

ISSN 1392-8295



UK
Uniwersytet
Jana Kochanowskiego w Kielcach

mokslo darbai transactions

RESPECTUS PHILOLOGICUS

2016 Nr. 29 (34)

RESPECTUS PHILOLOGICUS Nr. 29 (34)

MOKSLINIS TĘSTINIS LEIDINYS

Leidžia Vilniaus universiteto Kauno humanitarinis fakultetas ir Jano Kochanovskio universiteto Humanitarinis fakultetas Kielcuose du kartus per metus (balandžio 25 d. ir spalio 25 d.).

Leidinyje lietuvių, lenkų, anglų, rusų kalbomis skelbiami originalūs moksliniai straipsniai, atitinkantys svarbiausias fakultetų humanitarinių mokslinių tyrimų kryptis: lingvistikinius tyrinėjimus, literatūros naratyvų ir kontekstų problematiką, reklamos diskurso paveikumą, vertimo teoriją ir praktiką.

CZASOPISMO NAUKOWE

Wydawcy: Uniwersytet Wileński – Wydział Humanistyczny w Kownie oraz Uniwersytet Jana Kochanowskiego – Wydział Humanistyczny w Kielcach. Ukazuje się dwa razy w roku: 25 kwietnia i 25 października.

W czasopiśmie, w językach litewskim, polskim, angielskim i rosyjskim, zamieszczone są oryginalne artykuły naukowe, których tematyka odzwierciedla najważniejsze kierunki badań prowadzonych na wydziałach humanistycznych współpracujących uczelni: badania językoznawcze i literaturoznawcze wraz z ich kontekstami, badania nad reklamą oraz nad teorią i praktyką przekładu.

ONGOING ACADEMIC PUBLICATION

Published twice a year (April 25, October 25) by Vilnius University Kaunas Faculty of Humanities and The Jan Kochanowski University Faculty of Humanities in Kielce.

The periodical publishes original scholarly articles written in English, Lithuanian, Polish and Russian that correspond to the main humanitarian research fields of the faculties: linguistic researches, issues of literary narratives and contexts, influence of advertising discourse, theory and practice of translation.

DUOMENŲ BAZĖS / BAZY DANYCH / ABSTRACTING AND INDEXING

Elektronische Zeitschriftenbibliothek (2002)	Frei zugängliche E-Journals Universitätsbibliothek Regensburg
Arianta (2002)	Naukowe i branżowe polskie czasopisma elektroniczne
Balcan Rusistics (2004)	Biblioteka Uniwersytetu Śląskiego
C.E.E.O.L. (2005)	Russian Language, Literature and Cultural Studies Central and Eastern European Online Library DFG Nationallizenzen
EBSCO (2006)	Humanities International Complete Humanities Source Current Abstracts Humanities International Index TOC Premier
MLA (2007)	Modern Language Association International Bibliography
Index Copernicus (2008)	Index Copernicus International Journal Master List
Lituanistika (2011)	The Database of the Humanities and Social Sciences in Lithuania
PNB (2011)	Polska Bibliografia Naukowa
MIAR (2011)	Information Matrix for the Analysis of Journals
Linguistic Bibliography Online (2012)	BrillOnline Bibliographies, Netherlands
Ulrichs (2013)	Ulrich's Periodicals

PATIKRA / WERYFIKACJA / VERIFICATION



REDAKCIJOS ADRESAS / ADRES REDAKCJI / EDITORIAL OFFICE

Vilniaus universitetas	Vilnius University	Uniwersytet Jana Kochanowskiego
Kauno humanitarinis fakultetas	Kaunas Faculty of Humanities	Wydział Humanistyczny
Muitinės g. 8	Muitines 8	ul. Świętokrzyska 21D
LT-44280 Kaunas	LT-44280 Kaunas	25-406 Kielce
Lietuva	Lithuania	Polska

Straipsnius ir kitą mokslinę medžiagą prašome siųsti į žurnalo redakciją Kaune, taip pat el. paštu.

Articles and other scientific material send to the Editorial Office in Kaunas, as well as by e-mail.
Artykuły i inne materiały przeznaczone do publikacji prosimy nadsyłać na adres redakcji w Kownie, także pocztą elektroniczną.

El. paštas / E-mail respectus.philologicus@khf.vu.lt **Tel.** +370 37 750 536
Interneto svetainė / Strona internetowa / Homepage <http://www.rephi.khf.vu.lt>

REDAKTORIAI / REDAKCJA / EDITORS

Gabija Bankauskaitė-Sereikienė (lietuvių kalba / język litewski / Lithuanian language)
Beata Piasecka (lenkų kalba / język polski / Polish language)
Brygita Brasiene (anglų kalba / język angielski / English language)
Alla Diomidova (rusų kalba / język rosyjski / Russian language)

Print ISSN 1392-8295, Online ISSN 2335-2388

© Vilniaus universiteto Kauno humanitarinis fakultetas, 2016

© Uniwersytet Jana Kochanowskiego – Wydział Humanistyczny w Kielcach, 2016

RESPECTUS PHILOLOGICUS Nr. 29 (34)

REDAKCINĖ KOLEGIJA – KOLEGIUM REDAKCYJNE EDITORIAL BOARD

Vyriausioji redaktorė – redaktor naczelny – editor-in-chief

**Gabija BANKAUSKAITĖ-
SEREIKIENĖ**
gabija.sereikiene@khf.vu.lt

Vilniaus universiteto profesorė, dr., Lietuva
Profesor Uniwersytetu Wileńskiego, dr, Litwa
Professor, Vilnius University, Lithuania

Vyriausiosios redaktorės pavaduotojas – zastępca redaktora naczelnego – deputy editor-in-chief

Marek RUSZKOWSKI
gustaw1960@op.pl

Jano Kochanovskio universiteto Kielcuose profesorius, habil. dr., Lenkija
Profesor Uniwersytetu Jana Kochanowskiego w Kielcach, dr hab., Polska
Professor, Jan Kochanowski University in Kielce, Poland

Vyriausiosios redaktorės pavaduotoja – zastępca redaktora naczelnego – deputy editor-in-chief

Loreta ULVYDIENĖ
loreta.ulvydiene@khf.vu.lt

Vilniaus universiteto profesorė, dr., Lietuva
Profesor Uniwersytetu Wileńskiego, dr, Litwa
Professor, Vilnius University, Lithuania

Daiva ALIŪKAITĖ
daiva.aliukaite@khf.vu.lt

Vilniaus universiteto docentė, dr., Lietuva
Docent Uniwersytetu Wileńskiego, dr, Litwa
Associate Professor, Vilnius University, Lithuania

Danutė BALŠAITYTĖ
dbalsaityte@gmail.com

Vilniaus universiteto profesorė, dr., Lietuva
Profesor Uniwersytetu Wileńskiego, dr, Litwa
Professor, Vilnius University, Lithuania

Wiesław CABAN
caban@ujk.kielce.pl

Jano Kochanovskio universiteto Kielcuose profesorius, habil. dr., Lenkija
Profesor Uniwersytetu Jana Kochanowskiego w Kielcach, dr hab., Polska
Professor, Jan Kochanowski University in Kielce, Poland

Renato CORSETTI
renato.corsetti@gmail.com

Romos universiteto „La Sapienza“ profesorius, habil. dr., Italija
Profesor Uniwersytetu Rzymського „La Sapienza“, dr hab., Włochy
Professor, Sapienza University of Rome, Italy

Anatolij CHUDINOV
ap_chudinov@mail.ru

Uralo valstybinio pedagoginio universiteto profesorius, dr., Rusija
Profesor Uralskiego Państwowego Uniwersytetu Pedagogicznego, dr, Rosja
Professor, Ural State Pedagogical University, Russia

Janusz DETKA
januszdetka1@wp.pl

Jano Kochanovskio universiteto Kielcuose profesorius, habil. dr., Lenkija
Profesor Uniwersytetu Jana Kochanowskiego w Kielcach, dr hab., Polska
Professor, Jan Kochanowski University in Kielce, Poland

Margarita GAVRILINA
gavril@inbox.lv

Latvijos universiteto profesorė, dr., Latvija
Profesor Uniwersytetu Łotewskiego w Rydze, dr, Łotwa
Professor, University of Latvia, Latvia

Jonė GRIGALIŪNIENĖ
jone.grigaliuniene@gmail.com

Vilniaus universiteto profesorė, dr., Lietuva
Profesor Uniwersytetu Wileńskiego, dr, Litwa
Professor, Vilnius University, Lithuania

Laima KALĖDIENĖ
kaledienei@gmail.com

Mykolo Romerio universiteto profesorė, dr., Lietuva
Profesor Uniwersytetu Michała Rómera, dr, Litwa
Professor, Mykolas Romeris University, Lithuania

Asta KAZLAUSKIENĖ
a.kazlauskiene@hmf.vdu.lt

Vytauto Didžiojo universiteto profesorė, dr., Lietuva
Profesor Uniwersytetu Witolda Wielkiego, dr, Litwa
Professor (04H), Vytautas Magnus University, Lithuania

Aleksandras KRASNOVAS
aleksandras.krasnovas@khf.vu.lt

Vilniaus universiteto profesorius, dr., Lietuva
Profesor Uniwersytetu Wileńskiego, dr, Litwa
Professor, Vilnius University, Lithuania

- Jadvyga KRŪMINIENĖ**
jadvyga.kruminiene@khf.vu.lt
Vilniaus universiteto profesorė, dr., Lietuva
Profesor Uniwersytetu Wileńskiego, dr, Litwa
Professor, Vilnius University, Lithuania
- Kazimierz LUCIŃSKI**
ifs@ujk.edu.pl
Jano Kochanovskio universiteto Kielcuose profesorius, habil. dr., Lenkija
Profesor Uniwersytetu Jana Kochanowskiego w Kielcach, dr hab., Polska
Professor, Jan Kochanowski University in Kielce, Poland
- Lidia MICHALSKA-BRACHA**
lidia.bracha@ujk.edu.pl
Jano Kochanovskio universiteto Kielcuose profesorė, habil. dr., Lenkija
Profesor Uniwersytetu Jana Kochanowskiego w Kielcach, dr hab., Polska
Professor, Jan Kochanowski University in Kielce, Poland
- Gerald E. MIKKELSON**
gemikk@ku.edu
Kanzaso universiteto profesorius, dr., JAV
Profesor Uniwersytetu w Kansas, dr, USA
Professor, The University of Kansas, USA
- Olegas POLIAKOVAS**
oleg.poljakov@ffv.vu.lt
Vilniaus universiteto profesorius, dr., Lietuva
Profesor Uniwersytetu Wileńskiego, dr, Litwa
Professor, Vilnius University, Lithuania
- Adriana SERBAN**
adriana.serban@univ-montp3.fr
Polio Valéry universiteto Montpellier III docentė, dr., Prancūzija
Docent Uniwersytetu Paul Valéry, Montpellier III, dr, Francja
Senior lecturer, Paul Valéry University, Montpellier III, France
- Yuri STULOV**
yustulov@mail.ru
Minsko valstybinio lingvistikos universiteto profesorius, dr., Baltarusija
Profesor Państwowego Uniwersytetu Lingwistycznego w Mińsku, dr, Białoruś
Professor, Minsk State Linguistic University, Belarus
- Leona TOKER**
toker@mscc.huji.ac.il
Jeruzalės Hebrajų universiteto profesorė, habil. dr., Izraelis
Profesor Uniwersytetu Hebrajskiego w Jerozolimie, dr hab., Izrael
Professor, The Hebrew University of Jerusalem, Izrael

TURINYS / SPIS TREŚCI

I. DISKURSO IR NARATYVO TYRIMAI / BADANIA NAD DYSKURSEM I NARRACJĄ

Marek Ruszkowski (Lenkija / Polska). Fundamentalizm językowy.....	9
Ina Pukelytė (Lietuva / Litwa). Tarp dramos ir postdramos.....	18
Eglė Gabrėnaitė (Lietuva / Litwa). Politinio komentaro retorika: žanro transformacijos	25
Sylwia Konarska-Zimnicka (Lenkija / Polska). Socjologia a warsztat naukowy historyka. Przykład etnometodologii.....	33
Iwona Filipczak (Lenkija / Polska). Marginalization of South Asians Based on Race and Skin Color in Bharati Mukherjee's <i>Jasmine</i> and Chitra B. Divakaruni's <i>The Mistress of Spices</i>	46
Skirmantė Biržietienė (Lietuva / Litwa). Antikinė „lengvoji poezija“: epigramos ypatumai.....	57
Ewa Anna Piasta (Lenkija / Polska). Motywy religijne w powieści <i>Gody magdeburskie</i> Gertrudy von Le Fort.....	65

II. KALBA. VISUOMENĖ. KULTŪRA / JĘZYK. SPOŁECZEŃSTWO. KULTURA

Oleksandr Kapranov (Norvegija / Norwegia). Conceptual Metaphors in British Foreign Secretary's Twitter Discourse Involving Ukraine.....	75
Татьяна Бобкова (Ukraina / Ukraina). Классификация коллокаций: основные подходы и критерии	87
Martyna Król (Lenkija / Polska). Homonimia wśród nominatów dotyczących czasu i przestrzeni: polsko-rosyjskie studium kontrastywne.....	99
Ольга Солодушко (Lenkija / Polska). Семантическая характеристика зоонимов, мотивированных пестрой окраской и узорами на шерсти животного (на примере кличек коров в польском и украинском языках)	107
Elżbieta Michow (Lenkija / Polska). Polskie i bułgarskie frazeologizmy somatyczne motywowane przez nazwy części ciała służących do mówienia.....	121
Olga V. Domnich (Ukraina / Ukraina). Etymologic Features and Peculiarities of the Process of Indigenous Vocabulary Borrowing in American English.....	130
Erika Merkytė-Švarcienė (Lietuva / Litwa). Tarmių mentalinio žemėlapiu (ne)tiksłumo veiksniai: tarmės vartojimas vs. nevartojimas.....	139
Hona Mickienė, Rita Baranauskienė (Lietuva / Litwa). Veliunos apylinkių pravardžių semantikos ypatumai: pirminės pravardės.....	155
Lina Bikilienė (Lietuva / Litwa). Evaluative Language in Lithuanian Students' English Writing. A Study of <i>Important</i> and <i>Interesting</i>	177

**III MOKSLINIO GYVENIMO KRONIKA /
KRONIKA ŻYCIA NAUKOWEGO**

Konferencijų apžvalgos / Konferencje

Gabija Bankauskaitė-Sereikienė (Lietuva / Litwa). Žanras: kanonas ar transformacija?	188
Jurgita Astrauskienė, Danguolė Satkauskaitė (Lietuva / Litwa). Tarp eilučių: kalbos ir medijų sankirta.....	191

Knygų recenzijos / Recenzje książek

Алла Дюмидова (Lietuva / Litwa). Просто о сложном: новое пособие по фонетике русского языка.....	194
Aleksandras Krasnovas (Lietuva / Litwa). Apie lenkų tremtinius Sibire – išsamiai ir objektyviai.....	196

Anonsai / Zapowiedzi	199
-----------------------------------	-----

IV. REIKALAVIMAI STRAIPSNIAMS /

ZASADY OPRACOWYWANIA PUBLIKACJI	200
--	-----

CONTENTS

I. RESEARCH OF DISCOURSE AND NARRATIVE

Marek Ruszkowski (Poland). Linguistic Fundamentalism.....	9
Ina Pukelytė (Lithuania). Between Drama and Postdrama	18
Eglė Gabrėnaitė (Lithuania). The Rhetoric of Political Commentary: Genre Transformations	25
Sylvia Konarska-Zimnicka (Poland). Sociology and Methodology of Historical Research. Ethnomethodology.....	33
Iwona Filipczak (Poland). Marginalization of South Asians Based on Race and Skin Color in Bharati Mukherjee's <i>Jasmine</i> and Chitra B. Divakaruni's <i>The Mistress of Spices</i>	46
Skirmantė Biržietienė (Lithuania). Ancient "Simple Poetry": Peculiarities of the Epigram	57
Ewa Anna Piasta (Poland). Religious Motifs in the Novel <i>The Magdeburg Wedding</i> by Gertrud von le Fort	65

II. LANGUAGE. SOCIETY. CULTURE

Oleksandr Kapranov (Norway). Conceptual Metaphors in British Foreign Secretary's Twitter Discourse Involving Ukraine	75
Татьяна Бобкова (Ukraine). Classification of Collocation: The Main Approaches and Criteria	87
Martyna Król (Poland). Homonymy among the Nominates Concerning Time and Space: Polish-Russian Contrastive Study	99
Ольга Солодушко (Poland). Semantyc Analysis of Zoonyms Motivated by the Coloured Coat (Based on the Cow Names in Polish and Ukrainian).....	107
Elżbieta Michow (Poland). Polish and Bulgarian Somatic Phrases Justified with the Names of the Parts of the Body Designed for Speaking	121
Olga V. Domnich (Ukraine). Etymologic Features and Peculiarities of the Process of Indigenous Vocabulary Borrowing in American English.....	130
Erika Merkytė-Švarcienė (Lithuania). Un(accuracy) Factors of the Mental Map of Dialects: Using vs. Non-using a Dialect	139
Ilona Mickienė, Rita Baranauskienė (Lithuania). Peculiarities of Nickname Semantics in Veliuona Region: Primary Nicknames	155
Lina Bikelienė (Lithuania). Evaluative Language in Lithuanian Students' English Writing. <i>A Study of Important and Interesting</i>	177

III. SCIENTIFIC LIFE CHRONICLE

Conference review

Gabija Bankauskaitė-Sereikienė (Lithuania). Genre: Canon or Transformation? 188

Jurgita Astrauskienė, Danguolė Satkauskaitė (Lithuania). Between the Lines:
The Intersection of Speech and Media 191

Book reviews

Алла Дюмидова (Lithuania). In Simple Words about Complicated: The New Guide
to the Russian Phonetics 194

Aleksandras Krasnovas (Lithuania). The Polish Exiles in Siberia: In Detail and
Objectively..... 196

Announcements 199

IV. REQUIREMENTS FOR PUBLICATION 200

Татьяна Бобкова

Киевский национальный лингвистический университет

Кафедра общего и сравнительного языкознания

и новогреческой филологии

вул. Велика Васильківська, 73, Київ, Україна

Тел. +38 066 848 01 06

E-mail: tatva93@gmail.com

Область научных интересов автора: прикладная лингвистика,

корпусная лингвистика, корпусная лексикография, синтаксис

КЛАССИФИКАЦИЯ КОЛЛОКАЦИЙ: ОСНОВНЫЕ ПОДХОДЫ И КРИТЕРИИ

В статье анализируются системо-ориентированный и текст-ориентированный подходы к классификации коллокаций, различающиеся критериями идентификации и последовательностью применяемых процедур. Предложенная методика идентификации коллокаций позволяет на основе статистической обработки и лемматизации автоматически извлекать двусловные коллокации из корпуса украинских текстов. Основные проблемы классификации извлеченных коллокаций представляют трехсловные устойчивые сочетания, ошибочно идентифицируемые как двусловные, неразрешенная омонимия и совместное употребление слов в тексте, не связанных синтаксически. Применение корпусно-ориентированного подхода в данном исследовании обеспечило выявление непосредственно наблюдаемых в подкорпусе законодательных текстов лексических, грамматических и предикативных коллокаций, релевантных для современного украинского юридического дискурса. Использование большего корпуса текстов разных функциональных стилей и лингвистических фильтров обеспечит повышение эффективности результатов идентификации и классификации двусловных коллокаций.

КЛЮЧЕВЫЕ СЛОВА: коллокация, корпус текстов, классификация, идентификация, извлечение коллокаций, критерии коллокаций.

Интерес к изучению коллокаций объясняется высокой востребованностью этой информации для решения различных прикладных задач: автоматического анализа текста, машинного перевода, составления словарей, тезаурусов, построения семантических сетей. На современном этапе создание прикладных систем в области автоматического анализа текста основывается на обработке значительных объемов естественно-языкового материала, предоставляемого лингвистическими корпусами. Необходимость основательного изучения коллокаций объясняется также ориентированностью большинства прикладных систем анализа текста на традиционные лексемы.

Эмпирические данные исследований корпусов свидетельствуют о том, что 80 % текстов составляют не изолированные лексические единицы, а регулярно воспроизводимые структуры (Marcinkevičienė 2010: 62), являющиеся «важным источником информации как о языке, так и об окружающем нас мире» (Evert 2005: 15). Поэтому на современном этапе в создании прикладных лингвистических систем наблюдается перемещение акцента с отдельного слова на устойчивые сочетания. Указанные структуры не создаются участником коммуникации непосредственно в процессе речепроизводства, а извлекаются из памяти целиком, как готовые образования, подобные простым словам. В связи с этим особый интерес представляет описание и классификация высокочастотных устойчивых синтагматических последовательностей, для определения которых в современной лингвистике используется термин «коллокация» (Bartsch 2004; Evert 2005; Fontenelle 1994; Marcinkevičienė 2010; Seretan 2011; Sinclair 1991; Smadja 1990; Teubert 2009; Tognini-Bonelli 2000; Большакова *и др.* 2011; Гладка 2013; Дарчук 2013; Зацеркляний *и др.* 2013; Хохлова 2010; Шкурко 2012; Ягунова *и др.* 2011).

В самом общем понимании изучение явления предполагает выявление и описание его сущности, определение критериев идентификации и статуса среди подобных, установление набора единиц и классификацию их видов. Впервые в научный обиход термин «коллокация» был введен Дж. Р. Ферзом в рамках контекстуальной теории значения Лондонской школы структурной лингвистики (Seretan 2011: 9). Основной заслугой представителей британского контекстуализма (J. R. Firth, A. McIntoshas, M. A. K. Halliday) следует признать осознание необходимости эмпирического изучения языка непосредственно в его функционировании. В 1930 гг. понятие коллокации интерпретировалось как «привычные» словосочетания (Evert 2005: 15) или установление обычных, привычных позиций данного слова (Ягунова *и др.* 2011). И хотя в рамках британского контекстуализма не было сформулировано четкое определение термина «коллокация», концептуальные принципы указанного направления послужили основой для формирования в современной лингвистике контекстно-ориентированного и корпусного подходов к изучению феномена.

Несмотря на относительно недавнее вхождение термина «коллокация» в широкий научный обиход (Sinclair 1991), история данного вопроса насчитывает почти сто лет. Отдельные аспекты классификации коллокаций как устойчивых сочетаний и проявления синтагматических отношений в языке являлись объектом исследования лингвистов уже в первой половине XX в. (Ch. Bally, J. R. Firth, W. Porzig, L. Tesnière, В. В. Виноградов). В 1960-е годы с началом формирования парадигмы корпусной лингвистики концепция коллокации выходит на первый план прикладных исследований, а описание данных видов лексических структур становится основной целью разработки лингвистических корпусов текстов (Teubert 2009: 137–138). Об этом свидетельствует первая фундаментальная работа в области корпусного анализа – *English Collocation Studies* (J. McN. Sinclair, S. Jones, R. Daley), написанная в середине 1960-ых и опубликованная только в 2005 г. (in: Teubert 2009).

Следует отметить, что в славянской лингвистической традиции термин «коллокация» до этого времени еще окончательно не закрепился, поэтому в

современной литературе часто наблюдается употребление нескольких конкурентных наименований для обозначения данного понятия:

- устойчивое сочетание, клишированное сочетание, стабильный контекст (Левицкий 2006: 205),
- несвободное сочетание (Большакова *и др.* 2011: 25; Дарчук 2013: 120),
- фразеологизм, фразема, лексико-функциональное выражение (Иорданская 2007: 239),
- неоднословная номинация, сложная номинация, многословная конструкция, неоднословная конструкция (Большакова *и др.* 2011: 26–27),

С одной стороны, указанная нестабильность терминологии объясняется сосуществованием в современной лингвистике давно сформировавшихся традиций исследования коллокаций и новой парадигмы корпусной лингвистики (Ягунова *и др.* 2011). Именно исследования в области корпусной лингвистики и автоматического анализа текста способствовали распространению термина «коллокация» в такие отрасли языкознания, как лексикология, дериватология, синтаксис, психолингвистика, лексикография, теория перевода и лингвистика текста (Marcinkevičienė 2010: 77). Однако, в определенной степени распространенность данного термина определяется значениями, вкладываемыми в понятие коллокации разными лингвистами. Фактически использование самого термина и обозначаемого им общего понятия исследователями различных областей привело к серьезной терминологической путанице, связанной с концепцией коллокации на современном этапе (Evert 2005: 15).

Несомненно, термин «коллокация» является многомерным, поскольку призван отобразить комплекс лексических, грамматических и статистических признаков. Дискуссионный характер точного определения коллокации способствовал аспектуализации рассматриваемого понятия и выделения широкого, лингвистически мотивированного значения и узкого, отражающего статистический аспект изучения феномена (Fontenelle 1994: 47; Evert 2005: 16; Marcinkevičienė 2010: 40; Seretan 2011: 26). На современном этапе в широком смысле под коллокацией понимается сочетание двух слов или привычное совместное употребление определенных лексем (Marcinkevičienė 2010: 40; Sinclair 1991: 115–116). В узком значении коллокация – это слова, фиксируемые в тексте чаще рядом, чем согласно случайной вероятности отдельно (Sinclair 1991: 115–116). Таким образом, с другой стороны, описанная выше терминологическая нечеткость обусловлена многоаспектностью изучаемого феномена (Durrant *etc.* 2010: 133; Marcinkevičienė 2010: 89; Хохлова 2010: 4) и в связи с этим – неопределенным статусом коллокаций среди словосочетаний.

Достоверное описание коллокаций требует учета тройкой природы явления – лексической, грамматической и статистической – и применения новых, нетрадиционных для лингвистики методов анализа (Tognini-Bonelli 2000: 206). В современном языкознании вопрос о критериях классификации и статусе данных устойчивых сочетаний является дискуссионным. Так, в традиционных лингвистических исследованиях по степени фразеологизации, или «жесткости связи компонентов» коллокации класси-

фицируются как устойчивые сочетания (Дарчук 2013: 120; Иорданская 2007: 239) или объединяются в один класс со свободными словосочетаниями (Гладка 2013: 18).

Однако, согласно статистическим (Левицкий 2006: 206; Хохлова 2010: 4) и корпусным данным (Fontenelle 1994: 47; Evert 2005: 16; Большакова *и др.* 2011: 25; Хохлова 2010: 4; Ягунова *и др.* 2011), коллокации характеризуются промежуточным статусом устойчивых несвободных сочетаний, и рассматриваются как определенный континуум – шкала без четких границ между свободными и фразеологизированными сочетаниями (Fontenelle 1994: 48; Хохлова 2010: 10). Тройкая природа, регулярность и высокая частотность появления в текстах разных функциональных стилей требуют обоснования теоретических принципов исследований коллокаций, широкий спектр направлений которых включает:

- 1) определение сущности, критериев идентификации и статуса коллокаций среди устойчивых сочетаний (S. Bartsch, M. Benson, F. Hausmann, R. Marcinkevičienė, J. McH. Sinclair, W. Teubert, B. V. Левицкий);
- 2) разработку комплексной методики выделения в тексте и классификации статистически устойчивых сочетаний (S. Evert, R. Marcinkevičienė, V. Seretan, J. Sinclair, F. Smadja, B. V. Шкурко, E. V. Ягунова);
- 3) составление специализированных словарей коллокаций (M. Benson, F. J. Hausmann, R. Marcinkevičienė, J. McH. Sinclair).

Среди используемых для классификации коллокаций критериев выделяют «необходимые условия» и «континуумы» (Handl 2008: 50). К необходимым относят идентифицирующие критерии коллокации, предполагающие наличие, по крайней мере, двух слов, совместно употребляемых в одном контексте (Handl 2008: 51). Критерии-условия применяются при описании коллокаций на основе установленных классов, в частности для морфолого-синтаксической классификации (Benson *etc.* 2009: xix–xxx; Hausmann 1984). Критерии-континуумы являются существенными не столько для определения классификационных признаков коллокаций, сколько для степени их выраженности (Handl 2008: 50; Nation 2001: 332). Согласно S. Handl (Handl 2008: 50) к данным критериям следует отнести основные признаки коллокации: семантическую транспарентность, композиционность, сочетаемость, частоту и предсказуемость. Более разветвленная система степенных критериев в зависимости от частоты употребления и лексикализации коллокаций представлена в работе P. Nation (Nation 2001: 332), но, в целом, критерии-континуумы служат основанием для градуированной классификации коллокаций на материале текстов.

Как показывает анализ современной литературы, указанным типам критериев соответствуют системо-ориентированный и тексто-ориентированный подходы к описанию коллокаций, обнаруживающие различия в употреблении терминов и в последовательности применяемых методов и процедур исследования (Bartsch 2004: 27–64; Evert 2005: 15; Wermter 2008: 53–54). При этом, системо-ориентированный подход, включающий фразеологический, семантический, структурный, подход теории “Смысл↔Текст”, лексико-грамматический, синтаксический и конвенционный подход, предполагает анализ коллокации как лексической единицы в соотношении с другими языковыми единицами в аспекте структуры и семантики. Соответственно,

тексто-ориентированный подход охватывает статистический, лексико-композиционный, контекстно-ориентированный и корпусно-ориентированный подходы и выявляет закономерности поведения коллокации в тексте: насколько частотной является анализируемая лексическая единица, в каких позициях относительно других единиц встречается, какие сочетания являются характерными для изучаемого домена.

В соответствии с системо-ориентированным подходом исходными положениями классификации коллокаций являются установленные в ручном или автоматическом режиме ограничения по формальным и качественным признакам: функциональным моделям (Иорданская 2007: 239), комбинаторным характеристикам (Nation 2001: 332; Хохлова 2010: 16–17), лексическим вариантам (Ягунова *и др.* 2011) и морфолого-синтаксическим моделям (Seretan 2011: 27; Большакова 2011: 22; Гладка 2013: 16; Хохлова 2010: 3; Шкурко 2012: 31). Как правило, в традиционных исследованиях больше внимания уделяется описанию лексических коллокаций (Hausmann 1984: 399; Newmark 1988: 114–115), в то время как вопрос выделения грамматических коллокаций и их классов остается дискуссионным. Независимо от исследуемого материала (словаря, текста, корпуса) применение системо-ориентированного подхода позволяет классифицировать коллокации на основе системных – формальных, структурных и семантических признаков. Однако, как показывают результаты анализа аутентичных текстов, многообразие идентифицированных коллокаций не всегда вписывается в границы морфолого-синтаксических классификаций: учет реального порядка слов в украинском предложении почти вдвое увеличивает количество моделей сочетаемости (Дарчук 2013: 177–180).

Именно поэтому в современных корпусных исследованиях прибегают к градуированной или иерархической классификации коллокаций на основе функциональных признаков. В соответствии с тексто-ориентированным подходом исходными для классификации коллокаций становятся критерии частоты и релевантности коллокаций в определенном дискурсе (Nation 2001: 332; Seljan *etc.* 2012; Smadja *etc.* 1990; Ягунова *и др.* 2011). Собственно, благодаря накоплению статистических данных распознавания коллокаций в больших массивах текстов доказана неоднородность исследуемого феномена (Bartsch 2004: 40; Tognini-Bonelli 2004: 215), не поддающаяся систематизации и описанию традиционными лингвистическими методами. В частности, в детальном корпусном анализе нуждаются частотные грамматические реализации, используемые в качестве целостных единиц (Smadja *etc.* 1990: 252; Tognini-Bonelli 2000: 215). Под влиянием корпусных исследований традиционные морфолого-синтаксические классификации дополняются классом грамматических коллокаций (Benson *etc.* 2009).

Первой попыткой систематизации коллокаций на основе частоты следует считать классификацию коллокаций нисходящего и восходящего типа по J. Sinclair (Sinclair 1991: 115–116). Однако, применение указанной классификации для систематизации корпусных данных сопряжено с определенными трудностями вследствие отсутствия общепринятых стандартов статистической значимости, низких показателей частоты типичных коллокаций и характерной для определенного текста случайной сочетаемости лексических единиц. Все это приводит к нарушению границ установ-

ленных классов и выделению буферной зоны между коллокациями нисходящего и восходящего типа (Sinclair 1991: 115–116), хотя, в целом, результаты корпусного анализа способствуют разработке классификации на основе функциональных-статистических и сочетаемостных характеристик.

Безусловные преимущества применения корпусного подхода составляют: 1) большее покрытие текста (Seljan *etc.* 2012: 152; Smadja *etc.* 1990: 252), 2) фиксация вариативности формальных и семантических признаков коллокаций (Seljan *etc.* 2012: 152; Smadja *etc.* 1990: 252; Хохлова 2010: 4), 3) определение нерегулярных, заранее не установленных синтаксических конструкций (Bartsch 2004: 40; Fontenelle 1994: 40; Smadja *etc.* 1990: 2004, Seljan *etc.* 2012: 152) и 4) универсальность корпусной методики для классификации общезыковых и релевантных для определенного домена коллокаций (Seljan *etc.* 2012: 151; Ягунова *и др.* 2011). Однако, на практике использование тексто-ориентированного подхода ограничивается функциональными возможностями корпус-менеджера, в частности заданным списком морфолого-синтаксических моделей и необходимостью ручного редактирования результатов (Seljan *etc.* 2012: 152). Кроме того, разработанные для двусловных сочетаний статистические меры ассоциаций (MI, score, t-score и log-likelihood) не способны охватить все многообразие представленных в корпусе коллокаций, дифференцировать ядро и коллокат (Dugant *etc.* 2010: 33) и различные типы коллокаций (Ягунова *и др.* 2011). Именно поэтому для классификации украинских устойчивых сочетаний целесообразным представляется использование пакета программ корпуса текстов и специализированного программного обеспечения.

Цель данного исследования – выделить классы частотных коллокаций украинского юридического дискурса на материале подкорпуса законодательных текстов, являющегося составной частью Корпуса текстов украинского языка (Бобкова 2015: 22). В основу планирования подкорпуса законодательных текстов положены ключевые требования, предъявляемые к корпусным объектам: репрезентативность, сбалансированность лингвистического материала, ограниченность объема (1 млн. слов), стандартность и исследовательское предназначение (Бобкова 2015: 22–23). Выбор официальных документов для классификации коллокаций обусловлен максимальной насыщенностью в анализируемых текстах различных видов устойчивых сочетаний (Marcinkevičienė 2010: 61–62), что позволит использовать полученные списки в качестве основы для составления общего словаря украинских коллокаций.

Автоматическое формирование списка коллокаций украинского юридического дискурса опирается на подход Ф. Смаджа (Smadja *etc.* 1990), предполагающий анализ статистических параметров как исходных данных для определения синтаксических и семантических характеристик устойчивых сочетаний. Предложенный Ф. Смаджа подход к идентификации коллокаций существенно отличается поэтапным применением статистического анализа, лингвистических фильтров (позиционного, синтаксического, морфологического) и итоговым постредактированием экспертом-лексикографом в ручном режиме (Smadja *etc.* 1990: 253). Так как автоматическое извлечение коллокаций в данном исследовании ограничено возможностями

поискового аппарата Корпуса текстов украинского языка, то экстракция устойчивых сочетаний на основе частоты их употребления и расчет вероятностных характеристик реализуется с помощью процедуры запросов на языке SQL (автор программного обеспечения – В.М. Сорокин).

Использование статистического подхода для классификации коллокаций позволяет игнорировать ограничения относительно составляющих компонентов устойчивых сочетаний. В соответствии с корпусно-ориентированным подходом в данном исследовании рассматриваются как лексические, так и грамматические коллокации. При этом, основной моделью анализа является двухсловная коллокация, как наиболее частотная в текстах разных функциональных стилей (Перебийніс *и др.* 2008: 446). В связи с этим устойчивые сочетания с сочинительной связью не принимаются во внимание, поскольку указанные пары слов представляют собой часть многословного сочетания, состоящего, по крайней мере, из двух компонентов, соединенных сочинительным союзом (...*та міра, ...і контейнер*). Для осуществления классификации коллокаций, идентифицированных в корпусе текстов, прежде всего, необходимо отсортировать пары по типам связей между компонентами: 1) сочинительная – подчинительная, 2) предикативная – непередикативная и 3) лексическая – грамматическая. Благодаря возможности задать списком сочинительные союзы, данные устойчивые сочетания автоматически исключаются из списка потенциальных коллокаций. Устойчивые сочетания с подчинительной связью рассматриваются с точки зрения формальной структуры как свободные словосочетания (Дарчук 2013: 177–180).

Таким образом, в данном исследовании основными критериями идентификации коллокаций в подкорпусе законодательных текстов является: 1) наличие двух компонентов устойчивого сочетания; 2) установленный статистический порог частоты их совместного употребления – не меньше двух раз на 1 млн. слов; 3) контактность компонентов; 4) наличие подчинительной связи между компонентами сочетания. Разработанное программное обеспечение для извлечения и последующей классификации коллокаций протестировано на подкорпусе законодательных текстов. В результате обработки текстов подкорпуса и лемматизации компонентов извлеченных пар составлен предварительный список, включающий 64.361 потенциальную коллокацию. Для последующей классификации была отобрана тысяча самых частотных устойчивых сочетаний, точнее – 1011 пар слов с учетом одинакового показателя частоты употребления последних в списке 20 сочетаний: 59 раз на 1 млн. словоупотреблений.

Результаты анализа автоматически извлеченных потенциальных коллокаций показали, что при программной обработке украинских текстов существенные проблемы представляют: а) многословные устойчивые сочетания, два компонента которых ошибочно идентифицируется как двухсловная лексическая или грамматическая коллокация (*банк України, орган державний, в іноземний*), б) неразрешенная грамматическая (*вона територія, вони використання*) и лексико-грамматическая омонимия (*рад Україна, до скласти*), в) соответствующее установленному статистическому порогу совместное употребление в тексте двух

слов, не связанных между собой подчинительной связью (*Україна від, громадянин яка*). При этом максимальный процент (10,78 %) от общего количества ошибок идентификации коллокаций в подкорпусе законодательных текстов составляют многословные устойчивые сочетания, основные морфолого-синтаксические модели которых представлены в Таблице 1. Знаком (*) помечен предполагаемый отсутствующий компонент трехсловных и четырехсловных устойчивых сочетаний. Как показывает анализ таблицы, проблемы идентификации возникают при анализе как лексических, так и грамматических двухсловных коллокаций. Максимальный процент ошибок составляет идентификация более коротких трехсловных устойчивых сочетаний. Чаще всего – 35 раз на тысячу пар – двухсловные устойчивые сочетания смешиваются с трехсловной грамматической коллокацией с предложным управлением PrpAN* (*в установлений**, *на фондовый**). Тем не менее, в общем списке лексические коллокации чаще, чем грамматические, ошибочно идентифицируются как двухсловные.

Табл. 1. Основные ошибки идентификации многословных конструкций

Трехкомпонентные конструкции			Четырехкомпонентные конструкции		
модель	кол-во ошибок	пример	модель	кол-во ошибок	пример
PrpAN*	35	<i>в іноземний*</i>	N*N*NN	1	<i>**здорів'я України</i>
NAN*	33	<i>орган місцевий*</i>	N*A*NN	1	<i>**справа Україна</i>
AAN*	20	<i>вищій навчальний*</i>	PrpAN*N*	1	<i>в автономний**</i>
A*NN	6	<i>*сила Україна</i>			
PrpNN*	4	<i>*межа Україна</i>			
VAN*	3	<i>скласти верховний*</i>			
N*NN	2	<i>*міністр України</i>			
N*AN	1	<i>*об'єднаний нація</i>			
NPrpN*	1	<i>раз на*</i>			
PrpNumN*	1	<i>протягом три*</i>			
ВСЕГО:	106		ВСЕГО:	3	

При этом с двухсловными устойчивыми сочетаниями смешиваются 6 моделей трехсловных (NAN*, AAN*, A*NN, VAN*, N*NN, N*AN) и 2 модели четырехсловных лексических коллокаций (N*N*NN, N*A*NN), а также 3 модели трехсловных (PrpAN*, PrpNN*, NPrpN*) и 2 четырехсловных грамматических коллокации (PrpNumN*, PrpAN*N*). Предварительно можно заключить, что по результатам анализа украинских законодательных текстов в отношении контаминации чаще вступают лексические коллокации. Гипотетически можно предположить, что грамматические связи в двухсловных коллокациях характеризуются большей степенью жесткости и фразеологизированности. Кроме того, со снижением показателя частоты потенциальных коллокаций наблюдается закономерное увеличение количества ошибочно идентифицированных многословных сочетаний, что подтверждается данными о максимальной частоте употребления сочетаний с минимальным количеством составляющих компонентов (Перебийніс *и др.* 2008: 446).

Эффективность идентификации коллокаций также зависит от правильной лемматизации компонентов двухсловных устойчивых сочетаний, выполненной программными средствами Корпуса текстов украинского языка. По предварительным данным анализа тысячи самых частотных пар общий процент неразрешенной в автоматическом режиме омонимии составил 2,67 %. При этом основную проблему для системы представляет снятие лексико-грамматической омонимии пяти видов, представленных ниже в Таблице 2.

Табл. 2. Ошибки лемматизации компонентов коллокаций

Лексико-грамматическая омонимия			Грамматическая омонимия		
вид	кол-во ошибок	пример	вид	кол-во ошибок	пример
Prn-N	6	<i>той-том</i>	Prn-Prn	4	<i>її-вона</i>
N-V	6	<i>склад-скласти</i>			
N-A	4	<i>рада-рад</i>			
V-V	4	<i>бути-бутити</i>			
N-N	2	<i>стаття-стать</i>			
Conj-N	1	<i>того-тога</i>			
ВСЕГО:	23		ВСЕГО:	4	

Как показывает анализ таблицы, максимальный процент ошибок выявляется при идентификации частотных коллокаций с предложным управлением (*у том, до скласти*). Кроме описанных выше фактов наблюдается также совмещение или наложение нескольких типов ошибок. Например, ошибочная идентификация части многословной устойчивой конструкции вследствие неразрешенной грамматической омонимии PrpPrnN* – *в вона – в її (складі*)* или лексико-грамматической омонимии (Prn-Conj) – Prp*PrnN – (*у**) *тому число*. Наименьший процент от общего количества ошибок составляют случаи установления синтаксической связи между словами (1,09 %), находящимися в контактной позиции в тексте, но не связанными между собой синтаксически. В этом аспекте основной ошибкой является идентификация двухсловных грамматических коллокаций NPrp – 9 (*Україна від*), NConj – 1 (*громадянин яка*), ConjPrn – 1 (*яке вони*).

Применяемый в данном исследовании корпусно-ориентированный подход позволяет выявить непосредственно наблюдаемые в текстах модели двухсловных коллокаций, релевантные для украинского юридического дискурса (см. Табл. 3). При этом, не все установленные на материале подкорпуса законодательных документов морфолого-синтаксические модели вписываются в известные классификации коллокаций (M. Venson, F. J. Hausmann, P. Howarth, P. Newmark и др.). Безусловно, это объясняется в первую очередь характеристиками исследуемого текстового материала, в частности, функциональным стилем законодательных документов. Так, среди грамматических коллокаций максимальными показателями частоты характеризуются составные предлоги – PRP: *у разі* – 1.406, *згідно з* – 1.131.

Табл. 3. Классификация частотных коллокаций на материале законодательных текстов

Лексические коллокации		Грамматические коллокации		Предикативные коллокации	
модель	кол-во	модель	кол-во	модель	кол-во
AN	252	PrpN	215	ConjV	24
NN	134	NPrp	130	NV	5
VN	14	VPrp	40	NV(-но)	6
VV	5	PartV	11	V(-но)	5
VAdv	3	APrp	10	PrnN	3
AV	1	PRP	10	NA	2
AdvV	2	PrpA	5		
		AdvPrp	5		
		PartAdv	5		
		PartA	2		
		NPart	2		
ВСЕГО:	411	ВСЕГО:	435	ВСЕГО:	45

По предварительным результатам классификации устойчивых сочетаний большую часть (48,82 %) автоматически идентифицированных пар составляют грамматические коллокации, хотя модели лексических коллокаций реализуются в текстах чаще: AN (*договірний сторона* – 1902, *верховний рада* – 911, *національний банк* – 886), NN (*конституція Україна* – 1301, *закон Україна* – 1032, *президент Україна* – 821). В подкорпусе законодательных текстов грамматические коллокации представлены 11 моделями, при этом наиболее частотными являются модели предложного управления (PrpN, NPrp, VPrp). Лексические коллокации, составляющие 46,13 % устойчивых сочетаний, представлены 7 моделями сочетаемости. Максимальной частотой отличаются субстантивные модели сочетаемости, соответствующие неоднословным наименованиям (*кабінет міністр*, *новий редакція*) и терминологическим сочетаниям (*цінні папери*, *охорона здоров'я*), характерные для данного домена законодательных текстов. Незначительную часть извлеченных из подкорпуса законодательных текстов коллокаций составляют предикативные структуры (5,05 %), реализуемые в текстах 6 моделями, наиболее частотной из них является ConjV, в которой роль подлежащего выполняет союзное слово: *яке бути* – 423, *яке мати* – 268.

Полученные в результате классификации списки моделей двухсловных устойчивых сочетаний могут быть использованы для решения различных прикладных задач автоматического анализа украинского текста – морфолого-синтаксической и семантической разметки, снятия омонимии, и непосредственно для машинного обучения системы распознавания коллокаций. Автоматически составленные на материале подкорпуса законодательных текстов списки содержат коллокации, характерные не только для современного юридического дискурса, но и для украинского языка в целом, что определяет перспективы верификации результатов исследования на материале корпуса украинских текстов разных функциональных стилей.

Литература

BARTSCH, S., 2004. *Structural and Functional Properties of Collocations in English: A Corpus Study of Lexical and Pragmatic Constraints on Lexical Co-occurrence*. Tübingen: Gunter Narr Verlag.

BENSON, M., BENSON, E., ILSON, R., 2009. Your guide to collocations and grammar. In: *The BBI Combinatory Dictionary of English*. Amsterdam: John Benjamins.

DURRANT, Ph., DOHERTY A., 2010. Are high-frequency collocations psychologically real? *Corpus Linguistics and Linguistic Theory*, 6, 2, 125–155.

EVERT, S., 2005. *The Statistics of Word Cooccurrences Word Pairs and Collocations*. Ph. D. thesis, IMS, Stuttgart: University of Stuttgart. Available at: elib.uni-stuttgart.de/opus/volltexte/2005/2371/pdf/Evert2005phd.pdf [Accessed 19 May 2015].

FONTENELLE, Th., 1994. What on earth are collocations? *English today: the International Review of the English Language*, 10 (40), 4, 42–48.

HANDL, S., 2008. Essential collocations for learners of English. In: Ed. F. MEUNIER; S. GRANGER, *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins Publishing Company, 43–66.

HAUSMANN, F. J., 1984. Wortschatzlernen ist Kollokationslernen. *Praxis des neusprachlichen Unterrichts*, 31, 395–406.

HOWARTH, P., 1998. Phraseology and Second Language Proficiency. *Applied Linguistics*, 19 (1), 24–44.

MARCINKEVIČIENĖ, R., 2010. *Lietuvių kalbos kolokacijos*. Kaunas: Vytauto Didžiojo universiteto leidykla. Available at: http://vddb.laba.lt/fedora/get/LT-eLABa-0001:B.03~2010~ISBN_978-9955-12-656-0/DS.001.0.01.BOOK [Accessed 8 August 2014].

NATION, P., 2001. *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

NEWMARK, P., 1988. *Approaches to Translation*. London-New York-Toronto: Phoenix ELT.

SELJAN, S., GAŠPAR, A., 2012. First Steps in Term and Collocation Extraction from English-Croatian Corpus. In: ed. S. SELJAN. *Computational Language Analysis, Computer-Assisted Translation and e-Language Learning*. Zagreb: Zavod za informacijske studije, 149–156.

SERETAN, V., 2011. *Syntax-Based Collocation Extraction*. Berlin: Springer Science & Business Media.

SINCLAIR, J., 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.

SMADJA, F. A., McKEOWN, K. R., 1990. Automatically Extracting and Representing Collocations for Language Generation. In: *Proceedings on the 28-th Annual Meeting of the ACL*. Pittsburg: PA, 252–259.

TEUBERT, W., 2009. Linguistique de Corpus: un Alternative. In: *Semen. Critical Discourse Analysis I*. Besançon: PUF, 27, 130–152.

TOGNINI-BONELLI, E., 2000. Corpus classroom currency. *Darbai ir Dienos*, 24, 205–243. Available at: <http://donelaitis.vdu.lt/publikacijos/bonelli.pdf> [Accessed 4 April 2015].

WERMTER, J., 2008. *Collocation and Term Extraction Using Linguistically Enhanced Statistical Methods*. Jena: der Friedrich Schiller Universität. Available at: <http://d-nb.info/993920594/34> [Accessed 4 April 2015].

БОБКОВА, Т., 2015. Лексикографический корпус как источник для словарей нового типа. *Žmogus kalbos erdvėje*, 8, 15–26.

БОЛЬШАКОВА, Е. И., КЛЫШИНСКИЙ, Э. С., ЛАНДЭ, Д. В., 2011. *Автоматическая обработка текстов на естественном языке и компьютерная лингвистика*. Москва: МИЭМ.

ГЛАДКА, В. А., 2013. Структурно-синтаксичний підхід у вивченні колокацій (на матеріалі французької мови). In: *Наукові записки національного університету "Острозька академія". Серія "Філологічна"*, 39, 16–20.

ДАРЧУК, Н., 2013. *Комп'ютерне анування українського тексту: результати і перспективи*. Київ: Освіта України.

ЗАЦЕРКЛЯНИЙ, М. М., УЗЛОВ, Д. Ю., 2013. Об'єктно-орієнтований тезаурус і словник колокацій для бази знань криміналістичних інформаційних систем. *Системи обробки інформації*, 2, 183–186.

ИОРДАНСКАЯ, Л. Н., МЕЛЬЧУК, И. А., 2007. *Смысл и сочетаемость в словаре*. Москва: Языки славянских культур.

ЛЕВИЦКИЙ, В. В., 2006. *Семасиология*. Винница: Нова Книга.

ПЕРЕБИЙНИС, В. І., БОБКОВА, Т. В., 2008. Частота мовних одиниць як відображення їхніх системних характеристик. *In: Ред. В. В. ЛЕВИЦКИЙ. Проблеми загального, германського та слов'янського мовознавства*. Чернівці: Книги–XXI, 446–453.

ХОХЛОВА, М. В., 2010. *Исследование лексико-синтаксической сочетаемости в русском языке с помощью статистических методов (на базе корпусов текстов)*. Автореферат диссертации. Санкт-Петербург.

ШКУРКО, В. В., 2012. Лексикографічний агент екстракції колокацій у природномовному тексті. *Вісник Київського національного університету ім. Т. Шевченка. Серія: Літературознавство. Мовознавство. Фольклористика*, 28, 31–35.

ЯГУНОВА, Е. В., ПИВОВАРОВА, Л. М., 2011. От коллокаций к конструкциям. *In: Труды Института лингвистических исследований РАН*. Режим доступа: http://www.webground.su/data/lit/pivovarova_yagunova/Ot_kollokatsiy_k_konstruktsiyam.pdf [См. 14 April 2015].

Tatiana Bobkova

Kyiv National Linguistic University, Ukraine

Research interests: applied linguistics, corpus linguistics, corpus lexicography, syntax

CLASSIFICATION OF COLLOCATION: THE MAIN APPROACHES AND CRITERIA

Summary

The article deals with the system-oriented and text-oriented approaches to the classification of collocations, which differ in the criteria and the sequence of applied procedures. The developed methodology of collocation identification with the help of statistical analysis of the texts and lemmatization allows extracting the fixed two-word combinations from the corpus of Ukrainian texts automatically. Three-word co-occurrences, mistakenly identified as two-word, homonymy and positional co-occurrence of the words that are not related syntactically present the main problems of the collocation identification and classification. The appliance of the corpus-oriented approach provided the revelation of two-word lexical, grammatical and predicative collocations, which are relevant to the modern Law discourse, observed in the subset of the sub-corpus of the Ukrainian Law Acts. The use of a bigger corpus and linguistically-motivated filters guarantees the increasing efficiency of the results of collocation identification and classification.

KEY WORDS: collocation, text corpus, collocation classification, identification, extraction, collocation criteria.