автору вдається не тільки підкреслити характерну рису героя, але і надати особливого колориту зображуваній реальності, що ми і спостерігаємо у науково-фантастичній повісті О. Тесленка «Дьондюранг».

Таким чином, власна назва є багатоплановим компонентом художнього тексту, яка може реалізувати свої численні можливості як засіб творення образу відповідно до жанрових особливостей твору, авторських завдань і психологічних та інтелектуальних особливостей читача. Зважаючи на недостатню кількість наукових розвідок у галузі літературної ономастики, зокрема поетоніміки, вважаємо доцільним подальше вивчення власних назв та їхніх функцій на матеріалі науково-фантастичних творів українських письменників-фантастів XX століття.

Список використаних джерел

1. Ахманова О. С. Словарь лингвистических терминов / О. С. Ахманова. – М.: Советская энциклопедия, 1969. – 608 с.
2. Белей Л. Функціонально-стилістичні можливості української літературно-художньої антропонімії XIX – XX ст. / Л. О. Белей. – Ужгород, 1995. – 120 с.
3. Бондалетов В. Д. Русская ономастика / В. Д. Бондалетов. – М.: Просвещение, 1983. – 224 с.
4. Калінкін В. М. Теоретичні основи поетичної ономастики: автореф. дис. … д-ра філол. наук: 10.02.01 – рос. мова, 10.02.15 – заг. м-во. – К., 2000, – 24 с.
5. Михайлов В. Н. О специфике литературной ономастики / В. Н. Михайлов // Вопросы стилистики: Стилистика художественной речи: межвузовский научный сборник – Саратов: Изд. Сарат. ун-та, 1988. – С. 3-19.
6. [www.sviato.in.ua/names_a.phb]
7. [www.zname.org.ua/letter/m/?id=258]
8. [http://ukrslov.com/slovnyk_inshomovnyk_sliv/page/im.7655/]

Summary. *In the article the basic aspects of research of the proper names function in artistic texts are reflected. The purpose of this investigation is research of common onomastic's units in works of Ukrainian writers-fantasies, who were not under intent attention of researchers until now.*

Key word: *literary onomastics, proper name, function.*

Отримано: 7.10.2012 р.

УДК 81'276. 6.34 : 373.423 : 81 – 139

Г. В. Карнаух

ЗНЯТТЯ ГРАМАТИЧНОЇ НЕОДНОЗНАЧНОСТІ СЛОВОФОРМ ШЛЯХОМ АВТОМАТИЧНОЇ ОБРОБКИ ПРИРОДНОЇ МОВИ (НА МАТЕРІАЛІ ТЕКСТІВ НОРМАТИВНО-ПРАВОВИХ ДОКУМЕНТІВ)

Дослідження присвячено вивченню особливостей зняття граматичної неоднозначності словоформ на матеріалі текстів нормативно-правових актів. Запропоновано алгоритм ідентифікації словоформ, які можуть позначати і власні, і загальні назви. Виявлено специфіку поведінки лінгвістичного процесора при використанні триграм різних типів.

Ключові слова: *нормативно-правові документи, граматична омонімія, словоформа, граматичне значення, власна назва, загальна назва, триграми.*

Швидкий темп сучасного життя, спричинений бурхливим розвитком інноваційних технологій, супроводжується постійними змінами в мові, а це, у свою чергу, вимагає відображення цих змін (відповідно до вимог часу) у словниковій продукції. Тому оновлення фундаментального лексикону 11-томного Словника української мови на рубежі тисячоліть виявилося не лише актуальним мовознавчим завданням, але й узагалі нагальною потребою українського суспільства. Однак при цьому виникла «ціла низка конкретних питань – починаючи від визначення принципів укладання нового лексикону «поверх» старого і закінчуючи технологією та організацією виконання такого лексикографічного проекту» [3; 7, 3; 13]. Зрозуміло, що ефективне вирішення зазначених проблем викликало потребу створення певної узагальнювальної теоретичної схеми, спроможної надати уніфіковані засоби до їх розв'язання, а також необхідність побудови комп'ютерного інструментального комплексу для реалізації зазначеного проекту[1].

Головний редактор Словника української мови в 20 томах (СУМ-20)[2] академік В. М. Русанівський у Вступі до нього зазначав, що «нація усвідомлює себе тоді, коли вона створює порівняно повний словник своєї мови» [7, 6]. Саме тому основним завданням нового Словника, що функці-

онуватиме в паперовому й електронному варіантах, є максимально повне представлення лексичного, а головне – семантичного багатства сучасної української мови, «яке творилося від її зародження та відображене в письмових джерелах XIX, XX, XXI ст.» [7, 7].

Як відомо, для систематичного опису лексики потрібне її якомога повніше зібрання. Цю проблему вирішено завдяки створеному в Українському мовно-інформаційному фонді НАН України (УМІФ НАНУ) Українському національному лінгвістичному корпусу обсягом понад 64 млн слововживань [3]. Лінгвістичний корпус, який замінив собою традиційну лексичну картотеку, відіграв роль одного з основних джерел текстово-ілюстративного матеріалу до СУМ-20. Загалом, зазначена програмна бібліотека – це представлений в електронному вигляді, великий за обсягом, уніфікований, структурований, розмічений і філологічно компетентний масив мовних даних, доповнений системою керування – універсальними програмними засобами для пошуку різноманітної лінгвістичної інформації, перевагами якого є великий обсяг мовного матеріалу, що залучається до мовознавчого дослідження, комплексність, оперативність опрацювання зазначеного матеріалу та можливість прямого доступу до значної кількості лінгвістичних фактів [3; 7, 257].

Одним із важливих компонентів програмного забезпечення корпусу текстів є морфологічний аналіз словоформ, метою якого є визначення граматичних параметрів досліджуваної одиниці. Однак суттєвою перешкодою на шляху реалізації морфологічної розмітки корпусу є явище граматичної омонімії. Тому постає потреба створення програмних засобів, спрямованих на вирішення цієї проблеми [1; 3; 5; 6; 12].

Поставлене ще в минулому столітті, особливої гостроти це завдання набуває на сучасному етапі розвитку корпусної лінгвістики, оскільки виникає необхідність розроблення ефективних методик, які б ураховували сучасний стан лінгвістичних знань та комп'ютерних технологій.

Питання, пов'язані з вивченням й усуненням граматичної неоднозначності[3] словоформ, досліджено багатьма лінгвістами і на теоретичному, і на практичному рівнях. Теоретичний аспект (здебільшого в контексті загальних проблем мовної неоднозначності, омонімії та полісемії, зумовлених потребами лексикографії, лексикології та граматики) висвітлено в працях О. Потебні, Л. Авксентьєва, Ю. Апресяна, О. Ахманової, В. Виноградова, Г. Гнатюк, Т. Грязнухіної Н. Клименко, М. Кочергана, Л. Лисиченко, А. Лучик, М. Муравицької, В. Русанівського, О. Смирницького, О. Тараненка, Г. Уфімцевої, О. Шипнівської, В. Широкова, Д. Шмельова та ін. Практичний аспект (дослідження особливостей автоматичної ідентифікації текстових словоформ (машинний переклад, автоматичне коригування тексту тощо)) представлено у роботах Т. Аполлонської, С. Білокриницької, Т. Грязнухіної, А. Залізняка, Ю. Зеленкова, Т. Любченко, Ю. Марчука, Т. Молошної, О. Невзорової, Ю. Орєхова, Р. Піотровського[4] та ін.

Активна робота над розробленням формалізованих методів аналізу й усунення мовних неоднозначностей ведеться в УМІФі, в результаті якої створено програму автоматичної детермінації омонімічних одиниць із використанням статистичних методів [5; 6]. Присвоєння граматичних характеристик словоформам забезпечується Граматичним словником української мови при використанні алгоритму лематизації [11], внаслідок чого аналізована одиниця набуває набору граматичних параметрів (υ, g), де υ – частина мови, а g – граматичне значення. Омонімічна словоформа отримує декілька таких наборів (для кожного значення окремо) у кількості від 1 до n.

Метою цього дослідження є аналіз поведінки лінгвістичного процесора[5] при використанні допоміжних програмних функцій, спрямованих на збільшення відсотка точності зняття граматичної омонімії в текстах законодавчих документів шляхом автоматичного опрацювання природної мови.

Практичне використання зазначеного програмного засобу показало, що стилістичні ознаки, а також особливості орфографічного та пунктуаційного оформлення текстів суттєво впливають на результати усунення в ньому граматичної неоднозначності. Так, наявність граматичних (особливо орфографічних) помилок у тексті призводить до того, що аналізована одиниця не отримує набір граматичних характеристик, оскільки в граматичному словнику немає даних про неї. Більше того, з'являється імовірність неправильної детермінації інших словоформ у межах речення, адже внаслідок застосування принципу триграм[6] спрацьовує ланцюжковий ефект, коли ідентифікація (правильна чи ні) однієї одиниці впливає на маркування сусідніх.

Стилістичні особливості текстів при цьому також відіграють неабияку роль. Так, наприклад, для юридичних текстів, з огляду на специфіку законодавчої мови, можна виробити ряд додаткових функцій, які допоможуть видалити зайву інформацію з граматичного набору омонімічної словоформи. Оскільки граматичний словник досить широко представляє лексику української мови (включаючи розмовну, застарілу, рідковживану та ін.; власні назви нарівні з загальними), більшість словоформ визначаються програмою як омонімічні.

У писемному варіанті мови омоніми, що збігаються за звучанням, але належать до різних груп, виділених за принципом відношення власна / загальна назва, розрізняються на графічному рівні. І, оскільки запропонована програма автоматичної обробки текстів враховує декілька рівнів

мови (знаковий, фонетичний, лексичний, граматичний і т. ін. [6, с. 75]), існує можливість розрізнення графеми за написанням (велика / мала літера), а це, у свою чергу, дає змогу побудувати алгоритм зняття неоднозначності словоформ, що базуватиметься на принципі відношення власна / загальна назва[7].

Наприклад, словоформи[8]:

або:

1. А#бо – власна назва (друга назва міста Турку, Фінляндія), іменник середнього роду, однина (називний, родовий, давальний, знахідний, орудний, місцевий, кличний відмінки);

2. А#ба – власна назва (місто в Угорщині), іменник жіночого роду, кличний відмінок, однина;

3. або# – частка;

4. або# – сполучник.

батьки:

1. Батьки# – власна назва (населений пункт в Україні), множинний іменник, називний, знахідний, кличний відмінки;

2. ба#тько – загальна назва, іменник чоловічого роду, істота, множина, називний, кличний відмінки;

3. батьки# – загальна назва, множинний іменник, істота, називний, кличний відмінки.

Внаслідок використання алгоритму для словоформи ***або*** буде видалено значення 1і 2, а для ***батьки*** – 1.

Практичне застосування зазначеної функції викликало такі зміни у маркуванні нормативно-правового акту: кількість детермінованих одиниць збільшилася на 223, що складає 2,31 % від кількості маркованих словоформ і 1,59 % від загальної кількості словоформ у тексті. Для систем машинної обробки мовних матеріалів цей результат є суттєвим.

Крім того, такі ознаки законодавчої мови, як офіційність, конкретність змісту; чіткість, стислість, логічна послідовність, однозначність формулювань; несуперечність аргументації; усталеність та одноманітність форми документів; канцелярська лексика, відсутність емоційної образності, стандартні звороти (кліше); безособові та наказові форми; відсутність індивідуальних авторських рис; монологічна (рідше діалогічна) форма тексту [2, 55; 8; 10; 14] вимагають відповідного оформлення нормативно-правових документів на синтаксичному рівні. Тому доцільним є врахування розділових знаків, наявних у тексті при побудові допоміжного інструментарію (триграм у межах речення).

Застосування триграм 2-го типу (з урахуванням пунктуації) сприяло зменшенню кількості неправильно ідентифікованих словоформ порівняно з використанням триграм 1-го типу (без урахування), що привело до збільшення відсотка точності зняття граматичної неоднозначності словоформ у законодавчому тексті (видалено неправильне маркування для 191 одиниці), хоча і викликало усунення правильного маркування словоформ. Однак варто наголосити, що при розробленні систем автоматичного опрацювання мови важливим є не лише кількісна, а й якісна характеристика отриманих даних.

Таким чином, зменшення загальної кількості детермінованих одиниць (що включає правильне і неправильне маркування) не можна вважати негативними (небажаними) змінами. Усунення правильного маркування слід кваліфікувати як побічний результат, який порівняно з видаленням помилкового – набуває вторинного значення. У Таблиці 1 представлено детальний аналіз результатів, отриманих унаслідок застосування двох допоміжних програмних функцій одночасно, де (* +) – зіставлення двох текстів (позиція 1 відповідає тексту, в якому при усуненні граматичної омонімії використано алгоритм розрізнення написання словоформ (велика / мала літера) і 2 – застосування зазначеного алгоритму і триграм 2-го типу); «*» – неідентифікована словоформа, «+» – правильно ідентифікована словоформа, «–» – неправильно ідентифікована словоформа.

Таблиця 1

№		Кількість словоформ	Кількість словоформ, %
1.	Усунено неправильне маркування (– *)	158	1,12
2.	Усунено правильне маркування (+ *)	366	2,60
3.	Промарковано неправильно (помилки збігаються в обох текстах) (– –)	257	1,82
4.	Правильно промарковано неідентифіковані словоформи (* +)	271	1,92
5.	Неправильно промарковано неідентифіковані словоформи (* –)	166	1,18

6.	Правильно промарковано неправильно ідентифіковані словоформи (– +)		33	0,23
7.	Неправильно промарковано правильно ідентифіковані словоформи (+ –)		63	0,44
8.	Загальна кількість «позитивних» змін (сума № 1,4,6)		462	3,28
9.	Загальна кількість «негативних» змін	(сума № 2,5,7)	595	4,23
		(сума № 5,7)	229	1,62

Загальна інформація щодо кількості ідентифікованих омонімічних словоформ при застосуванні різних допоміжних програмних засобів при машинній обробці текстів подана в Таблиці 2.

Таблиця 2

№		Без застосування допоміжних програмних функцій	Із застосуванням алгоритму власна / загальна назва	Із застосуванням алгоритму власна / загальна назва та триграм 2-го типу
1.	Загальна кількість словоформ	14066	14066	14066
2.	Кількість унікальних словоформ	2885 (20,51 %)	2885 (20,51 %)	2885 (20,51 %)
3.	Загальна кількість маркованих словоформ	9424 (66,99 %)	9647 (68,58 %)	9514 (67,63 %)
4.	Кількість унікальних маркованих словоформ	1888 (13,42 %)	1939 (13,78 %)	1876 (13,33 %)
5.	Кількість неомонімічних словоформ	4016 (28,55 %)	4174 (29,67 %)	4174 (29,67 %)
6.	Кількість омонімічних словоформ	10050 (71,45 %)	98,92 (70,33 %)	98,92 (70,33 %)

Отже, аналіз отриманих даних показав: врахування специфічних особливостей текстів, на матеріалі яких здійснюється зняття граматичної неоднозначності словоформ машинним способом, позитивно впливає на результати маркування досліджуваних об'єктів. Вдале практичне застосування створених на основі цих особливостей допоміжних програмних функцій дозволяє констатувати, що вони здатні продуктивно взаємодіяти при одночасному їх використанні.

Очевидно, що обсяги мовного матеріалу в межах наукових досліджень, зокрема лінгвістичних, весь час зростатимуть, тому необхідність оперативного, комплексного, а головне – якісного його опрацювання, і, відповідно, потреба подальшого вдосконалення систем автоматичної обробки природної мови залишаються актуальними й надалі.

Примітки

1. Тлумачний словник української мови в 20-ти томах є одним із центральних проектів програми створення Національної словникової бази, ініційованої Указом Президента України від 7 серпня 1999р. № 967., координатором якого є Український мовно-інформаційний фонд НАН України.
2. Частину Вступу до СУМ-20, яку встиг написати В. М. Русанівський наведено в колективній монографії «Лінгвістичні та технологічні основи тлумачної лексикографії» [7, 6-8].
3. Неоднозначність або ж багатозначність трактується як родове поняття, що поєднує полісемію та омонімію [4, 93].
4. Усі прізвища подано за алфавітним принципом.
5. Сукупність штучних моделей природної мови, алгоритмів і програм, що описують будову та функціонування цих моделей, та технічні засоби, що реалізують цю модель [9, 10].
6. Кожне речення в тексті поділяється на трійки словоформ (триграми), які складаються з одиниці, що досліджується, і двох сусідніх (попередньої та наступної). Отже, при статистичному аналізі будь-якої словоформи враховуються показники всіх трьох складників триграми.
7. Принцип дії зазначеного алгоритму детально описано в роботі Карнаух Г. В. Велика і мала літери як засіб розрізнення неоднозначних словоформ при автоматичній дизамбігуації / Г. В. Карнаух // Проблеми граматики і лексикології української мови : Збірник наукових праць. – К. : НПУ ім. М. П. Драгоманова, 2011. – Вип. 8. – С. 26-36.
8. Приклади наведено з Конституції України [Конституція України. Закон від 28.06.1996 № 254к/96-ВР / [Верховна Рада України]. – Режим доступу : http://zakon2.rada.gov.ua/laws/.

Список використаних джерел

1. Карнаух Г. В. Велика і мала літери як засіб розрізнення неоднозначних словоформ при автоматичній дизамбігуації / Г. В. Карнаух // Проблеми граматики і лексикології української мови : Збірник наукових праць. – К. : НПУ ім. М. П. Драгоманова, 2011. – Вип. 8. – С. 26-36.
2. Колеснікова І. Є. Українське ділове мовлення : навч. посібник / І. Є. Колеснікова, Ю. Ф. Прадід / [за ред. Ю. Ф. Прадіда]. – Сімферополь : ВД «АРІАЛ», 2009. – 210 с.
3. Корпусна лінгвістика / [В. А. Широков, О. В. Бугаков, Т. О. Грязнухіна та ін.] ; під ред. В. А. Широкова. – К. : Довіра, 2005. – 407 с.
4. Кочерган М. П. Слово і контекст / Кочерган М. П. – Львів : Вища школа, 1980. – 182 с.
5. Крыгин М. Снятие грамматической омонимии в тексте с помощью статистических методов / Крыгин М., Шкурко В. Афанасьева О. // Прикладна лінгвістика та лінгвістичні технології : MegaLing-2009. – К. : Довіра, 2009. – 527 с.
6. Крыгин М. Текст на естественном языке как объект статистического анализа / М. Ю. Крыгин// Біоніка інтелекту : наук.-техн. журнал. – 2000. – № 1 (72). – С. 75-82.
7. Лінгвістичні та технологічні основи тлумачної лексикографії / [В. А. Широков, В. М. Білоноженко, О. В. Бугаков та ін.] ; під ред. В. А. Широкова. – К. : Довіра, 2010. – 296 с.
8. Мільченко О. С. Семантичні девіації в нормативно-правових текстах: дис. … канд. філол. наук: 10.02.01 / Мільченко Ольга Сергіївна. – К., 2011. – 178 с.
9. Пиотровский Р. Г. Инженерная лингвистика и теория языка / Пиотровский Р. Г. – Ленинград : Наука, 1979. – 112с.
10. Чулінда Л. І. Аспекти використання лінгвістичних знань в юриспруденції / Л. І. Чулінда // На терені юридичної і філологічної науки : [Зб. наук. праць, присвячений 50-річчю від дня народження і 25-річчю наук.-пед. діяльності професора Прадіда Ю. Ф.]. – Сімферополь : Еліньо, 2006. – С. 89-93.
11. Шевченко И. В. Электронный грамматический словарь украинского языка. / Шевченко И. В., Рабулец А. Г., Широков В. А. // Труды Международной конференции «MegaLing-2005. Прикладная лингвистика в поиске новых путей», 27 июня – 2 июля2005 г. – Меганом, 2005. – С. 124-129.
12. Шипнівська О. О. Структурно-семантичні та функціональні характеристики міжчастиномовної морфологічної омонімії сучасної української мови: дис. … канд. філол. наук: 10.02.01 / Шипнівська Ольга Олександрівна. – К., 2007. – 238 с.
13. Широков В. А. Елементи лексикографії / В. А. Широков. – К. : Довіра, 2005. – 304 с.
14. Язык закона / [под. ред. А. С. Пиголкина]. – М. : Юрид. лит., 1990. – 192 с.

Summary. *This research describes the investigation of particulars for elimination of grammatical ambiguity of word-forms on the basis of legal acts. It is offered some algorithm of word forms identification which may designate both proper and common names. It is found specificity of the behavior of linguistic processor while using trigrams of different types.*

Keywords: *legal documents, grammatical ambiguity, word-form, grammatical meaning, proper name, common name, trigram.*

Отримано: 7.10.2012 р.

УДК 811.161.2’373.7

Т. В. Кедич

ОСОБЛИВОСТІ ІНДИВІДУАЛЬНО-АВТОРСЬКОГО ВИКОРИСТАННЯ ФРАЗЕОЛОГІЗМІВ В АСПЕКТІ КОНТЕКСТУАЛЬНОЇ РЕАЛІЗАЦІЇ В ІСТОРИЧНІЙ ПРОЗІ ДРУГОЇ ПОЛОВИНИ ХХ СТОЛІТТЯ

У статті досліджено особливості індивідуально-авторського використання фразеологічних одиниць в історичних романах Р. Іваничука, Р. Іванченко, Ю. Мушкетика, П. Загребельного. Фразеологізми проаналізовано в аспекті їхньої контекстуальної реалізації.

Ключові слова: *фразеологічна одиниця, фразеологічний вираз, стійкі сполучення слів, контекст.*

Проблема дослідження фразеологічного багатства мови в художніх текстах тісно пов’язана з такими дисциплінами, як стилістика, лінгвістика тексту та перебуває на перетині двох наук