. P — passive — пасивне;
8. S — symbol — символічне;
9. Ab — активне переносне значення, що накладається на основне пряме;
10. Ba — основне пряме значення, на яке накладається активне переносне.

Пропонований мінімальний набір релевантних класифікаційних критеріїв та їх кореляція можуть бути представлені в такий спосіб:

Таблиця 2

№	I	II	III	IV	V
кри-терій	Назва твору	Ім'я концепту	Тип значення (прям./непряме)	Основне/другор. Активне/пасивне	Лінгвальна реалізація концепту
те-ги	Дім на горі DNG	C	L / F	M / B A / P Ab/Ba	Конкретна цитата з твору
н.п.	DNG	R (road)	L	B	Ніхто не знав про його хід цією дорогою...
н.п.	DNG	R	L	M	По тому вони побачили батька, що повертався з вулиці.
н.п.	DNG	R	F	P	... дорога, по якій вони йшли, засвітилася синім вогнем.
н.п.	DNG	R	F	A	Хто зна, може, в такий спосіб її дух долав отак кілометр за кілометром синю дорогу — час відлетілий та забутий...
н.п.	DNG	R	F	Ab	Вона подивилася далі, собі під ноги, і знову побачила там шматок синьої дороги, по якій йшла вона і простували й інші.

н.п.	DNG	R	L	Va	Вже давно думала про дорогу. Ввижалася їй нескінченним білим змієм, бо простягалась у визорі звідсіля, від їхнього передмістя, туди, в далекий безмежний світ, думати про який було страшно й лячно.
------	-----	---	---	----	--

Таким чином, спостерігаються кореляції F із L і, відповідно, A / P із M / B.

При цьому  $S = Va$ ,  $S = Ab$ ,  $Va = Ab = S$ .

Відповідно до зазначеного, корпусне анотування концептів має такий вигляд: **Наша присутність — це і є та дорога, яку я так часто останнім часом бачу {DNG/R/F/A}** («Дім на горі», концепт «Шлях»), переносне значення, активне переносне значення).

Таким чином, концептні смисли реалізують значення S (Va, Ab) та FA, тобто символічні значення і активні переносні значення. Можливо, символічні значення, визначені як поєднання активних переносних і основних прямих значень, є прикметною ідеостиллю Валерія Шевчука, який свідомо накладає абстрактні глибинні смисли на конкретні образи у своїх творах. Тому ефективність пропонованої схеми аналізу на матеріалі творів інших авторів потребує перевірки.

#### Література

1. Автоматизация анализа научного текста / В. А. Вербицкий, Т. А. Грязнухина, Н. П. Дарчук и др. – К.: Наукова думка, 1984. – 256 с.
2. Демська-Кульчицька О. Основи національного корпусу української мови. – К., 2005. – 219 с.
3. Іващенко В. Концепт-символ «кобзар» у життєтворчості Т. Г. Шевченка // Матеріали IV Міжнародного семінара «Шевченківський Петербург». – СПб., 2005. – 186 с.
4. Никитин М. Курс лингвистической семантики: Учебное пособие для студентов, аспирантов и преподавателей лингвистических дисциплин в школах, лицеях, колледжах и вузах. – СПб., 1996. – 760 с.

**Ł. Grabowski\***

Institute of East-Slavonic Studies University of Opole (Opole, Poland)  
УДК 81'322

### COMPARATIVE CHARACTERISTICS OF POLISH AND RUSSIAN LANGUAGE CORPORA

*The aim of this paper is to present and compare the most representative language corpora of Polish and Russian according to selected criteria determining both their potential as a source of linguistic data for various types of linguistic analyses and their availability for researchers. Moreover, the present paper indicates areas for improvement as far as the possibilities offered by the corpora and access to them are concerned.*

#### 1. Introduction

Corpus linguistics, which is both a branch of linguistics and a methodology of research, concentrates on the study of texts with the use of dedicated computer software [4:143]. Thus, any collection of texts stored in electronic form is called a corpus. The linguistic data stored in the corpus is the actual collection of electronic words, which are classified in terms of types and tokens [2:16]. This typology is important in that a token is every running word-form (segment) the corpus is composed of, whereas a type is a group of the same tokens (eg. the sentence 'Моя мама живёт в городе, а моя тётя живёт в деревне' consists of 8 types and 11 tokens).

Since language corpora, unlike single and coherent texts, are objects of complex and multidimensional character, the answer to the question how similar or different they are will indubitably be complex and multidimensional itself. Thus, the study aimed at comparing corpora that represent two different languages will be even more challenging and complex. Corpora as objects of comparison will always be similar in some aspects and different in others. For the comparison to render objective observations, one has to adopt a framework that would serve as criteria for such a study. Although there have been multiple quantitative methods based on word frequencies and ngram frequencies [1], purely statistical approach to a comparison of different multilingual corpora falls beyond the scope of the current paper. Thus, its aim is to present possibilities offered by the most representative commercial corpora representing Polish and

\* © Ł. Grabowski, 2006