

УДК 81'276.6.34:373.423:81-139

## НОРМАТИВНО-ПРАВОВІ ДОКУМЕНТИ ЯК ОБ'ЄКТ ЗНЯТТЯ ГРАМАТИЧНОЇ ОМОНІМІЇ ШЛЯХОМ АВТОМАТИЧНОЇ ОБРОБКИ ТЕКСТІВ

*Карнаух Г. В.*

*Український мовно-інформаційний фонд НАН України, м. Київ*

Дослідження присвячено вивченню особливостей зняття граматичної неоднозначності словоформ на матеріалі текстів нормативно-правових актів. З огляду на специфіку мови юридичних документів, створено лексико-граматичні фільтри, що сприяють зростанню кількісних показників усунення омонімії статистичними методами. Здійснено аналіз якісно-кількісних характеристик отриманих даних. Встановлено, що поява більшості омонімічних одиниць зумовлена їх різною функціональною належністю.

**Ключові слова:** нормативно-правові документи, граматична омонімія, лексико-граматичні фільтри, статистичні методи, словоформа, граматичне значення.

**Постановка проблеми.** Найновіші лінгвістичні дослідження часто охоплюють предметне поле декількох наукових дисциплін або здійснюються в рамках різних парадигм [2; 4; 6; 11–14; 18; 20; 21]. З огляду на це останнім часом чітко окреслився системно-функціональний підхід до вивчення мови, який дозволяє здійснювати всебічний аналіз усіх параметрів досліджуваного об'єкта [1; 6; 10; 11; 17; 18; 20]. Саме у межах цього підходу варто розглядати комунікативну функцію мови. При цьому детальне вивчення дієвих механізмів, правил та особливостей комунікації дозволяє розробити уніфіковані норми, моделі спілкування в певних сферах соціального життя [11; 13, с. 70, 73]. Це актуально не лише для лінгвістики, а й для кібернетики, інформатики, оскільки побудова універсального коду для адекватного сприйняття і надання потрібної інформації є важливою складовою сучасних технологій створення мов машинного програмування, пошукових систем, а також систем автоматичного опрацювання мови (переклад, автокорекція та ін.) [6; 11; 18; 20].

Протесуттєвою перешкодою при автоматичній обробці текстів є явище неоднозначності<sup>1</sup> словоформ, зокрема – граматичної омонімії [2; 3; 6; 8–10; 18]. Одним із шляхів вирішення цієї проблеми є створення лінгвістичного процесора<sup>2</sup>, принцип дії якого базується на застосуванні статистичних методів<sup>3</sup> [2–4; 8; 9]. Крім того, результати машинної обробки досліджуваного об'єкта суттєво залежать від його особли-

<sup>1</sup>Неоднозначність або ж багатозначність трактується як родове поняття, що поєднує полісемію та омонімію [7, с. 93].

<sup>2</sup>Сукупність штучних моделей природної мови, алгоритмів і програм, що описують будову та функціонування цих моделей, та технічні засоби, які реалізують цю модель [12, с. 10].

<sup>3</sup>Такі програмні засоби створено в Українському мовно-інформаційному фонді НАН України (УМІФ НАНУ). Вони дозволяють вивчати мову на всіх її рівнях та вирішувати цілу низку завдань (здійснювати автоматичне зняття омонімії, аналіз статистичних параметрів текстів, автоматичний синтаксичний аналіз, створювати частотні словники, порівнювати тексти за різними параметрами) [9].

востей: функціонального стилю, граматичних характеристик, орфографічного та пунктуаційного оформлення тексту.

Основним завданням цього дослідження є вивчення поведінки лінгвістичного процесора при знятті граматичної омонімії статистичними методами в текстах нормативно-правових документів з метою створення додаткових програмних функцій, покликаних збільшити кількість правильно ідентифікованих слів<sup>1</sup>.

Аналіз даних, отриманих внаслідок застосування зазначеної програми на матеріалі тексту законодавчого акту<sup>2</sup>, показав, що поява значної кількості омонімічних одиниць зумовлена їх різною функціональною належністю. Це пояснюється тим, що Граматичний словник української мови<sup>3</sup>, який надає інформацію про частиномовну належність і граматичне значення кожної словоформи тексту<sup>4</sup> (завдяки застосуванню алгоритму лематизації), майже повністю репрезентує лексику української мови, а також усі можливі комбінації символів у межах слів, відповідно – і всі допустимі граматичною системою словоформи, серед яких і рідковживані [9].

Тому, враховуючи специфіку мови законодавчих документів, виникає потреба видалення певних лем чи граматичних значень, тобто створення лексико-граматичних фільтрів, формування яких ґрунтується на детальному вивченні особливостей мови зазначених текстів.

І. Колеснікова і Ю. Прадід визначають такі ознаки офіційно-ділового стилю: 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 [5, с. 55].

Автори наукової праці «Мова закону», наголошуючи, що законодавчий текст не описує подій, фактів, наукових понять, почуттів і переживань людей, не характеризує ставлення до певних осіб, явищ дійсності, а владно приписує суб'єктам права певну модель поведінки, формулює вимоги, загальнообов'язкові норми [21, с. 18], виділяють такі мовні особливості нормативно-правових актів:

1. Офіційний характер, документальність мовного вираження думки законодавця.

2. Ясність і простота мови нормативно-правових актів.

3. Максимальна точність мови законодавства (досягнення найбільшої відповідності між ідеєю, думкою законодавця й утіленням цієї думки в законодавчій формулі).

3.1. Вимоги до термінології:

3.1.1. Застосування термінів із чітко окресленим змістом.

<sup>1</sup>Словоформи, які при попередній обробці отримали неправильний набір граматичних значень або залишилися неідентифікованими.

<sup>2</sup>Дослідження проводилося на матеріалі тексту Закону України «Про Кабінет Міністрів України» [Про Кабінет Міністрів України. Закон від 07.10.2010 № 2591-VI / [Верховна Рада України]. – Режим доступу: <http://zakon3.rada.gov.ua/laws/show/2591-17>].

<sup>3</sup>Веб-сервіс розподіленої програмної лексикографічної системи «Граматичний словник української мови», створений науковим колективом УМФУ НАНУ, забезпечує морфологічний аналіз словоформ – один з основних етапів процесу зняття граматичної омонімії шляхом автоматичної обробки текстів [17].

<sup>4</sup>Якщо словоформа є омонімічною, вона отримує так званий набір граматичних характеристик (граматичні параметри всіх можливих значень аналізованої одиниці) [2, с. 27–28; 8, с. 518].

3.1.2. Використання слів та висловів, зазвичай, у більш вузькому чи спеціальному значенні, порівняно з тим, яке вони мають у літературній мові.

3.1.3. Уживання слів і виразів у їх прямому і первинному значенні.

3.1.4. Відмова від багатозначних термінів.

3.1.5. Єдність термінології, що застосовується в усьому законодавстві чи у великому його розділі.

3.2. Вимоги до стилю викладу:

3.2.1. Логічна послідовність і стрункність висловлення думки законодавця, її смислова завершеність.

3.2.2. Одноманітність способів викладу одностипних формулювань (хоча з точки зору інформативності вони можуть бути зовсім різними).

4. Економність використання в текстах нормативних актів мовних засобів.

5. Логічна зв'язність і послідовність викладу (адже мові законодавства властива раціональність, однозначність змісту, вмотивованість висновків).

6. Експресивна нейтральність нормативного тексту.

7. Безособовий, неіндивідуальний характер стилю законодавчих положень (у цьому плані мова законодавства кардинально відрізняється від мови художньої літератури й науки (особливо гуманітарної), де дозволяється вираження авторського ставлення до викладеного матеріалу) [20].

Враховуючи зазначені вище особливості законодавчої мови, до фільтру можна ввести такі леми та граматичні значення: вигуки (*a* – уживається для вираження здогаду, здивування і т. ін.); частки з ремаркою *підсил.* (*i* – уживається для підкреслення, виділення значення слова, перед яким стоїть); усі леми з ремаркою *розм., діал.* (*віда* – звістка, чутка, повідомлення); рідковживані або нехарактерні для мови законодавства слова (*банка* – скляна, металева і т. ін. посудина переважно циліндричної форми; підвищення морського дна, мілина в морі; дошка, яка надає шлюпці або човну поперечної міцності й одночасно служить сидінням для веслярів), серед яких і словоформи, що є власними назвами (*Або* – назва міста у Фінляндії); кличний відмінок однини і множини для іменників чоловічого, жіночого, середнього роду (*адміністрації* – кличний відмінок, множина); особові форми наказового способу (*плати* – відшкодовувати вартість одержаного або використаного), форми 1-ої особи для дієслів (*вищим* – видавати уривчасті, різкі, пронизливі звуки). Детальний опис одиниць, які ввійшли до фільтру, подано в таблиці 1.

Застосування зазначеного програмного засобу на практиці дало такі результати: із загальної кількості словоформ (10595) у Тексті 1 – 2745 неомонімічних одиниць, а в Тексті 2<sup>1</sup> – 3821 (отже, 1076 одиниць унаслідок використання фільтру перейшли в розряд однозначних). Крім того, у Тексті 2 промарковано на 187 словоформ більше, ніж у Тексті 1 (7355 і 7186 одиниць відповідно). Таким чином, різниця в кількості правильно ідентифікованих словоформ між аналізованими текстами становить 1263 одиниці (11,92 % від загальної кількості) (див. табл. 2).

---

<sup>1</sup>У Тексті 2 при знятті граматичної неоднозначності словоформ застосовано лексико-граматичний фільтр, у Тексті 1 зазначена програмна функція не застосовувалась.

## Список унікальних словоформ, що ввійшли до лексико-граматичного фільтру

Таблиця 1

№ з/п	Унікальна словоформа	Грамматична характеристика словоформи			Примітки	
		лема	граматичне значення	код (числовий)		
1.	а	а	вигук	72,0		
		а	частка	74,0	<i>спонук.</i>	
2.	або	Аба	Іменник жіночого роду, кличний відмінок, одна	5,7	власна назва (місто в Угорщині)	
		Або	іменник середнього роду, одна	називний відмінок	13,1	власна назва (друга назва м. Турку, Фінляндія)
				родовий відмінок	13,2	
				давальный відмінок	13,3	
				знахідний відмінок	13,4	
				орудний відмінок	13,5	
				місцевий відмінок	13,6	
				кличний відмінок	13,7	
або	частка	74,0	<i>розм.</i>			
3.	але	але	вигук	72,2		
		але	частка	74,3	<i>рідко</i> (з відтінком сумніву)	
4.	банками	банка	іменник жіночого роду, орудний відмінок, множина	5,12	мілина, лавка в човні	
5.	без	без	іменник чоловічого роду, знахідний відмінок, одна	8,4	<i>діал.</i> бузок	
6.	бере	бере	іменник жіночого роду	називний відмінок, одна	5,1	назва групи столових сортів груші, плоди яких мають видовжену форму та соковитий м'якуш.
				родовий відмінок, одна	5,2	
				давальный відмінок, одна	5,3	
				знахідний відмінок, одна	5,4	
				орудний відмінок, одна	5,5	
				місцевий відмінок, одна	5,6	
				кличний відмінок, одна	5,7	
				називний відмінок, множина	5,8	
				родовий відмінок, множина	5,9	
				давальный відмінок, множина	5,10	
				знахідний відмінок, множина	5,11	
				орудний відмінок, множина	5,12	
				місцевий відмінок, множина	5,13	
				кличний відмінок, множина	5,14	
	бер	іменник чоловічого роду, кличний відмінок, одна	8,7	одиниця дози рентгенівського випромінювання		
	Бер	іменник чоловічого роду, кличний відмінок, одна	33,7	власна назва (прізвище)		
7.	веде	веда	іменник жіночого роду, кличний відмінок, одна	5,7	пам'ятка літератури (інд.)	
		вед	іменник чоловічого роду, кличний відмінок, одна	8,7	представник народності	

**НОРМАТИВНО-ПРАВОВІ ДОКУМЕНТИ ЯК ОБ'ЄКТ ЗНЯТТЯ  
ГРАМАТИЧНОЇ ОМОНІМІЇ ШЛЯХОМ АВТОМАТИЧНОЇ ОБРОБКИ ТЕКСТІВ**

8.	вищу	вищати	дієслово недоконаного виду, теперішній час, перша особа, однина	36,11		
9.	від	віда	іменник жіночого роду, родовий відмінок, множина	5,9	діал. звістка, повідомлення	
10.	всій	всіяти	дієслово доконаного виду, наказовий спосіб, друга особа, однина	35,2		
11.	для	дляти	дієслово недоконаного виду, теперішній час, третя особа, однина	36,15		
12.	до	до	іменник жіночого роду	називний відмінок, однина	13,1	нота
				родовий відмінок, однина	13,2	
				давальний відмінок, однина	13,3	
				знахідний відмінок, однина	13,4	
				орудний відмінок, однина	13,5	
				місцевий відмінок, однина	13,6	
				кличний відмінок, однина	13,7	
				називний відмінок, множина	13,8	
				родовий відмінок, множина	13,9	
				давальний відмінок, множина	13,10	
				знахідний відмінок, множина	13,11	
				орудний відмінок, множина	13,12	
місцевий відмінок, множина	13,13					
кличний відмінок, множина	13,14					
13.	за	за	прислівник	71,4	на підтримку	
		за	частка	74,5		
14.	меті	мет	іменник чоловічого роду, місцевий відмінок, однина	8,6	діал. момент	
15.	на	на	вигук	72,6		
		на	частка	74,3	розм.	
16.	освіти	освітити	дієслово доконаного виду, наказовий спосіб, друга особа, однина	35,2		
17.	переговори	переговорити	дієслово доконаного виду, наказовий спосіб, друга особа, однина	35,2		
18.	перед	перед	іменник чоловічого роду	називний відмінок, однина	8,1	
				знахідний відмінок, однина	8,4	
19.	під	під	іменник чоловічого роду	називний відмінок, однина	8,1	западина; нижня горизонтальна поверхня; місце для стіжка; горище
				знахідний відмінок, однина	8,4	
20.	підстав	підставити	дієслово доконаного виду, наказовий спосіб, друга особа, однина	35,2		
21.	після	після	прислівник	71,8		
22.	плати	платити	дієслово недоконаного виду, наказовий спосіб, друга особа, однина	36,2		

23.	постанови	поста- нови- ти	дієслово доконаного виду, наказовий спосіб, друга особа, однина	35,2		
24.	розмір	розмі- ряти	дієслово доконаного виду, наказовий спосіб, друга особа, однина	35,2		
25.	стану	стати	дієслово доконаного виду, майбутній час, перша особа, однина	35,5		
26.	так	така	іменник жіночого роду, родовий відмі- нок, множина	5,9	грошова одиниця Бангладеш	
		так	прислівник	71,2		
		так	частка	74,3		
27.	тих	тихну- ти	дієслово недоконаного виду, минулий час, чоловічий рід, однина	36,19		
28.	у	у	вигук	72,19		
29.	чи	чи	частка	74,19	<i>nit.</i>	
30.	чиї	чий	іменник чоловічо- го роду	місцевий відмінок, однина	8,6	
				називний відмінок, множина	8,8	
				знахідний відмінок, множина	8,11	
				кличний відмінок, множина	8,14	
31.	що	що	частка	74,8	<i>розм.</i>	
32.	як	як	частка	74,8	<i>експрес.-підсил.</i>	
		як	іменник чоловічого роду, істота, назив- ний відмінок, однина	7,1	велика жуйна рогата тварина	
33.	яка	як	іменник	родовий відмінок, однина	7,2	велика жуйна рогата тварина
			чоловічого роду, істота	знахідний відмінок, одни- на	7,4	
		якати	дієслово недоконаного виду, теперіш- ній час, третя особа, однина	36,15		

## Загальна характеристика Текстів 1 і 2

Таблиця 2

№		Без застосування фільтру		Із застосуванням фільтру	
			%		%
1.	Загальна кількість словоформ	10595		10595	
2.	Кількість унікальних словоформ	1872	17,69	1872	17,69
3.	Загальна кількість маркованих словоформ	7186	67,82	7355	69,42
4.	Кількість унікальних маркованих словоформ	1230	11,60	1234	11,65
5.	Кількість неомонімічних словоформ	2745	25,91	3821	36,06
6.	Кількість омонімічних словоформ	7850	74,09	6774	63,94

Для зручності роботи з текстом, в якому здійснюється зняття граматичної неоднозначності словоформ статистичними методами, запропонований лінгвістичний процесор надає можливість формувати індивідуальний лексико-граматичний фільтр для кожного окремого аналізованого об'єкта. Однак, спираючись на лінгвістичний матеріал, можна констатувати, що до фільтру, створеного в межах цього дослідження, входить група омонімічних словоформ, яка буде завжди вилучатися із загально-

го реєстру омонімів при автоматичній обробці текстів нормативно-правових актів. Тому доцільним вважаємо створення словника омонімічних словоформ, які є однозначними в межах юридичних документів.

**Висновки.** Отже, основне завдання законодавчої мови (владно приписувати суб'єктам права певну поведінку, формулювати вимоги, загальнообов'язкові норми) диктує певні правила та особливості оформлення нормативно-правових документів на всіх рівнях мови. Це у свою чергу дозволяє розробити певні програмні засоби автоматичного опрацювання текстів зазначеного виду, зокрема, створити лексико-граматичні фільтри, що сприятимуть збільшенню кількості правильно ідентифікованих неоднозначних словоформ. Оскільки результати апробації зазначеного програмного засобу демонструють динаміку зростання позитивних змін (збільшення кількості правильно ідентифікованих словоформ, зменшення – ідентифікованих неправильно), актуальним залишається поповнення списку словника омонімічних словоформ, що стають однозначними в контексті юридичної мови, для створеного фільтру, а також формування нових функцій цього типу.

### **Список літератури**

1. Афанасьєва О. С. Лексико-семантичні девіації у текстах нормативно-правових актів / О. С. Афанасьєва // Вісник Прикарпатського університету. Філологія (Мовознавство). – Випуск XXI – XXII. – Івано-Франківськ : Прикарпат. нац. ун-т ім. В. Стефаника, 2009. – С. 77–80.
2. Карнаух Г. В. Велика і мала літери як засіб розрізнення неоднозначних словоформ при автоматичній дизамбігуації / Г. В. Карнаух // Проблеми граматики і лексикології української мови. – Вип. 8. – К. : НПУ ім. М. П. Драгоманова, 2011. – С. 26–36.
3. Карнаух Г. В. Деякі аспекти зняття граматичної неоднозначності словоформ у текстах художнього стилю (на матеріалі прозових та поетичних творів) / Г. В. Карнаух // Філологічні студії. Науковий вісник Криворізького державного педагогічного університету. – Випуск 6. – Кривий Ріг : КДПУ, 2011. – С. 585–592.
4. Карнаух Г. В. До питання про зняття граматичної неоднозначності словоформ у системах автоматичної обробки текстів (на матеріалі прозового твору) / Г. В. Карнаух // Українське мовознавство. Міжвідомчий науковий збірник. – Випуск 41/1. – К. : КНУ ім. Т. Шевченка. – С. 106–110.
5. Колеснікова І. Є. Українське ділове мовлення : [навч. посібник] / І. Є. Колеснікова, Ю. Ф. Прадід ; [за ред. Ю. Ф. Прадіда]. – Сімферополь : ВД «АРИАЛ», 2009. – 210 с.
6. Корпусна лінгвістика / В. А. Широков, О. В. Бугаков, Т. О. Грязнухіна та ін. ; [під ред. В. А. Широкова]. – К. : Довіра, 2005. – 407 с.
7. Кочерган М. П. Слово і контекст (лексична сполучуваність і значення слова) : [монографія] / Михайло Петрович Кочерган. – Л. : Вид-во при Львів. держ. ун-ті видавничого об'єднання «Вища школа», 1980. – 182 с.
8. Крыгин М. Снятие грамматической омонимии в тексте с помощью статистических методов / М. Крыгин, В. Шкурко, О. Афанасьєва // Прикладна лінгвістика та лінгвістичні технології : MegaLing–2009. – К. : Довіра, 2009. – 527 с.
9. Крыгин М. Текст на естественном языке как объект статистического анализа / М. Ю. Крыгин // Біоніка інтелекту : [наук.-техн. журнал]. – 2000. – № 1 (72). – С. 75–82.
10. Любченко Т. П. Лексикографічні системи граматичного типу та їх застосування в засобах автоматичного опрацювання мови : дис. на здобуття наукового ступеня кандидата техн. наук : спец. 05.13.01 – К., 2011. – 294 с.
11. Мільченко О. С. Семантичні девіації в нормативно-правових текстах : дис. на здобуття наукового ступеня кандидата філол. наук : спец. 10.02.01 «українська мова» / О. С. Мільченко. – К., 2011. – 178 с.
12. Пиотровский Р. Г. Инженерная лингвистика и теория языка : [монография] / Раймунд Генрихович Пиотровский. – Л. : Наука. Ленингр. отд-ние 1979. – 111 с.

13. Токарська А. С. Комунікативна природа права і правова професійна діяльність / А. С. Токарська // 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 : Елінь, 2006. – С. 69–73.
14. 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 : Елінь, 2006. – С. 89–93.
15. Чулінда Л. І. Особливості тексту нормативно-правового акта / Л. І. Чулінда // Науковий вісник Національної академії внутрішніх справ України. – 2003. – № 2. – С. 82–89.
16. Чулінда Л. І. Текст нормативно-правового акта: загальні підходи до аналізу / Л. І. Чулінда // Науковий вісник Національної академії внутрішніх справ України. – 2002. – № 3. – С. 120–126.
17. Шевченко И. В. Электронный грамматический словарь украинского языка / И. В. Шевченко, А. Г. Рабулец, В. А. Широков // Труды Международной конференции «MegaLing–2005. Прикладная лингвистика в поиске новых путей», 27 июня – 2 июля 2005 г. – Меганом, 2005. – С. 124–129.
18. Шипнівська О. О. Структурно-семантичні та функціональні характеристики міжчастиномовної морфологічної омонимії сучасної української мови: дис. на здобуття наукового ступеня кандидата філол. наук : спец. 10.02.01 «українська мова» / О. О. Шипнівська. – К., 2007. – 238 с.
19. Шипнівська О. О. Функціонування міжчастиномовних морфологічних омонімів в українських текстах / О. О. Шипнівська // Мовознавство. – 2005. – № 6 (233). – С. 70–80.
20. Широков В. А. Феноменологія лексикографічних систем : [монографія] / Володимир Анатолійович Широков. – К. : Наук. думка, 2004. – 327 с.
21. Язык закона / [под. ред. А. С. Пиголкина]. – М. : Юрид. лит., 1990. – 192 с.

**Карнаух А. В.** Нормативно-правовые документы как объект снятия грамматической омонимии путем автоматической обработки текстов / А. В. Карнаух // Ученые записки Таврического национального университета имени В. И. Вернадского. Серия «Филология. Социальные коммуникации». – 2012. – Т. 25 (64), № 4, ч. 2. – С. 415–422.

Исследование посвящено изучению особенностей снятия грамматической неоднозначности словоформ на материале текстов нормативно-правовых актов. С учетом специфики языка юридических документов созданы лексико-грамматические фильтры, которые способствуют росту количественных показателей устранения омонимии статистическими методами. Проанализированы качественно-количественные характеристики полученных данных. Установлено, что появление большинства омонимичных единиц обусловлено их разной функциональной принадлежностью.

**Ключевые слова:** нормативно-правовые документы, грамматическая омонимия, лексико-грамматические фильтры, статистические методы, словоформа, грамматическое значение.

**Karnaukh G.** The legal documents as an object of elimination of grammatical of grammatical homonymy by automatic text processing / G. Karnaukh // Scientific Notes of Taurida V. I. Vernadsky National University. – Series: Philology. Social communications. – 2012. – Vol. 25 (64), No 4, part 2. – P. 415–422.

This research describes the investigation of particulars for elimination of grammatical ambiguity of word-forms on the basis of legal acts. Considering the specificity of language of judicial documents it was developed lexical and grammatical filters which increase the growth of quantitative indicators for elimination of grammatical ambiguity by means of statistical methods. The qualitative and quantitative characteristics of data obtained were analyzed. It was found that the appearance of majority of homonyms is caused by its different functional belonging.

**Keywords:** legal documents, grammatical ambiguity, lexical and grammatical filters, statistical methods, word-form, grammatical meaning.

*Поступила в редакцию 14.10.2012*