

ВЕРИФІКАЦІЯ КОГНІТИВНОЇ ТЕРМІНОСТРУКТУРИ *REFUGEE/БИЖЕНЕЦЬ* В АНГЛО-УКРАЇНСЬКОМУ ПАРАЛЕЛЬНОМУ КОРПУСІ ЮРИДИЧНИХ ТЕКСТІВ

Матвєєва С. А.

ВСТУП

Сучасний когнітивний підхід до розуміння термінологічної одиниці як результату пізнання зумовлює введення до наукового обігу поняття *когнітивної терміноструктури*, під якою розуміємо одиницю свідомості, у якій сконцентовано фрагмент професійної картини світу та яка має вербальну реалізацію у вигляді термінів.

Широкі можливості для здійснення верифікації когнітивної терміноструктури надають корпуси текстів, серед яких – створений автором цієї роботи англо-український паралельний корпус юридичних текстів. Теоретичне значення та практична мета розробки такого паралельного корпусу полягають у можливості проведення компаративного лінгвістичного аналізу на репрезентативному масиві відібраних за певними критеріями текстів. Такий підхід до аналізу термінологічних одиниць двох та більше мов представляє фактично нову парадигму лінгвістичного знання та дослідницької методології, які створюють підґрунтя для формування компаративно-корпусного підходу до аналізу й галузевого перекладу текстів близько-й неблизькоспоріднених мов. Отже, корпусна лінгвістика як така й високий індекс верифікаційного потенціалу математичної обробки корпусних даних у поєднанні із семантичним аналізом забезпечують якісно новий рівень об'єктивної верифікації мовних фактів у межах компаративно-корпусного підходу до аналізу матеріалу англійської та української мов у зіставному аспекті.

Корпусна верифікація використання одиниць в текстах юридичного спрямування заснована на показниках частотності, часу відповіді на запити в автоматичних системах обробки текстів тощо, що дозволяє відстежувати сучасний стан й зміни у функціонуванні одиниць у дискурсі, корегувати перекладацьку діяльність та вносити лінгвістичні та лексикографічні корективи на основі своєчасного аналізу процесів і результатів трансформацій міжмовних зв'язків термінологічних одиниць певної галузі.

Алгоритм корпусної верифікації базується на процедурі вилучення досліджуваних одиниць (за різними показниками в залежності від

дослідницького завдання) з корпусного масиву англомовних текстів, параметризації та зіставлення цих одиниць із еквівалентами з текстів-перекладів українською мовою. Залучення корпусного аналізу для визначення семантико-типологічних характеристик юридичних термінів, їх порівняння й зіставлення з еквівалентами мови-перекладу матиме результатом удосконалення лексикографічних та перекладацьких матеріалів двох мов – англійської та української, гармонізування термінів і понять у галузі права, а також лінгвістичного забезпечення інформаційних систем, комп'ютерної термінографії та методології корпусної лінгвістики.

Для проведення повного циклу корпусної верифікації термінологічних одиниць англійської та української мов нами запропоновано система послідовних етапів проведення комплексного лінгвістичного, фреймового аналізу й корпусної верифікації отриманих даних:

- 1) вибірка лексем та словосполучень, що вербалізують когнітивну терміноструктуру REFUGEE в корпусних текстах англійською мовою; укладання реєстру;
- 2) вибірка лексем та словосполучень, що вербалізують когнітивну терміноструктуру БІЖЕНЕЦЬ у корпусних текстах українською мовою; укладання реєстру;
- 3) семантико-дистрибутивна верифікація корпусних та лексикографічних даних;
- 4) відбір колокацій із досліджуваними термінами в корпусних текстах англійською та українською мовами;
- 5) тезаурсна верифікація вербалізаторів когнітивної терміноструктури REFUGEE / БІЖЕНЕЦЬ.

Для досягнення перекладацьких цілей пропонується додати ще два етапи – (6) вибірка еквівалентів перекладу відібраних одиниць англійської мови українською з укладанням реєстру та (7) верифікація даних перекладацьких словників в англо-українському корпусі юридичних текстів. Така перекладацька верифікація потребує спеціального вивчення та стане предметом окремої розвідки.

1. Семантико-дистрибутивна верифікація вербалізаторів когнітивної терміноструктури REFUGEE/БІЖЕНЕЦЬ

Сучасне юридичне тлумачення терміна *refugee* зафіксовано в тексті *Конвенції про статус біженців*, прийнятої Організацією Об'єднаних націй 1951 року. Відповідно до цієї Конвенції, біженцем вважають “any person who ... owing to wellfounded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group or political opinion, is outside the country of his nationality and is unable or, owing to such fear, is unwilling to avail himself of the protection of that country; or who, not having a nationality and being outside the country of

his former habitual residence as a result of such events, is unable or, owing to such fear, is unwilling to return to it”¹ (‘особу, яка має обґрунтований страх переслідування через расу, релігію, національність, належність до певної соціальної групи чи політичну думку, перебуває за межами країни свого громадянства і не має можливості або через такий страх не бажає скористатися захистом цієї країни; чи особу без громадянства, що перебуває за межами країни свого попереднього звичайного проживання внаслідок таких подій і не має можливості або через такий страх не бажає повертатися до неї’).

Антропоцентричність і соціальне навантаження феномена біженства пояснює багатогалузевий характер термінів *refugee* та *біженець* – вони функціонують одразу в декількох терміносистемах, набуваючи різних тлумачень і відтінків термінологічної семантики в залежності від екстралінгвістичних підстав, які зумовлюють багатоманітність акцентів на певних складових цього явища (суспільних, економічних, психологічних тощо) і, відповідно, визначають його термінологічний зміст.

Побудовані на основі даних лексикографічних джерел та текстових масивів моделі когнітивної терміноструктури REFUGEE / БІЖЕНЕЦЬ (див. рис. 1 та 2) є типізованими моделями слотової структури, в яких кожен слот наповнений різними лексичними змінними, що вербалізують фрагменти професійної картини світу правника. Семантико-дистрибутивна верифікація досліджуваних термінологічних одиниць дозволяє виявити семантичні компоненти, які збігаються в значеннях лексем, а також змінні компоненти та зробити висновки про причини відмінностей таких складових у побудованих моделях й, відповідно, лексичних одиницях на основі врахування ступеня розробленості терміна в національній (англійській та українській) метамові юриспруденції.

Застосування процедур лінгвістичної реконструкції концептів і когнітивних терміноструктур для тлумачення термінів у межах галузевих дискурсів дозволяє проаналізувати термінологічне значення на основі вивчення словникових дефініцій (термінологічного й нетермінологічного характеру), врахування дистрибуції та впливу екстралінгвістичних елементів, утворення символічних смислів, які концентруються навколо кожної окремої одиниці й цілого тексту, в якому ця лексема функціонує, а також аналіз контекстуального оточення через залучення корпусного підходу та аналізу тезаурусних зв'язків. Така реконструкція має результатом побудову генералізованих моделей причинно-наслідкової семантики юридичних термінів

¹ Convention and Protocol relating to the Status of Refugees. URL : <https://www.unhcr.org/protection/basic/3b66c2aa10/convention-protocol-relating-status-refugees.html>.

біженець і *refugee*, які демонструють наявність спільних і відмінних семантичних формантів відповідно до рівнів нарощення значення одиниць, що досліджуються (див. рисунок 3, де 1 – семми, виявлені при лексикографічному аналізі даних нетермінологічних тлумачних словників; 2 – семми, виявлені при лексикографічному аналізі даних термінологічних тлумачних словників; 3 – семми, виявлені при фреймовому аналізі даних корпусних текстів юридичного дискурсу).

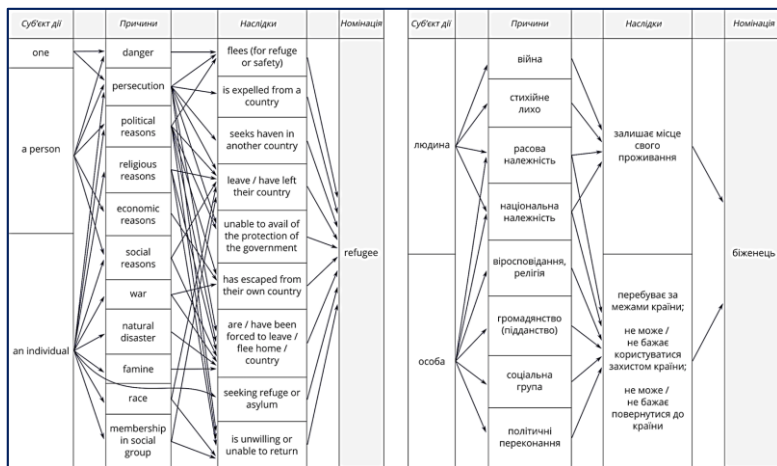


Рис. 1. Семантичне наповнення дискретних термінологічних одиниць *refugee* та *біженець* (за матеріалами словників юридичної термінології)

Спеціальні корпуси текстів дозволяють вивчати показники частотності використання термінів у текстах, контекстуальне оточення, зрушення значень, нормативне й індивідуально-стилістичне вживання слів, а також ті зв'язки, які ці терміни вибудовують у конкретному тексті й у цілому юридичному дискурсі. Іншими словами, завдяки корпусній організації матеріалу та можливості застосування корпусних інструментів роботи з одиницями мови для кожної термінологічної одиниці може бути організована поетапна верифікація її валідності з різних позицій, вагомих для конкретних цілей дослідження (напр., специфіка функціонування, дискурсивні характеристики, набуття, закріплення чи втрата статусу терміна тощо), якими доповнюються дані лексикографічних джерел та контекстуального аналізу.

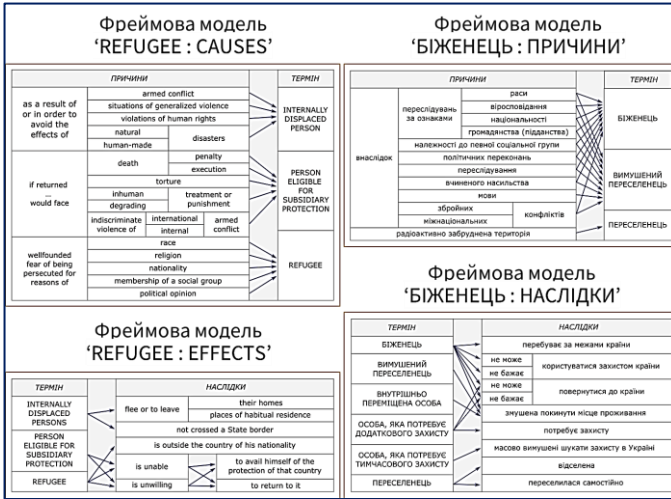


Рис. 2. Фреймові моделі когнітивної терміноструктури REFUGEE / БІЖЕНЕЦЬ (за матеріалами англо-українського паралельного корпусу юридичних текстів)

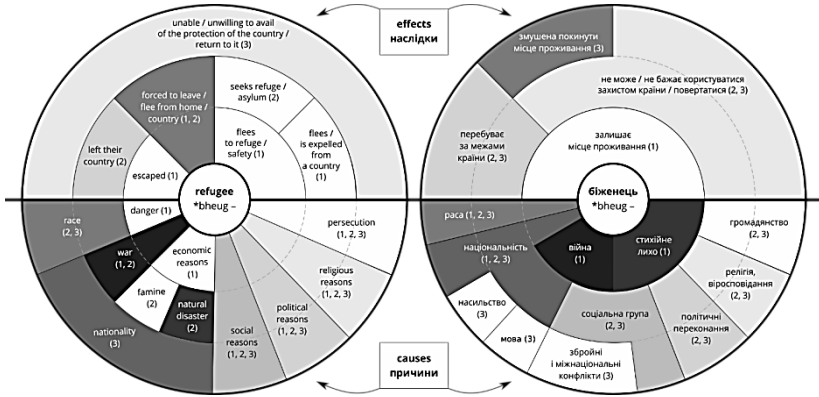


Рис. 3. Генералізовані моделі причинно-наслідкової семантики юридичних термінів *refugee* та *біженець*

Оскільки всі корпуси містять різну кількість одиниць та слововживань, виникла потреба введення певного об'єктивного критерію, яким став показник відносної частотності. Цей показник може бути застосований для виявлення найбільш активної

у функціональному плані термінологічної одиниці на позначення когнітивної терміноструктури REFUGEE / БІЖЕНЕЦЬ у корпусі текстів, оскільки використовується для демонстрації частки мовної одиниці в цілому тексті.

Частотність вживання терміноодиниці розраховується нами як відношення числа входжень досліджуваної одиниці до загального обсягу слів у корпусі. Результат множимо на 1 000 000 для зручності операційних дій із отриманими числовими даними (оскільки у результаті поділу отримуємо дрібну частотність) – ipm (instances per million / відносна частотність, помножена на мільйон), наприклад (для терміна *refugee*):

$$\text{refugee: } 95 / 1\,135\,827 * 1\,000\,000 = 84$$

За цією ж формулою проведені підрахунки частотності вживання всіх наявних у корпусі термінологічних одиниць, що використовуються для вербалізації когнітивної терміноструктури REFUGEE / БІЖЕНЕЦЬ:

REFUGEE (див. таблицю 1):

$$\text{internally displaced person: } 69 / 1\,135\,827 * 1\,000\,000 = 61;$$

$$\text{fugitive: } 61 / 1\,135\,827 * 1\,000\,000 = 54;$$

$$\text{asylum seeker: } 28 / 1\,135\,827 * 1\,000\,000 = 25;$$

$$\text{escapee: } 20 / 1\,135\,827 * 1\,000\,000 = 18;$$

$$\text{displaced person: } 8 / 1\,135\,827 * 1\,000\,000 = 7;$$

$$\text{person eligible for subsidiary protection: } 6 / 1\,135\,827 * 1\,000\,000 = 5;$$

$$\text{runaway: } 5 / 1\,135\,827 * 1\,000\,000 = 4;$$

БІЖЕНЕЦЬ (див. таблицю 2):

$$\text{біженець: } 94 / 1\,135\,827 * 1\,000\,000 = 83;$$

$$\text{внутрішньо переміщена особа: } 91 / 1\,135\,827 * 1\,000\,000 = 80;$$

$$\text{переміщена особа: } 68 / 1\,135\,827 * 1\,000\,000 = 60;$$

$$\text{переселенець: } 12 / 1\,135\,827 * 1\,000\,000 = 11;$$

$$\text{особа, яка потребує додаткового захисту: } 8 / 1\,135\,827 * 1\,000\,000 = 7;$$

$$\text{вимушений переселенець: } 5 / 1\,135\,827 * 1\,000\,000 = 4;$$

$$\text{особа, яка потребує тимчасового захисту: } 3 / 1\,135\,827 * 1\,000\,000 = 3.$$

За результатами проведених кількісних підрахунків з'ясовано, що найвищий показник частотності в англійській мові отримуємо для термінологічної одиниці *refugee* (84), найнижчий – *runaway* (3):

В українській мові найвищий ступінь частотності в корпусі виявлено для терміна *біженець* (83), найнижчий – для терміна *особа, яка потребує тимчасового захисту* (3):

Така унормована одиниця забезпечує об'єктивний підхід до порівняння частотності синонімічних та асоціативно пов'язаних одиниць у межах одного корпусу, а також будь-яких одиниць в різних корпусах.

Дистрибутивну верифікацію досліджуваних одиниць проводимо з побудовою конкордансів у різних доступних корпусах текстів. Конкорданс одиниці залежить від типу корпусу, з якого вилучаються дані.

Таблиця 1

**Показники частотності вживання термінів-вербалізаторів
когнітивної терміноструктури REFUGEE в корпусі**

Одиниця	Частотність	Час пошуку
<i>refugee</i>	84	0,20
<i>internally displaced person</i>	61	0,20
<i>fugitive</i>	54	0,21
<i>asylum seeker</i>	25	0,24
<i>escapee</i>	18	0,24
<i>displaced person</i>	7	0,25
<i>person eligible for subsidiary protection</i>	5	0,25
<i>runaway</i>	4	0,26

Таблиця 2

**Показники частотності вживання термінів-вербалізаторів
когнітивної терміноструктури БІЖЕНЕЦЬ у корпусі**

Одиниця	Частотність	Час пошуку
<i>біженець</i>	83	0,19
<i>внутрішньо переміщена особа</i>	80	0,19
<i>переміщена особа</i>	60	0,21
<i>переселенець</i>	11	0,22
<i>особа, яка потребує додаткового захисту</i>	7	0,27
<i>вимушений переселенець</i>	4	0,27
<i>особа, яка потребує тимчасового захисту</i>	3	0,29

Для аналізу терміноодиниці *refugee* та *біженець* використовуємо дані одномовних корпусів англійської, української мов та пілотного англо-українського паралельного корпусу юридичних текстів. Вибірки зроблено з умовою глибини контексту – 4 одиниці ліворуч + 4 одиниці праворуч; наведено перші 10 рядків результатів пошуку.

of legislation , including the Refugee Relief Act and the 1980	Refugee Act	He argues that the American commitment to refugees in the
, instructed his envoy to the United Nations to oppose	refugee agency	devoted to Palestinian Arabs . He wrote : " We should
instituted two other major policy changes : Previously when	refugee applied	for asylum , the INS asked the State Department for a
Here , two boys play in an abandoned car	refugee camp	. # PHOTO (BLACK & WHITE) : Members of
, three days before the fall of Saigon . From	refugee camp	in Arkansas , she followed in her father 's footsteps and
, (Voiceover) : We found Abu Hassan	refugee camp	with his family , at least what remained of his family
hiding through plains and mountains to enter another hell	refugee camp	in Pakistan . // " Dead ? How ? " Soofia
, an Afghan refugee living under a plastic sheet	refugee camp	, he was never going to find his own way on
chased from her home in Sudan , now living	refugee camp	with her father Tahir (ph) . Her face makes
" moonscape " called Hearth and Home at Liberty into	refugee camp	with sustainable , small-plot farms called ... Bush

**Рис. 4. Конкорданс терміна *refugee*
з Корпусу сучасної американської англійської мови²**

² The Corpus of Contemporary American English. URL : <https://www.english-corpora.org/coca>.

remain in their place of asylum , although a Rwandan special	refugee	commission , to help and monitor refugees , was formed in Kigali
, turning Bernice Rubens 's clear-eyed novel about a German	refugee	plano teacher and an 11-year-old jewish prodigy into a star
Consensys had taken on Interactive System Corp and SunSoft Inc	refugee	Doug Miller as vice president of European sales . +1 416 940
she had been forced to leave her home and became a	refugee	in a village in Slovenia near Ljubljana . Now , for a
scales of compensation payments when there is any balancing of	refugee	debts at that final Middle East peace conference . The Custodian
of war , when petrol is scarce : Armed soldiers and	refugee	families were crowded round the foot of the steps leading up to
, Villages in the south , as well as the Palestinian	refugee	camps , were attacked . In May 1988 , two thousand Israeli
Danish bureaucrat blew the whistle on Mr Hocke as that	refugee	education fund ¶ was only supposed to be for Mr Hartling and
has retained its name but now functions mainly as a women	refugee	coordinating body ¶ particularly in Mexico and Nicaragua . # The
of Macbeth . They are playing this week at the Nitcheu	refugee	camp , Malawi ¶ , having previously graced venues in Sierra Leone

Рис. 5. Конкорданс терміна *refugee* з Британського національного корпусу³

<< наміром набути статусу біженця	біженця	незаконно перетнули державний кордон >>
<< надання їм статусу біженця	біженця	відповідно до Закону України >>
<< України " Про біженців	біженців	" (2557-14) >>
<< або набуття статусу біженця	біженця	, якщо вони в >>
<< притулку чи статусу біженця	біженця	. >>
<< задоволення життєвих потреб біженців	біженців	, >>
<< яким надано статус біженця	біженця	в Україні, мають >>
<< (грошову допомогу біженцям	біженцям	; компенсації на медикаменти >>
<< Кожна Сторона надасть біженцям	біженцям	, визначення яких наводиться >>
<< Конвенції про статус біженців	біженців	, підписаній у Женеві >>

Рис. 6. Конкорданс терміна *біженець* із Корпусу текстів української мови⁴.
Зона пошуку – законодавчі тексти

року , " Про імміграцію " від 7 червня 2001 року , " Про біженців	біженців	" від 21 червня 2001 р . , Правила в'їзду іноземців в
Складовою цього режиму є адміністративно-правовий режим біженців	біженців	. Законодавство про біженців також регулює певні відносини
адміністративно-правовий режим біженців . Законодавство про біженців	біженців	також регулює певні відносини імміграційного процесу .
також регулює певні відносини імміграційного процесу . Біженець	біженець	- особа , яка не є громадянином України і внаслідок цілком
засначених побоювань . Строки звернення за набуттям статусу біженця	біженця	. Особи , які мають намір набути статус біженця і перетнули
статусу біженця . Особи , які мають намір набути статус біженця	біженця	і перетнули державний кордон України у порядку , встановленому
органу міграційної служби із заявами про надання їм статусу біженця	біженця	. Особи , які з наміром набути статус біженця намагалися
їм статусу біженця . Особи , які з наміром набути статус біженця	біженця	намагалися незаконно перетнути або незаконно перетнули
органу внутрішніх справ із заявами про надання їм статусу біженця	біженця	, а також надати посадовим особам Державної прикордонної
перетинання державного кордону України . Умови , за яких статус біженця	біженця	не надається . Статус біженця не надається особі : яка

Рис. 7. Конкорданс терміна *біженець* із Генерального регіонально анотованого корпусу української мови⁵. Зона пошуку – юридичні тексти

³ British National Corpus. URL : <https://www.english-corpora.org/bnc>.

⁴ Корпус текстів української мови. URL : <http://www.mova.info>.

⁵ Генеральний регіонально анотований корпус української мови (ГРАК) / М. Шведова, Р. фон Вальденфельс, С. Яригін, А. Рисін, В. Старко, М. Возняк, М. Крук та ін. Київ, Львів, Єна, 2017–2020. URL : uacorp.us.

Застосований далі паралельний корпус поруч із даними щодо семантико-дистрибутивних характеристик досліджуваних одиниць дозволяє одразу простежити перекладацьку інформацію, оскільки пов'язані різномовні речення з'являються у тому самому контексті в одній колекції текстів мовами, що досліджуються (див. табл. 3.).

Таблиця 3

Конкорданс термінів *refugee* та *біженець* із пілотного англо-українського паралельного корпусу юридичних текстів

Текст англійською			Текст українською		
On	Refugees	and Persons in Need	“Про	біженців	та осіб, які потребують
the status of a	refugee,	a person in need	у сфері визнання особи	біженцем,	особою, яка потребує додаткового
the legal status of	refugees,	Persons in need of	також встановлення правового статусу	біженців	та осіб, які потребують
of the Law, a	refugee	is a person who	до статті 1 Закону	біженець	– це особа, яка не
who is not a	refugee	according to the 1951	особа, яка не є	біженцем	відповідно до Конвенції про
to the Status of	Refugees,	the 1967 Protocol relating	до Конвенції про статус	біженців	1951 року і Протоколу
to the Status of	Refugees,	and this Law, but	і Протоколу щодо статусу	біженців	1967 року та Закону,
A	refugee	or a person in		Біженець	чи особа, яка потребує
A	refugee	or a person in		Біженець	чи особа, яка потребує
the status of a	refugee	or a person in	із заявою про визнання	біженцем	або особою, яка потребує

Зведені дані з п'яти фрагментів, відібраних із різних корпусів за однаковими пошуковими критеріями, демонструють такі кількісні показники сполучуваності досліджуваного терміна (з метою демонстрації наводяться найближчі сполучення (досліджувана одиниця + одна одиниця праворуч чи ліворуч), кількість використань яких в аналізованих фрагментах корпусів більше за 1):

refugee: загальна кількість слововживань – 30 – 100%:

сполучуваність праворуч із одиницями самостійних частин мови:

refugee camp – 9 – 30%;

сполучуваність праворуч із одиницями службових частин мови:

refugee or – 3 – 10%;

refugees and – 2 – 6,67%;

сполучуваність ліворуч із одиницями службових частин мови:

a refugee – 16 – 53,33%;

of refugee – 4 – 13,33%;

біженець: загальна кількість слововживань – 30 – 100%:

сполучуваність праворуч із одиницями самостійних частин мови:

біженець – особа – 3 – 10%;

біженець відповідно – 2 – 6,67%;

сполучуваність праворуч із одиницями службових частин мови:

біженець та – 2 – 6,67%;

біженець чи – 2 – 6,67%;

сполучуваність ліворуч із одиницями самостійних частин мови:

статус біженця – 15 – 50%;

сполучуваність ліворуч із одиницями службових частин мови:

про біженців – 5 – 16,67%.

Результати аналізу сполучуваності лексем *refugee* та *біженець* продемонстрували використання цих одиниць з іменниками із значеннями явищ, що супроводжують життя людини в статусі біженця (напр., *refugee camp, refugee education; життєві потреби біженців* тощо), законодавче та соціальне забезпечення біженців (напр., *refugee act, refugee agency, refugee commission, refugee coordinating body, legal status of refugee; адміністративно-правовий режим біженців, визнання біженцем, законодавство про біженців, статус біженця, грошова допомога біженцям* тощо). Результати спостережень за терміноодиницями, що аналізуються, в корпусних текстах дає підстави стверджувати про використання лексем відповідно до їх термінологічних значень і нормативних уживань: виявлені словосполучення відбивають основні семантичні та асоціативні акценти термінів. Специфіка дистрибуції дозволяє уточнювати юридичне термінологічне значення досліджуваних одиниць та за рахунок побудування асоціативних зв'язків деталізувати сполучуваність одиниць у їх термінологічному та нетермінологічному використаннях.

Семантична та дистрибутивна верифікації терміноодиниць через роботу з пошуковими механізмами корпусного типу, застосування математичних підрахунків для обробки даних та інформації забезпечують високу валідність отриманих результатів та наближають часто описові лінгвістичні висновки до точних і об'єктивно підтверджених. При цьому без відповіді залишається питання нульового результату, тобто відсутності у результатах корпусного пошуку певних одиниць, що може свідчити не тільки про відсутність таких одиниць у мові, але й про недостатній рівень репрезентативності залученого корпусу.

Разом із семантичною верифікацією комплексний корпусний аналіз термінологічних знаків передбачає верифікацію з огляду на зв'язки кожної одиниці: синонімічні, гіперо-гіпонімічні та асоціативні.

2. Тезаурусна верифікація вербалізаторів когнітивної терміноструктури REFUGEE/БІЖЕНЕЦЬ

Завданням будь-якого тезаурусу є максимальне охоплення змістового наповнення когнітивної терміноструктури і термінів, що її оточують у межах конкретної предметної галузі, за рахунок побудування зв'язків різних рівнів між лексичними одиницями, що забезпечить якісне розмежування значень аналізованих синонімічних та асоційованих лексичних одиниць. Ця проблема – класифікація зв'язків у певній групі та пошук одиниць із заданою кількістю властивостей / характеристик, що перетинаються, – є одним із найскладніших завдань з огляду на велику кількість варіантів реалізації будь-яких семантичних відношень і зв'язків.

Метою тезаурусної верифікації термінів є розширення базової онтології юридичної терміносистеми через корпусний пошук та валідацію релевантних термінологічних варіантів базової термінологічної одиниці.

Семантична відстань між синонімами вимірюється з опорою на корпусні дані. Для підвищення об'єктивності структурування синонімічного ряду лексеми *refugee* корпусному аналізу передують збір лексикографічних даних словників і тезаурусів щодо синонімічних ланцюгів і місця кожного синоніма в них. Для цього пропонуємо узагальнену таблицю синонімів із зазначенням місця кожної одиниці (на прикладі англійської лексеми *refugee*).

Узагальнений кількісний аналіз синонімічного фонду лексеми *refugee* виявляє такі дані: найуживаніші синоніми за кількістю входжень – *exile* (4 входження), *displaced person* (3 входження), *runaway* (3 входження); найближчі за місцем до домінанти (за алфавітом): *alien*, *deportee*, *exile*, *fugitive*. Поєднання двох критеріїв виявляє лексему *exile* найчастотнішим і найближчим синонімом до терміна *refugee* за даними проаналізованих лексикографічних джерел.

Для проведення розрахунку семантичної близькості вербалізаторів когнітивної терміноструктури REFUGEE / БІЖЕНЕЦЬ застосовано квантитативну методику П. Жаккара, за допомогою якої можна удосконалити рівень об'єктивності даних та технології автоматизованого створення онтологій.

Подібність множин за Полем Жаккаром є бінарною мірою подібностей і вираховується як міра спільної частини, поділена на міру об'єднання множин⁶:

$$K_J = C / [(A + B) - C],$$

де С – це кількість спільних елементів значень одиниць, що порівнюються, А – кількість елементів значення першої одиниці, що порівнюється, В – кількість елементів значення другої одиниці, що порівнюється.

Таблиця 4

Синоніми лексики *refugee*

	<i>OT</i>	<i>MW</i>	<i>Th</i>	<i>CCELD</i>	<i>Kipfer</i>
<i>alien</i>			1		Дані не враховують через розташування одиниць за алфавітом, а не семантичною близькістю до домінанти
<i>asylum seeker</i>				2	
<i>boat person</i>			23		
<i>castaway</i>			9		
<i>defector</i>			10		
<i>deportee</i>		1			
<i>derelict</i>			11		
<i>deserter</i>			12		
<i>displaced person</i>	4		2	4	
<i>dp</i>	5		8		
<i>emigrant</i>			3		
<i>émigré</i>		2	27	3	
<i>escapee</i>	3		13	7	
<i>evacuee</i>		3	4		
<i>exile</i>	6	4	5	1	
<i>expat</i>		5			
<i>expatriate</i>		6	6		
<i>expellee</i>			24		
<i>foreigner</i>			7		
<i>foundling</i>			14		
<i>fugitive</i>	1		15	6	
<i>homeless person</i>			25		
<i>leper</i>			16		
<i>maroon</i>			17		
<i>migr</i>	7				
<i>outcast</i>			18		
<i>outlaw</i>			19		
<i>prodigal</i>			20		
<i>renegade</i>			21		
<i>runaway</i>	2		22	5	
<i>stateless person</i>			26	8	

⁶ Leskovec J., Rajaraman A., Ullman J. D. Mining of Massive Datasets. Cambridge University Press, 2014. P. 74.

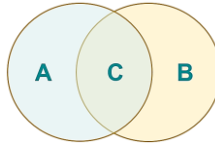


Рис. 8. Умовна схема перетину та об'єднання двох множин

Цей показник використовуємо для формального визначення міри змістового збігу вербалізаторів когнітивної терміноструктури. Якщо цей коефіцієнт дорівнює нулю, множини не мають спільних елементів, а якщо дорівнює одиниці – множини є рівними.

Із текстового масиву англо-українського паралельного корпусу юридичних текстів вилучено чотири вербалізатори когнітивної терміноструктури REFUGEE: *refugee – asylum seeker, internally displaced persons, person eligible for subsidiary protection* і шість вербалізаторів когнітивної терміноструктури БІЖЕНЕЦЬ: *біженець – вимушений переселенець, внутрішньо переміщена особа, особа, яка потребує додаткового захисту, особа, яка потребує тимчасового захисту, переселенець*.

Для підрахунку коефіцієнтів подібності до цих термінологічних одиниць застосовано методику П. Жаккара (порівняння проводимо з іменами когнітивної терміноструктури – *refugee* та *біженець*). Результати наведено в таблицях 5 і 6:

Таблиця 5

Коефіцієнт семантичної подібності термінологічних одиниць, що вербалізують когнітивну терміноструктуру REFUGEE

№	<i>refugee &</i>	Коефіцієнт Жаккара
1	<i>asylum seeker</i>	0
2	<i>internally displaced person</i>	0
3	<i>person eligible for subsidiary protection</i>	0

Таблиця 6

Коефіцієнт семантичної подібності термінологічних одиниць, що вербалізують когнітивну терміноструктуру БІЖЕНЕЦЬ

№	<i>біженець &</i>	Коефіцієнт Жаккара
1	<i>вимушений переселенець</i>	0,8
2	<i>внутрішньо переміщена особа</i>	0,5
3	<i>особа, яка потребує додаткового захисту</i>	0,3
4	<i>особа, яка потребує тимчасового захисту</i>	0,2
5	<i>переселенець</i>	0

Визначені відмінності в утворенні когнітивних терміноструктур англійської та української мов зумовлені галузевими факторами та чинниками юридичних систем, що склалися історично, а саме, відмінностями між англо-саксонським і романо-германським типами правових систем. Як відомо, ці два класичні типи правових систем мають різні джерела права (для романо-германського типу основним можна вважати нормативно-правові акти, для англо-саксонської – правовий прецедент), що призводить до різних підходів до тлумачення термінів і понять юридичної сфери. Більше того, «на відміну від юристів країн англо-американського права, що мислять конкретно, «прецедентно» і створили науку загального права на основі судових рішень, юристи романо-германського права мислять поняттями правових інститутів і галузей права. Їх правова наука – абстрактна, одним із визначальних ознак якої є правовий формалізм»⁷. Звідси й такі відмінності в процесах номінації й дефініції: в англійській мові визначення кожного дискретного терміна відокремлене від інших і не допускає перехресних сем та нечіткості визначень, у той час як в українській мові розповсюджена термінологічна синонімія і дублетність.

Отримані результати у виявлених подібностях і відмінностях семантики досліджуваних вербалізаторів дозволять запропонувати нові, більш точні й правильні тлумачення, словникові описання й перекладацькі еквіваленти цих одиниць. Подальше поєднання корпусів спеціальних текстів (одно- і дво- / багатомовних) і словників (синонімів), інтегрованих в автоматизовані системи, дозволить зливати дані цих двох джерел в автоматичному режимі, що пришвидшить пошук одиниць, підвищить точність зіставлення й удосконалисть якість створених на основі цих процедур онтологій.

ВИСНОВКИ

Компаративно-корпусний аналіз вербалізаторів когнітивної терміноструктури REFUGEE / БІЖЕНЕЦЬ продемонстрував подібність в отриманих даних двох мов у семантичній реалізації, сполучуваності та контекстуальному функціонуванні досліджуваних термінів, що свідчить на користь ознак спільності когнітивної основи та універсальності сприйняття явищ, що називають аналізовані лексичні одиниці.

Пріоритетними цілями використання корпусів у лінгвістичних дослідженнях є можливість швидкої обробки великого обсягу текстів та отримання точних кількісних результатів. Більше того, поєднання корпусного підходу з традиційним аналізом словника дозволяє отримати систематизовані та доведені математичним шляхом дані.

⁷ Скаун О.Ф. Теорія права і держави. Київ : Алерта; КНТ; ЦУЛ, 2010. С. 330.

Корпусна лінгвістика як така й високий індекс верифікаційного потенціалу математичної обробки корпусних даних у поєднанні із семантичним (лексикографічним та фреймовим) аналізом забезпечують якісно новий рівень об'єктивної верифікації мовних фактів у межах компаративно-корпусного підходу до аналізу матеріалу англійської та української мов у зіставному аспекті.

Семантично-дистрибутивна верифікація термінологічних одиниць через роботу з пошуковими механізмами корпусного типу, застосування математичних підрахунків для обробки даних та інформації (зокрема, формули частотності вживання теоріоодиниць) забезпечують високу валідність отриманих результатів та наближають часто описові лінгвістичні висновки до точних і об'єктивно підтверджених. За результатами проведених кількісних підрахунків з'ясовано, що найвищий ступінь частотності в англійській мові отримуємо для термінологічної одиниці *refugee* (84), найнижчий – *runaway* (3). В українській мові найвищий ступінь частотності виявлено для терміна *біженець* (83), найнижчий – для терміна *особа, яка потребує тимчасового захисту* (3).

Разом із семантичною верифікацією та комплексним корпусним аналізом термінологічних знаків проведено корпусну верифікацію зв'язків кожної одиниці: синонімічних, гіперо-гіпонімічних та асоціативних. Для проведення розрахунку семантичної близькості вербалізаторів когнітивної терміноструктури REFUGEE / БІЖЕНЕЦЬ застосовано квантитативну методику П. Жаккара. Підрахунки продемонстрували, що в англійській мові визначення кожного дискретного терміна відокремлене від інших і не допускає перехресних сем та нечіткості дефініцій (коефіцієнт семантичної подібності термінологічних одиниць, що вербалізують когнітивну терміноструктуру REFUGEE, дорівнює 0); в українській мові розповсюджена термінологічна синонімія і дублетність (коефіцієнт семантичної подібності термінологічних одиниць, що вербалізують когнітивну терміноструктуру БІЖЕНЕЦЬ, варіюється від 0 до 0,8).

Отже, компаративно-корпусний підхід до аналізу одиниць доводить вплив контексту на зміни в концептуальному значенні і модифікацію всієї когнітивної терміноструктури, а залучення екстралінгвістичного контексту дозволяє уточнювати особливості вербальної реалізації терміноструктури залежно від специфіки сфери її функціонування та лінгвокультурної специфіки професійної спільноти.

АНОТАЦІЯ

Статтю присвячено виконанню верифікації когнітивної терміноструктури REFUGEE / БІЖЕНЕЦЬ у корпусі текстів. Дослідницею запропоновані семантико-дистрибутивна й тезаурусна

верифікації вербалізаторів когнітивної терміноструктури REFUGEE / БІЖЕНЕЦЬ в англо-українському паралельному корпусі юридичних текстів. Автор доводить, що поглиблене вивчення когнітивних структур юридичного дискурсу (серед яких – REFUGEE / БІЖЕНЕЦЬ) із застосуванням процедур лінгвістичної реконструкції (шляхом теоретичного осмислення явища когнітивної терміноструктури та його дослідження через мовні факти) та кількісно-якісної верифікації термінів-вербалізаторів в англо-українському паралельному юридичному корпусі текстів забезпечує необхідну теоретико-методологічну базу для подальшого відтворення фрагментів професійної картини світу та виявлення закономірностей і відмінностей у засобах вербалізації когнітивних терміноструктур в англійській та українській мовах. Перспективу дослідження складає апробація корпусної верифікації даних перекладацьких словників.

ЛІТЕРАТУРА

1. British National Corpus. URL : <https://www.english-corpora.org/bnc>.
2. Convention and Protocol relating to the Status of Refugees. URL : <https://www.unhcr.org/protection/basic/3b66c2aa10/convention-protocol-relating-status-refugees.html>.
3. Leskovec J., Rajaraman A., Ullman J. D. Mining of Massive Datasets. Cambridge University Press, 2014. 495 p. URL : <https://doi.org/10.1017/cbo9781139924801>.
4. The Corpus of Contemporary American English. URL : <https://www.english-corpora.org/coca>.
5. Генеральний регіонально анотований корпус української мови (ГРАК) / М. Шведова, Р. фон Вальденфельс, С. Яригін, А. Рисін, В. Старко, М. Возняк, М. Крук та ін. Київ, Львів, Єна, 2017–2020. URL : uas Corpus.org.
6. Корпус текстів української мови. URL : <http://www.mova.info>.
7. Скакун О.Ф. Теорія права і держави. Київ : Алерта; КНТ; ЦУЛ, 2010. 520 с.

Information about the author:

Matvieieva S. A.,

Doctor of Philological Sciences, Associate Professor,
Professor at the Department of Applied Language Studies,
Comparative Linguistics and Translation
National Pedagogical Dragomanov University
8/14, Turhenivska str., Kyiv, 01054, Ukraine