

*салатовий, темно-синій, жовто-зелений, біло-блакитний, чорно-червоний, червоно-темний, рожево-коричневий, синьо-рожевий.*

У доповіді було досліджено лексичний рівень роману «Ворошиловград», у якому ідіостиль С. Жадана проявляється у вживанні великої кількості розмовно-просторічних слів, жаргонізмів та обценної лексики та наскрізному протиставленні темного та світлого чере, що пояснюється вибр антонімічні пари та оксюмори.

Дослідження ідіостилю Сергія Жадана показало його внесок як автора у розвиток сучасної української літературної мови та проявило увесь спектр мовних виражально-зображальних засобів, якими він послуговується для творення реалістичної і водночас метафоричної картини художньої дійсності.

#### **Список літератури**

1. Жадан С. Ворошиловград : роман / Сергій Жадан. – Х. : Фоліо, 2010. – 441 с.
2. Ставицька Л. О. Українська мова без табу : словник нецензурної лексики та її відповідників : обценізви, евфемізви, сексуалізви / Л. О. Ставицька. – К. : Критика, 2008. – 454 с.

***Юлія Старишко***

*Український мовно-інформаційний фонд*

*Національної академії наук України*

#### **ТЕХНОЛОГІЯ ОБРОБКИ ТЕКСТУ НА РІВНІ ГРАФЕМ**

Співробітниками Українського мовно-інформаційного фонду Національної академії наук України (УМІФ НАН України) розробляється система природномовного аналізу корпусного типу [4]. Ця система одночасно є платформою нового типу, що відповідає наступним принципам:

- мультиформатність;
- орієнтованість на віртуалізацію;
- використання Grid-технологій;

- використання методів машинного навчання;
- кросплатформність;
- сервісорієнтованість;
- більш повне завантаження апаратних потужностей;
- гнучкість розподілу ресурсів;
- орієнтованість на сучасні запам'ятовуючі пристрої постійного зберігання;
- розширення можливостей пошуку (наприклад, фасетний пошук, пошук із символом «\*»), обробка пошукового запиту з неточністю);
- новітні методи та підходи до створення структур зберігання даних, алгоритмів оперування масивами в пам'яті, індексування, сортування та об'єднання масивів текстових даних тощо.

Обробка тексту на рівні графем – початковий етап обробки тексту, у процесі якого визначаються такі елементи графематичної структури (слова, пунктуаційні знаки, числові комплекси та ін. Виділяють такі основні функції графематичного аналізу: розбиття тексту на графеми, визначення меж речень, розпізнання слів і службових графем (наприклад, пунктуаційних знаків), розпізнавання прізвищ, власних назв, абревіатур та скорочень тощо) [6, с. 69].

З метою розробки технології аналізу тексту на рівні графем здійснено дослідження наявних на сьогоднішній день теоретичних і практичних підходів. Найбільш значущими у цій галузі є напрацювання науково-дослідницької групи проекту ООО «ЭР СИ О» [5] (А. Ермаков, В. Плешко); АОТ [2] (А. Сокірко, С. Толдова, Н. Кецагіс, І. Зданська й інші); ДІАЛІНГ (А. Сокірко, І. Ножов, Б. Кобріцов, М. Рубінштейн, Г. Дурново, О. Шалімов та ін.); системи ФРАП [3]; напрацювання Консонціуму Юнікод [1] (Unicode Standard Annex #29: Unicode Text Segmentation). Основою технології аналізу на рівні графем є напрацювання співробітників УМІФ та правила сегментації графем за граничними символами з вищезгаданого стандарту Юнікод.

Здійснений нами аналіз праць провідних дослідників у галузі автоматичної обробки тексту дозволяє зробити висновок, що використання методу так званих

словників графематичних неоднозначностей із застосуванням правил сегментації тексту є найбільш оптимальним при графематичному аналізі. Виділяємо такі словники: Словник прізвищ, Словник службових слів в іншомовних особових назвах (напр.: *ван, да, де, ді, дю, дер, ед, ель, ла, ле, фон, пор.* Людвіг *ван* Бетховен, *де ла* Куева, *фон дер* Гольц; скорочена частка *д* та ірландська частка *о* пишуться з власними іменами через апостроф: *Д'Аламбер, Д'Артаньян, Д'Обіньє*; частки *Ван-, Мак-, Сан-, Сен-*, які передують прізвищам і, разом із тим, становлять їх невід'ємну частину: *Ван-Дейк, Мак-Магон, Мак-Клор, Сан-Мартін, Сен-Сімон*); Словник графічних скорочень (*с.-г.* — сільськогосподарський; *півд.-зах.* — південно-західний); Словник скорочень (*р.* — річка або рік, *ст.* — століття) та ін.). Зазначений перелік, безумовно, збільшиться в процесів створення.

Як приклад наведемо правила занесення претендентів до Словника прізвищ:

- а) якщо крапка стоїть після великої літери або велика літера і крапка, велика літера і крапка), а наступні графеми — велика та мала літери (наприклад, *Т. Г. Шевченко*);
- б) якщо перше слово з *п* графем (але не менше однієї), а наступні графеми — велика літера з крапкою (або: велика літера і крапка, велика літера і крапка) (наприклад, *Шевченко Т. Г.*).

Ці комплекси є претендентами до вищезазначеного словника. Претендентами, оскільки це правило не може охоплювати всі випадки. Наприклад: «Додаток А. Додаток А. Відомості про авторів винаходу» — відповідно до описаного вище правила *Додаток А* стане претендентом до словника прізвищ. Тому необхідним є створення та підключення до системи словника прізвищ, що значно полегшить верифікацію одержаних, відповідно до описаного правила, претендентів на внесення до Словника прізвищ.

Розроблена нами технологія виокремлення графематичних неоднозначностей та створення на їх основі відповідних словників є необхідною процедурою, що має передувати предморфологічному аналізу.

### Список літератури

1. Unicode Standard Annex #29: Unicode Text Segmentation [Електронний ресурс]. — Режим доступу: <http://www.unicode.org/reports/tr29/>. — Назва з екрана.
2. Автоматична обробка тексту [Електронний ресурс]. — Режим доступу: <http://aot.ru/>. — Назва з екрана.
3. Леонтьева Н.Н., Система французско-русского автоматического перевода (ФРАП): лингвистические решения, состав, реализация: Сб. науч. трудов / Моск. пед. ин-т иностр. языков. — 1986 [на обл.: 1987]. — Вып. 271 МП и ПЛ [Вып. 21]. — С. 6—25
4. Надутенко М. В. Віртуалізовані лексикографічні системи та їх застосування у прикладній лінгвістиці [Текст]: дис. канд. техн. наук : 10.02.21 / Надутенко Максим Вікторович ; НАН України, Нац. б-ка України ім. В. І. Вернадського. — Київ, 2016. — 284 с. : рис.
5. Проект ООО «ЭР СИ О» [Електронний ресурс]. — Режим доступу: [http://www.rco.ru/?page\\_id=4416](http://www.rco.ru/?page_id=4416). — Назва з екрана.
6. Седунов А. А. Модель графематического анализа в системе обработки естественного языка // Вестник Воронежского государственного университета. — 2007. — № 2. — С. 69—66
7. Широков В.А. та ін. Корпусна лінгвістика /В.А. Широков, О.В. Бугаков, Т.О. Грязнухіна, О.М. Костишин, М.Ю. Кригін; НАН України, Укр. мов.-інформ. фонд. — К.: Довіра, 2005. — 472 с.

*Ксенія Тараненко*

*Університет митної справи та фінансів*

### **ФУНКЦІЙНИЙ ПОТЕНЦІАЛ АНТОНІМІВ У КОМУНІКАЦІЇ**

Пріоритетною тенденцією розвитку мовознавства ХХІ ст. є вивчення одиниць, що закріпилися в мові, однак повною мірою реалізуються лише в мовленні, перебуваючи в постійній залежності від особистості мовця. Усе частіше мовознавці звертаються до вивчення прагматичних особливостей мовних одиниць, оскільки їхні значення в лінгвальній системі та в процесі комунікації істотно відрізняються. Через це спостерігається певне зміщення