

ФОКІН С. Б.

Київський національний університет імені Тараса Шевченка

ДИСТРИБУТИВНИЙ АНАЛІЗ ПРИ УКЛАДАННІ ДВОМОВНИХ ПЕРЕКЛАДНИХ СЛОВНИКІВ (на прикладі українсько-іспанських відповідників поля “освіта”)

Статтю присвячено одному з методів підбору міжмовних еквівалентів для перекладних словників, а саме, дистрибуції. Якщо для двох мов, які перебувають у контакті (як іспанська та українська), вже є базовий перекладний словник, проблемним є пошук еквівалентів радше для периферійних, ніж ядерних одиниць лексико-семантичних полів. Метод дистрибуції в поєднанні з використанням електронно згенерованого конкордансу та кількісними підрахунками дає можливість значно поповнити кількість перекладних еквівалентів периферії поля.

Ключові слова: еквівалент, двомовна лексикографія, дистрибуція, конкорданс, лексико-семантичне поле “освіта”.

Статья посвящена одному из методов подбора межъязыковых эквивалентов для переводных словарей, а именно, дистрибуции. Если для двух языков, пребывающих в контакте (как испанский и украинский), имеется базовый переводной словарь, проблемным окажется поиск эквивалентов скорее для периферийных, чем ядерных единиц лексико-семантических полей. Метод дистрибуции в сочетании с использованием сгенерированного компьютером конкорданса и количественными подсчетами дает возможность пополнить количество переводных эквивалентов периферии поля.

Ключевые слова: эквивалент, двуязычная лексикография, дистрибуция, конкорданс, лексико-семантическое поле “образование”.

The article is focused on distribution method for searching interlingual lexical equivalents. If there is a basic dictionary for a pair of languages in contact, the problematic equivalences will be mostly peripheral units of a lexical-semantic field rather than nuclear ones. Distribution method along with computer generated concordance and quantitative measuring allows enriching the list of translational equivalents in bilingual dictionaries.

Key words: equivalent, bilingual lexicography, distribution, concordance, lexical-semantic field “education”.

Актуальність дослідження зумовлюється як прикладною, так і загальнотеретичною проблемами: 1) відсутністю великого українсько-іспанського словника; 2) нерозробленістю ґрунтовної методологічної бази для укладання двомовних перекладних словників, зокрема, принципів встановлення еквівалентних відношень між лексичними одиницями різних мов. Стосовно останнього, яскравим свідченням серйозності й глибини проблеми є твердження В. П. Беркова про те, що чітких і конкретних інструкцій з підбору еквівалентів для перекладних словників немає й бути не може [1:140]. Не позбавлена

певних полемічних моментів, однак, ця позиція вченого відбиває загальний стан речей у двомовній лексикографії.

Деякі методи пошуку еквівалентів для двомовних перекладних словників, такі як калька, запозичення, переносне вживання слова або закріплення за рідним словом відповідного йому значення іноземного слова, згадує Л. В. Щерба [2:294]. Однак не можна не визнати, що ці методи чинні тільки тоді, коли вже є базовий двомовний перекладний словник або наявні білінгви з обох мов, які здатні оцінити близькість значень слів цих мов, або ж, принаймні, є контакт зі спільною мовою-посередником. При відсутності вказаних умов потрібен інший підхід. Так, для малодосліджених мов американських індіанців Ю. Найда розробляє окрему методику укладання словника [3:45–71], але при цьому для дослідника головним залишається опис значення слів, розмежування різних значень слова, а не методика підбору еквівалентів, адже на цьому етапі не принципово, чи робиться словник у вигляді еквівалентів, чи у вигляді металінгвістичного опису.

Таким чином, у проблемі підбору еквівалентів для двомовних перекладних словників слід розмежовувати два принципово відмінних випадки: 1) пошук еквівалентів для ядер лексико-семантичних полів та понять базисного рівня, коли ще немає базового перекладного словника; 2) пошук еквівалентів для периферійних явищ лексико-семантичних полів.

Перша проблема становить винятковий інтерес, але ми наразі зосереджуємось на лексикографічних аспектах тих мов, які перебувають у контакті, як іспанська та українська. Оскільки ядерні поняття різних мов виявляють значну схожість [4:8; 5:39], а поняття базисного рівня, за твердженням С. Г. Шафрікова, вважаються найвірогіднішими кандидатами на мовні універсалії [6:57], встановлення відношень еквівалентності між ними має відносно легко вирішуватись емпіричним шляхом при співпраці групи носіїв обох мов. Зокрема, Ю. Найда при укладанні словників мов американських індіанців на тих чи інших етапах не обходиться без опитування інформантів [3:58]. Але для іспансько-української мовної пари ця проблема не на часі: є і освічені білінгви-філологи, є і опубліковані словники, які, хоч і замалі за обсягом, охоплюють переважну частину понять базисного рівня.

Значно гостріше проблема пошуку еквівалентів постає для периферійних понять. Немає потреби доводити, що будь-якій білінгв

або будь-який елементарний перекладний словник розв'яже проблему перекладу такого поняття базисного рівня як “праця / працювати”. В той самий час, периферійні компоненти поля “праця”, виражені такими сполученнями як “кропітка праця”, “премія за результатами року”, “Сизифова праця”, “трудова книжка” в повній мірі словниками не розв'язуються, а білінгви, якщо й набирають відносно повний відповідний вокабуляр у двох мовах, йдуть до цього довгим емпірично-тернистим шляхом. Словники ж, як уже сказано, часто не охоплюють периферію, а пошук перекладного еквівалента шляхом опитування інформантів метамовним описом нам видається малопродуктивним, адже, окрім загальних мовних знань, необхідна обізнаність з предметною галуззю, і при спілкуванні з іноземцями ми неминуче наштовхуємось на відповідь: “Не знаю, як це сказати, таке поняття мені не зустрічалося”; носій мови, на відміну від словника, може забути, переплутати.

З огляду на сказане, для периферійних одиниць лексико-семантичних полів доцільною і необхідною є розробка комплексної методики встановлення еквівалентних відношень у двомовних перекладних словниках. Не маючи намір подужати таку багатопланову проблему в рамках статті, наразі ставимо за мету описати один з можливих методів, а саме, дистрибутивний аналіз за допомогою згенерованого комп'ютером словника-конкордансу, який, хоч і не вичерпно, розв'язує проблему встановлення еквівалентних відношень між значною частиною лексичних одиниць різних мов.

Метод дистрибуції, створений у рамках структуралізму, не позбавлений окремих недоліків і не є строгим. Він полягає у встановленні ознак і функцій властивостей певних мовних одиниць залежно від їхнього оточення [7:122]. Однак слід визнати, що в поєднанні з кількісними підрахунками його достовірність підвищується [8:138]. Саме тому можна стверджувати, що в час бурхливого розвитку комп'ютерних знарядь обробки мовного матеріалу цей метод переживає друге народження. Одним з яскравих підтверджень цього слід вважати успішне дослідження А. Шайкевича з застосуванням дистрибутивно-статистичного аналізу [7:123]. Оскільки згаданий метод спрямований на вивчення мовних одиниць у зв'язку з їхнім оточенням, саме в ньому ми вбачаємо потужний інструмент для дослідження усталених лексичних сполучень.

Фактичним матеріалом дослідження обрано одне зі складних з погляду перекладу лексико-семантичних полів – “освіта”. Складність посилюється тим, що в іспанській мові ядро поля представлено щонайменше п’ятьма синонімами, межа між якими часто виявляється дифузною: “enseñanza”, “educación”, “formación”, “instrucción” і “estudios”. Подальший аналіз показує, що не менш важливим конститuentом поля також є іменник “carrera”. В українській мові їм відповідає базисне поняття, виражене частково синонімічними лексемами “освіта” і “навчання”. Значну близькість між зазначеними синонімами красномовно ілюструє, наприклад, можливість легкої взаємозамінюваності цих іменників у сполученні “освітній рівень”. Згідно з даними конкордансу, згенерованого на основі корпусу сучасної іспанської мови *CREA (Corpus de Referencia del Español Actual)* Іспанської королівської академії обсягом у 410 мільйонів слововживань, у поєднанні з іменником “nivel” можливі й узуальні всі п’ять сполучень: “nivel de formación” (48 вживань), “nivel de educación” (103 вживання), “nivel de instrucción” (49 вживань), “nivel de enseñanza” (34 вживань), “nivel de estudios” (98 вживання), “nivel de carrera” (3 вживання) [9]. Аналогічно можуть взаємно заміщуватись усі шість іменників у сполученні з прикметником “universitario”, як видно з таблиці 2. Як показав Корпус текстів української мови лабораторії комп’ютерної лінгвістики КНУ імені Тараса Шевченка обсягом у 3 мільйони слововживань, “рівень освіти” має 11 вживань, “рівень навчання” – 2 [10]. Таким чином, теоретично можлива взаємна перекладність цих сполучень в обох напрямках.

Натомість при перекладі інших словосполучень та термінів допускається вживання лише окремих з них. Наприклад, “formación económica”, “carrera económica” вживаються на позначення економічної освіти, а “estudios económicos” означає і “економічну освіту”, і “дослідження в галузі економіки”; “educación física” – усталений термін на позначення “фізкультури”, а “formación física” вживається і на позначення освіти з фізики, і фізичного розвитку; “educación económica”, “formación económica” (економічна освіта).

Таким чином, і при перекладі з української на іспанську, і при укладанні великого українсько-іспанського словника ми неминуче зіштовхуємось з проблемою вибору одного з шести вказаних синонімів. Це вживання не в останню чергу буде детерміноване лексико-семантичною сполучуваністю слова, його правою та лівою

дистрибуцією. Тому наявність у словнику еквівалентів не лише для одиниць ядра поля, а й для термінологічних та інших усталених сполучуваностей значно допоможуть в оперативному пошуку необхідного варіанта перекладу.

Оскільки такі сполучення передають ті чи інші відтінки, нюанси, зв'язки ядерної одиниці, саме вони охоплюють значну частину периферії. Крім того, на відміну від окремих слів, сполучення характеризуються значно меншою багатозначністю, оскільки в ньому ліва або права дистрибуція ключового слова часто детермінують той лексико-семантичний варіант, у якому воно вжито. Приміром, аналізуючи в такий спосіб дистрибуцію терміна “estudios”, яке може означати і “дослідження”, і “освіта”, переконуємось, що тільки останній лексико-семантичний варіант реалізується в сполученні “cursar los estudios” (у корпусі *CREA* Іспанської королівської академії 6 таких уживань). Хоча подекуди детермінант і не знімає багатозначності. Так, “realizar los estudios” у третині випадків означає “отримати освіту”, а в решті – “провести дослідження”. Саме тому для успішного використання дистрибутивного аналізу, в якому враховується лише механічне поєднання мовних одиниць, без урахування їхнього змісту та причин цього поєднання, необхідне залучення контекстуального смислового аналізу та кількісних даних.

Таким чином, метод дистрибуції при аналізі лексичного значення усталених сполучень не позбавлений недоліків і, безперечно, він даватиме й певний непоказовий і негативний матеріал. Але оскільки левову частину роботи виконують комп'ютери, негативні результати не призведуть до витрат дорогоцінного робочого часу та людських зусиль. Коротко опишемо перебіг дослідження.

Як уже сказано, перевага комп'ютерного застосування методу дистрибуції полягає в можливості його органічного поєднання з кількісною обробкою. Тому наразі була порахована частота вживання досліджуваних синонімів та здійснена інтерпретація частотних даних. Зокрема, в усьому корпусі абсолютна частота вживання досліджуваних конститuentів лексико-семантичного поля “освіта” була такою:

<i>formación:</i>	17 730
<i>educación:</i>	18 282
<i>instrucción:</i>	2 706
<i>estudios:</i>	20 769
<i>enseñanza:</i>	7 843

carrera: 17 660

Корпус *CREA* дозволяє звузити коло пошуку за тематичним критерієм, тому ми також проаналізували абсолютну частоту вживання цих слів у текстах спеціалізованої підмови освіти. Пошук дав такі результати:

estudios: 1 247

educacion: 5 209

instrucción: 69

formación: 1 366

enseñanza: 2 241

carrera: 415

Цікаво відзначити, що у спеціалізованій підмові термін “*estudios*” характеризує ідіостиль двох авторів, а порівняно з іншими він вживається відносно рідко (займає четверте місце за частотою, і то, навіть більша частина, 1 053 випадки у спеціалізованій підмові припадають на пресу), хоча взагалі в корпусі є лідером цього поля. Отже, виходячи з такої контрастної зміни частотності, можна сказати, що “*estudios*” характеризує мову нефаківців галузі освіти.

Наступним етапом було виділення типової дистрибуції іменників “*enseñanza*”, “*educación*”, “*formación*”, “*instrucción*”, “*estudios*” і “*carrera*”. Для цього фіксуємо перші 100 випадків вживання цих слів у конкордансі *CREA* Іспанської королівської академії. Іменник “*instrucción*”, наприклад, показав таку дистрибуцію (для економії обсягу результати наводимо лише частково):

juzgado de instrucción,
instruccion cívica,
instruccion básica,
revista de instrucción,
tres meses de intrucción,
mujeres sin ninguna instrucción,
recibir instrucción,
instrucción primaria,
su instrucción o sus modales,
práctica de instrucción militar,
director de la instrucción del Estado Mayor,
tradicional instrucción naval,
viaje de instrucción,
proyectiles de instrucción,
viceseretario de la instrucción pública,
durante la instrucción militar,

recibir instrucción,
centros de instrucción,
una acertada instrucción escolar,
instrucción de los alumnos,
instrucción de los estudiantes,
derecho a la instrucción,
consecuente instrucción premilitar,
grado de instrucción,
el nivel de la instrucción,
carecía de la instrucción necesaria,
sistema de instrucción,
reformar la instrucción,
en materia de instrucción, тощо.

Наступним етапом є відбір тих сполучень та термінів, які становлять інтерес з погляду перекладу. Теоретично усі сполучення важливі, але з огляду практичних потреб словника, який звичайно ж не є неосязним, вибираємо ті, які є найтиповішими, а також ті, що, з погляду дослідника, можуть становити явний інтерес (наприклад, де проявляються якась міжмовна асиметрія). Виявити останні можна шляхом калькування і контекстуальної перевірки, чи відповідає припущене значення контексту. Після цього порівнюємо частоту вживання аналогічних сполучуваностей для інших іменників лексико-семантичного поля і отримуємо такі таблиці:

Таблиця 1

**Ліва дистрибуція конститuentів лексико-семантичного поля
 “освіта”**

	formación	educación	instrucción n	enseñanz a	estudios	carrera
abandon* la	4	1	1	4	50	42
brind* la ~	1	1	1	1	0	3
centr* de ~	135	133	48	290	165	0
curs* la/los~	1	4	0	3	19	23
dar ~	5	7	2	1	0	1
dar la/los ~	0	0	1	1	0	1
dar una/unos~	14	9	3	0	0	3
dejar la ~	0	0	0	2	14	7
dejó la ~	0	0	0	0	7	11
graduado* en ~	0	0	0	0	1	0
hacer la	2	5	1	3	14	15
interrump* la/los ~	0	1	0	0	9	4
interrump* su/sus ~	0	0	0	0	24	5
Ministerio de ~	0	1783	23	1	0	0
nivel* de ~	59	147	57	75	102	3

obtener ~	0	3	0	1	0	0
obtener la/los ~	1	0	0	0	0	0
obtuvo ~	0	0	0	0	0	0
obtuvo la ~	0	0	0	0	0	0
program* de ~	155	172	9	44	71	1
proyectil* de ~	0	0	1	0	0	0
realiz* ~	3	5	0	0	430	3
realiz* la/los ~	7	6	3	1	69	4
realiz* una ~	3	0	0	1	5	18
reconocimient* de la ~	1	0	0	2	2	0
reform* la/los	0	5	1	3	0	2
segunda	8	1	4	50	0	46
título* de ~	3	4	4	1	3	0

Таблиця 2

**Права дистрибуція конститuentів лексико-семантичного поля
“освіта”**

	formació n	educación	instrucción	enseñanz a	estudios	carrera
~ a distancia	15	72	1	30	2	1
~ completa/os	10	6	2	1	12	5
~ cristiana/os	22	20	1	5	0	0
~ de adultos	11	42	0	3	0	0
~ de calidad	4	55	0	12	7	0
~ de los maestros	7	0	0	0	0	0
~ de maestros	20	0	1	0	0	0
~ de pago	0	1	0	0	0	0
~ del profesorado	110	0	0	0	1	0
~ económica	17	0	0	0	50	1
~ en ciencia	1	1	1	1	0	0
~ especializada	36	8	2	2	23	0
~ estatal/es	0	2	0	27	3	1
~ física	10	425	0	0	4	0
~ gratis	0	2	0	0	1	0
~ gratuita/os	2	32	1	20	3	1
~ infantil/es	4	64	0	3	0	0
~ militar/es	51	15	49	14	19	154
~ pagada/os	1	2	0	0	1	0
~ privada/os	0	44	0	138	5	1
~ profesional	609	17	2	14	23	319
~ pública/os	3	201	42	169	2	13
~ reglada/os	2	5	0	9	4	0
~ universitaria	80	109	3	93	329	119
~ superior	19	1310	9	454	245	1
~ secundaria	4	118	1	130	67	1

Зазначимо, що зірочки після основи слова означають будь-яке можливе сполучення літер, яке для флективних мов найчастіше відповідає флексії.

Окремі випадки, хоч і відрізняються разючою нерівномірністю числових даних, ми не враховували через явну зміну лексико-семантичного варіанта слова в контексті детермінанта. Так, “formación del profesorado” – не навчання, а “підготовка професорсько-викладацького складу”, “instrucción militar” – переважно не військове навчання, а “військова інструкція”. Такі знахідки лексикографу теж у необхідно брати до уваги, але в нашому випадку вони виходять за рамки об’єкта дослідження.

Таким чином, на підставі порівняння частоти вживання тих чи інших сполучень, відсіювання рідких та відсутніх комбінацій, контекстуального аналізу можливої двозначності сполучення, можна зробити підсумок. Це “сухий залишок” корисного матеріалу, який, хоч і незначний на тлі опрацьованої сукупності даних, здатний значно поповнити словникові статті “навчання” і “освіта” в українсько-іспанському словнику:

безкоштовне навчання	educación gratuita, enseñanza gratuita
вища освіта	educación superior, enseñanza superior, estudios superiores
військова освіта	formación militar
дати освіту	dar (una) educación, dar (una) formación
державна освіта	educación pública, enseñanza pública
дистанційне навчання	educación a distancia, enseñanza a distancia
дитяча освіта	educación infantil
друга освіта	segunda enseñanza, segunda formación
економічна освіта	formación económica, carrera económica
кинути навчання	abandonar la carrera, abandonar los estudios, dejar los estudios
Міністерство освіти	Ministerio de educación
освіта для дорослих	formación de adultos, educación de adultos
отримати освіту	cursar los estudios, obtener la educación, realizar los estudios
перервати навчання	interrumpir los estudios
приватна освіта	enseñanza privada, educación privada
професійна освіта	educación profesional, enseñanza profesional; formación profesional
реформувати освіту	reformar la educación, reformar la enseñanza
середня освіта	educación secundaria, enseñanza secundaria, estudios

secundarios

спеціалізована освіта formación especializada

християнська освіта formación cristiana, educación cristiana

Отже, при укладанні двомовних перекладних словників необхідні принципово відмінні підходи до ядерних та периферійних понять лексико-семантичних полів. Для українсько-іспанської та інших мовних пар, у яких уже укладено базові двомовні словники, найактуальнішою проблемою є розширення кількості еквівалентів для одиниць периферії того чи іншого лексико-семантичного поля за допомогою комплексної методики.

Оскільки значна частина периферії представлена усталеними сполученнями базисних лексем з іншими лексичними одиницями, а сполучення характеризуються значно нижчою багатозначністю, ніж окремі слова, їх відносно легко досліджувати за допомогою методу дистрибуції. Дистрибутивний аналіз за даними конкордансу, в поєднанні з кількісною обробкою, калькуванням та смисловим контекстуальним аналізом, дозволяє виявити узуальні сполучення й перевірити їхнє значення. Особливо продуктивною й доречною така методика видається для лексико-семантичних полів, ядро яких представлено декількома близькими синонімами. **Перспективи дослідження.** Описану методику доцільно застосувати й до інших лексико-семантичних полів з аналогічною міжмовною асиметрією. Скажімо, українському терміну “суд” відповідають “corte”, “juzgado”, “tribunal” на позначення судових органів та “juicio” на позначення судового процесу.

ЛІТЕРАТУРА

1. Берков В. П. Двухязычная лексикография / Валерий Павлович Берков. – М. : Транзиткнига, 2004. – 236 с.
2. Щерба Л. В. Опыт общей теории лексикографии / Лев Владимирович Щерба // Языковая система и речевая деятельность. – Л. : Наука, 1974. – С. 290–297.
3. Найда Ю. Анализ значения слов и составление словарей / Юджен Найда // Новое в зарубежной лингвистике. – 1962. – Вып. 2. – С. 45–71.
4. Найда Ю. Наука переводить / Юджен Найда // Вопросы языкознания. – 1970. – № 4. – С. 3–14.
5. Апресян Ю. Д. Лексическая семантика / Юрий Дереникович Апресян – М. : Наука, 1974. – 367 с.
6. Шафриков С. Г. Типология лексических систем и лексико-семантических универсалий / С. Г. Шафриков. – Уфа : Издание Башкирск. ун-та, 2004. – 80 с.
7. Селіванова О. О. Сучасна лінгвістика: термінологічна енциклопедія / Олена Олександрівна Селіванова. – Полтава : Довкілля-К, 2006. – 716 с.
8. Лингвистический энциклопедический словарь / гл. ред. В. Н. Ярцева. – [2-е изд., доп.]. – М. : Большая Российская энциклопедия, 2002. – 709 с.

ДЖЕРЕЛА ІЛЮСТРАТИВНОГО МАТЕРІАЛУ

9. *Corpus de Referencia del Español Actual de la Real Academia Española* [Електронний ресурс]. – Режим доступу : <http://corpus.rae.es/creanet.html>
10. *Corpus* текстів української мови [Електронний ресурс]. – Режим доступу : <http://www.mova.info/corpus.aspx?11=209>