

ТЮРКСЬКІ МОВИ

УДК 811.512.161:811.161:[004.4+003.034]:341.76
DOI <https://doi.org/10.32782/2710-4656/2024.5.1/35>

Челік К. С.

Таврійський національний університет імені В. І. Вернадського

ЕКСТРАКЦІЯ СТІЙКИХ СЛОВЕСНИХ КОМПЛЕКСІВ З ТЕКСТІВ СУЧАСНОГО ТУРЕЦЬКО-УКРАЇНСЬКОГО ДИПЛОМАТИЧНОГО ДИСКУРСУ

Стаття присвячена дослідженню турецькомовних текстів сучасного турецько-українського дипломатичного дискурсу з метою екстракції актуальних стійких словесних комплексів. Оскільки досі немає єдиного визначення поняття «стійкий словесний комплекс», в роботі окреслено існуючі підходи та представлено власне розуміння стійкого словесного комплексу, як семантично цілісних стійких сполук слів, що природно виникли в мовленні й можуть виконувати номінативну, комунікативну чи експресивну функцію. Далі розглянуто місце стійких словесних комплексів у дипломатичному дискурсі та обґрунтовано важливість укладення й регулярного оновлення їх реєстру з огляду на безперервну зміну актуальності лексики міжнародних відносин. В практичній частині роботи описано алгоритм здійсненої екстракції актуальних стійких словесних комплексів турецько-українського дипломатичного дискурсу на матеріалі турецькомовних текстів, оприлюднених в період з 1.01.2023 по 11.07.2024 на офіційних веб-сайтах Міністерства закордонних справ Туреччини, Представництва Президента Турецької Республіки, Посольства Турецької Республіки в Україні та Посольства України в Турецькій Республіці. Процедура, проведена з залученням програмного забезпечення Sketch Engine та AntConc 4.3.1 складалася з етапів формування колекції текстів; самостійного пошуку та фіксації частотності стійких словесних комплексів у сформованій колекції; укладення корпусу текстів та проведення ідентифікації колокацій в корпусі за допомогою корпусних менеджерів; екстракція стійких словесних комплексів з отриманого переліку колокацій; валідація отриманих результатів та формування переліку актуальних стійких словесних комплексів турецько-українського дипломатичного дискурсу. Окрім досягнення основної мети, в процесі дослідження було визначено основні переваги й недоліки використання корпусних менеджерів для ідентифікації колокацій в турецькомовних текстах; виявлено деяку невідповідність транслітерації українських власних назв оригіналам в опрацьованому матеріалі; виділено тексти, які можуть бути використані як навчальні матеріали.

Ключові слова: турецько-український дипломатичний дискурс, стійкий словесний комплекс, стійкі словесні комплекси турецького дипломатичного дискурсу, обробка турецької мови в Sketch Engine, обробка турецької мови в AntConc.

Постановка проблеми. Стійкі словесні комплекси є ключовою лексикою дипломатичного дискурсу, основна проблема перекладу яких полягає у перманентній зміні їх актуальності: з розвитком глобалізаційних процесів, окремі поняття виходять з ужитку, натомість народжуються нові, що відображається у мові міжнародних відносин виникненням і асиміляцією з інших інституційних дискурсів термінів, усталених виразів, загальних і власних назв. Вирішенням цієї проблеми стане формування та регулярне оновлення реєстру стій-

ких словесних комплексів дипломатичного дискурсу, призначення якого ми вбачаємо в подальших лексичних і лексикографічних працях та використанні на практичних заняттях дисциплін з перекладу.

Аналіз останніх досліджень і публікацій. В останні десятиріччя дослідженню лексичного складу саме турецького дипломатичного дискурсу було присвячено дві роботи О. Локоти, які описують кількісно-якісні та формально-структурні характеристики термінології [10; 11], а також ряд

кваліфікаційних студентських робіт, що вказує на зростання інтересу до цієї теми. Темі міжкультурної взаємодії в контексті турецько-українського перекладу відведено місце у посібнику Г. Спотар-Аяр «Культура усного та писемного мовлення: діловий та академічний дискурс (турецька мова)» [13]. Інші українські дослідники – В. Пілик, О. Польова, О. Шулькевич дотично торкалися турецької дипломатичної лексики в дослідженнях політичного та публіцистичного дискурсів турецької мови. Серед турецьких науковців спроба дослідження мови міжнародних відносин в перекладацькому аспекті була зроблена Г. Унсал у статті «Мова дипломатії та переклад» [36], фундаментальна лексикографічна праця «Словник міжнародних відносин і дипломатії. Англійська – французька – турецька» належить А. Дагу [21]. Стійкі словесні комплекси сучасного турецько-українського дипломатичного дискурсу досі не були предметом наукового вивчення, проте самі стійкі сполуки слів почали привертати увагу науковців ще задовго до того, як сформувалася наука «фразеологія» і дотепер залишаються предметом наукових студій. Досліджуючи фразеологію, неможливо оминати увагою підручник Я. Барана та М. Зимомрі «Теоретичні основи фразеології» [2], дисертацію Я. Барана «Фразеологія у системі мови» [3] та дисертацію Л. Корнієнко «Українська фразеологія другої половини ХХ – початку ХХІ століть: тлумачні та перекладні словники» [9] у яких зібрані та проаналізовані підходи до вивчення, тлумачення, класифікації стійких словосполучень, починаючи від Ш. Баллі та О. Потебні до М. Жуйкової та Ж. Краснобаєвої-Чорної. Вагомий внесок у визначення меж предметів вивчення фразеології здійснило дисертаційне дослідження Ж. Колоїз «Структурно-функціональні різновиди усталених конструкцій у мові публіцистики Сергія-Єфремова», серед основних завдань якої – з'ясування понять «фразеологізм» та «усталена конструкція»; диференціація усталених словесних комплексів фразеологічного типу, усталених словесних комплексів нефразеологічного типу, усталених фразеологізованих схем [8]. У турецькому мовознавстві поняття фразеологізм розглядають, здебільшого, в широкому його розумінні: Д. Аксан, О. Аксой, С. Булут, Х. Гьокдаї, Ш. Ельчін, Ч. Ерол, В. Хатібоглу, та інші, відтак в останні десятиріччя увага турецьких дослідників змістилась від теоретичного аналізу до методики викладання та перекладу фразеології: М. Джанбулат, Ф. Йилмаз, Б. Сарикая, Н. Шімшек, та інші.

Постановка завдання. Метою дослідження є екстракція та формування реєстру стійких сло-

весних комплексів сучасного турецько-українського дипломатичного дискурсу.

Виклад основного матеріалу. У мовознавчій науці існують різні погляди на сутність поняття «стійкий словесний комплекс» (далі – ССК) і немає одностайної думки щодо того, чи можна вважати усі типи ССК елементами фразеологічної системи [2, с. 24], хоча їх завжди розглядали в контексті фразеології. Поняття «фразеологізм» мовознавці теж тлумачать по-різному, проте усі вказують на такі його риси як стійкість, наприклад: «відтворюване, цілісне за значенням, стійке за складом і структурою словосполучення» у В.-Ужченко [15, с. 7] та семантичну цілісність, наприклад: «стійкі словосполучення двох і більше слів, що створюють семантичну цілісність» у Л.-Скрипник [12, с. 7]. ССК нефразеологічного типу, також характеризуються стійкістю та семантичною цілісністю, проте на відміну від фразеологізмів, завжди є граматично нерозкладними [1, с. 81]. До них відносять титули, посади, географічні назви, назви країн та організацій, лексично нечленні загальні складні назви та термінологію різних галузей науки, техніки, промисловості, сільського господарства, культури, мистецтва [2, с. 9; 14, с. 92]. Відтак, ССК нефразеологічного типу «називають предмети, явища, процеси, а усталені словесні комплекси фразеологічного типу їх насамперед характеризують і тільки через таку характеристику називають, мотивуючи водночас висловлювану оцінку» [8, с. 6]. Разом з тим, дехто з науковців говорить, що окрім номінативної, ССК можуть виконувати і номінативно-експресивну, експресивно-номінативну або експресивну функцію [2, с. 122]. Окрім семантичної цілісності, аналіз літератури вказує на ще одну спільну рису ССК різних типів – їх природне виникнення в мовленні, відтак, етикетні конструкції та протокольні формули теж належать до класу ССК, проте виконують комунікативну функцію. Таким чином, до ССК ми зараховуємо усі семантично цілісні стійкі сполуки слів, що природно виникли в мовленні й можуть виконувати номінативну, комунікативну чи експресивну функції.

Дипломатичний дискурс (далі – ДД) містить велику кількість як дипломатичних, так і асимільованих внаслідок інституційної взаємодії ССК, які в мові перекладу відтворюються конкретними еквівалентами і ніяк інакше [4; 13; 35]. «Чи варто сумніватися, що в епоху гібридних війн за сенси і наративи навіть необережно названа “кордоном” адміністративна межа між Кримом та материковою Україною може “обмежити” укра-

їнський суверенітет Херсонською областю, тоді як необачно згадана “анексія” півострова замість його “тимчасової окупації” дозволить черговому “Спутнику” заявити про “визнання” приєднання Криму до Росії як доконаного факту» [4] зазначає перекладач, дипломат, нині старша дослідниця Atlantic Council і Центру досліджень сучасної Туреччини Карлтонського університету – Євгенія-Габер. Відтак, компетентність дипломатичного перекладача полягає не в умінні знаходити перекладацькі рішення, а в досконалому знанні відповідників актуальних понять та виразів ДД, що зумовлює необхідність постійного відстеження оновлень його лексичного складу.

Сучасним способом ідентифікації та/або статистичного аналізу об’єктів дослідження в базі даних є автоматизований пошук, що у лінгвістичних дослідженнях здійснюється за допомогою корпусних менеджерів – пошукових систем, які дозволяють знайти в текстових корпусах основи слів, словоформи, колокації (словоформи, які часто вживаються разом [18]), відслідкувати їх використання в контексті, отримати статистичні результати, тощо. Корпус в мовознавстві, загалом, визначають як збережену на машинному носії вибірку текстів, що може містити лінгвістичну розмітку, основними ознаками якої є автентичність та репрезентативність [5; 6; 7; 24; 28], однак з швидким технологічним розвитком поняття текстового корпусу також змінюється [26].

Для екстракції з корпусу текстів ССК за допомогою програмного забезпечення необхідні лише дві функції: ідентифікація колокацій та статистичні дані їх частотності, якими оснащені усі корпусні менеджери. Завдяки вільному доступу й підтримці багатьох мов, серед корпусних менеджерів найпопулярнішими є веб-ресурс Sketch Engine та застосунок AntConc. Їх загальні характеристики, як, наприклад: можливість створення користувачького корпусу, використання декількох статистичних метрик, тощо – мало різняться між собою [17; 37]. Вагомою відмінністю Sketch-Engine від останньої на час проведення дослідження версії AntConc 4.3.1. є наявність в базі Sketch-Engine 48-ми готових до використання корпусів (серед яких повний український та основний турецький корпус), коли AntConc 4.3.1. містить корпуси лише англійської мови. Однак цей факт не впливає на якість ідентифікації колокацій в користувачькому корпусі, оскільки технологія порівняння програмою досліджуваного корпусу з еталонним, яка вважається одним із найпотужніших видів аналізу на сьогодні [19], використовується зараз лише для

пошуку ключових слів та термінів [17; 27]. Відтак, з метою визначення оптимального інструменту для регулярного оновлення реєстру ССК турецько-українського ДД, ми вирішили протестувати обидві програми. Результати автоматизованого пошуку потребуватимуть постредагування, оскільки колокація і ССК не є тотожними поняттями. Для підтвердження достовірності отриманих даних, перелік ССК, отриманий напівавтоматизованим способом, буде порівняно з переліком ССК, виділених вручну.

Формування колекції текстів та самостійне виокремлення ССК. Актуальність ССК ДД безперервно змінюються відповідно до соціально-політичної ситуації в регіоні, в Україні такі кардинальні зміни востаннє відбулися з початком повномасштабного російського вторгнення. Аналізуючи офіційні джерела, а саме: офіційні веб-сайти Міністерства закордонних справ Туреччини [31], Інтернет-представництва Президента Турецької Республіки [29], Посольства Турецької республіки в Україні [30] та Посольства України в Турецькій Республіці [34], зміни у лексичному складі було зафіксовано з початком 2023 року, тому до вибірки було включено 33 тексти, оприлюднені в період з 1.01.2023 по 11.07.2024, темою яких є двосторонні відносини України та Турецької Республіки. До них належать: заяви та коментарі Міністерства закордонних справ Туреччини, уривки з інтерв’ю Міністра закордонних справ Туреччини, пресконференції Президента Туреччини, коментарі прессекретаря Президента Туреччини, уривок з програми діяльності Міністерства Закордонних справ Туреччини на 2024 рік, висвітлення Міністерством останніх подій, інформація для преси. Тексти були скопійовані та збережені на комп’ютері окремими файлами в форматі .txt з кодуванням Mac OS. Вони не містили переносів, таблиць чи інших вбудованих графічних елементів, які потребують редагування для коректного здійснення корпусним менеджером токенизації. Під час формування колекції текстів, вручну було знайдено та зафіксовано частотність 60 ССК (окрім географічних назв та назв країн), які зустрілися більш ніж в одному тексті (табл. 1).

Створення корпусу та проведення процедури ідентифікації колокацій у веб-ресурсі Sketch-Engine. Для створення користувачького корпусу в програмі Sketch Engine потрібно обрати функцію “NEW CORPUS” та завантажити файли у будь-якому з .csv, .doc, .docx, .htm, .html, .ods, .pdf, .tar.bz2, .tar.gz, .tei, .tgz, .tmx, .txt, .vert, .xlf, .xliff, .xml, .zip форматів, інші необхідні дії вико-

нуються програмою автоматично. В Sketch Engine є два інструменти, що ідентифікують колокації без заданого об'єкту пошуку: “Word Sketch As a List” та “N-grams”. Перший – видає перелік колокацій, відображаючи граматичні зв'язки між словами [38], другий – перелік лише самих колокацій, але має більше опцій налаштування пошуку: “N-gram length” – визначення кількості токенів колокації (“2-grams”, “4-grams”, “2 – 6-grams”, тощо); “Attribute” – вибір атрибута, за яким будуть згенеровані колокації (слово/основа/тег/частина мови/лема+частина мови); “Frequency” – частотність колокації, “Include nonwords” – врахування лексем, які не починаються з букви алфавіту; “Nest n-grams” – згрупування в списку результатів колокацій з їхніми суб-колокаціями, “A=a” – врахування або неврахування регістру, “Exclude these words” – виключення з діапазону пошуку заданих основ/слів/словосполучень [25]. В опціях налаштувань немає регуляції мінімальної кількості текстів у яких зустрічається колокація, що ускладнить виокремлення ССК, але ці дані вказані в таблиці результатів у графі “Document frequency”. Скориставшись інструментом “N-grams”, ми налаштували пошук колокацій за основою – “stem” (атрибут рекомендований для роботи з аглютинативними мовами), мінімальну частоту вживання – два рази, довжину n-gram (колокації) – від 2 до 4 токенів, оскільки найдовший ССК виокремлений вручну – чотирислівний (табл. 1). За цими параметрами Sketch-Engine виявив у заданому корпусі 1345 колокацій (рис. 1). У таблиці результатів (рис. 1) “Frequency” – частотність словосполучення, “DOCF” – кількість текстів у яких воно зустрічається.

Попри те, що ми використовували атрибут “stem”, в результатах все ще бачимо проблему токенизації власних назв: *nın toprak bütün, nın yeniden*, тощо; а також низку слів, які малоімовірно можуть входити до складу ССК ДД і не входять до ССК, виділених нами вручну: *Recep, bu, Sayın*, тощо (рис. 1). Для конкретизації результатів пошуку ми застосували функцію “Exclude these words”. При роботі з невеликим корпусом важко назвати раціональним час, витрачений на аналіз результатів пошуку колокацій та виокремлення шумових слів/токенів, проте в нашому випадку – коли ми плануємо регулярне оновлення отриманого реєстру, список стоп-токенів полегшить роботу в майбутньому. До переліку ми внесли відмінкові закінчення *-i, -ı, -uı, -uü, -ın, -in, -un, ün, -nın, -nın, -nun, -nün, -a, -e, -ya, -ye, -na, -ne, -da, -de, -nda, -nde, -dan, -den, -ndan, -nde, -ile*, які Sketch-Engine ідентифікує як окремі токени у випадках їх написання через апостроф, скорочення *No, Hk*; звертання *Sayın*; питальні частки *mi, mi, mi, mü*; сполучники *ki, yanı sıra, ve, ile, ama, de, da*; післяйменники *üzere, şekilde, dahil, için, itibarıyla*; вказівні займенники *bu, şu, o*; прикметники *ilgili, türlü, yönelik*; іменники *Cumhurbaşkanı, Cumhuriyet, Federasyon*; вставні слова *dolayısıyla, tabii, özellikle yani*; основу *et-/ed-*, щоб виключити складені дієслова та імена людей: *Recep, Tayyip, Erdoğan, Volodymyr, Zelenskyu, Volodimir, Zelenski, Zelenskiy, Olena, Zelenska, Ömer, Bolat, Yevhen, Perebiynis, Burak, Akçapar, Hakan, Fidan, Dmitro, Kuleba, Putin, Neriman, Celal, Mustafa, Kırımoglu*. Таким чином ми знизили кількість ідентифікованих колокацій з 1345 до 716 (рис. 2).

N-gram	Frequency ↑	DOCF ↑	N-gram	Frequency ↑	DOCF ↑
1 devam et	37	16 ...	11 Ukrayna ve	11	8 ...
2 dışişleri bakan	29	7 ...	12 nın yeniden	11	6 ...
3 cumhurbaşkanı Erdoğan	21	8 ...	13 ulus hukuk	11	5 ...
4 devlet başkan	19	9 ...	14 yeniden inşa	11	6 ...
5 bu konu	16	4 ...	15 bakan yardım	10	3 ...
6 kırım Tatar	16	9 ...	16 sayın Putin	10	2 ...
7 Türkiye cumhuriyet	15	10 ...	17 Ukrayna dışişleri bakan	10	5 ...
8 sayını bakan	15	9 ...	18 cumhurbaşkanı Recep Tayyip	9	6 ...
9 teşekkür et	14	6 ...	19 her zaman	9	7 ...
10 başış zirve	14	5 ...	20 nın toprak bütün	9	6 ...
11 toprak bütün	14	10 ...	21 dışişleri bakan yardım	9	2 ...
12 ve Ukrayna	13	12 ...	22 bakan Hakan	9	4 ...
13 Ukrayna devlet	13	9 ...	23 ve bu	9	8 ...

Рис. 1. Результати ідентифікації 2–4 grams за атрибутом stem в програмі Sketch-Engine

N-gram	Frequency	DOCF	N-gram	Frequency	DOCF	N-gram	Frequency	DOCF
dişişleri bakan	29	7	dişişleri bakan yardım	9	2	tahli koridor	7	4
devlet başkan	19	9	her zaman	9	7	herhangi bir	6	4
konm Tatar	16	9	el al	8	4	sanip çık	6	1
barış zirve	14	5	tahli girişim	8	5	barış formül	6	6
toprak bütün	14	10	Tatar Türk	8	5	bir an önce	6	5
Ukrayna devlet	13	9	serbest brak	8	5	bağ ol	6	6
bir ara gel	12	6	çaba sarf	8	4	her iki	6	4
son er	12	4	bir şekil	8	6	an önce	6	5
bir ara	12	6	konm Tatar Türk	8	5	telefon görüşme	6	4
ara gel	12	6	bir an	8	7	çaba sür	5	4

Рис. 2. Результати ідентифікації 2–4 grams за атрибутом stem з застосуванням переліку стоп-слів у програмі Sketch-Engine

Під час фіксації статистичних результатів необхідно перевіряти вживання знайдених ССК в контексті за допомогою інструменту “KWIC” (Key Word In Context), оскільки Sketch-Engine часто ідентифікує різні варіанти одного словосполучення як окремі колокації (рис. 3).

Створення корпусу та проведення процедури ідентифікації колокацій в застосунку AntConc 4.3.1. Для створення користувацького корпусу в AntConc потрібно завантажити файли форматів .txt, .docx або .pdf. AntConc пропонує більше опцій для формування корпусу, ніж Sketch-Engine, зокрема “Token definition” – налаштування токенизації. Ми активізували опцію “Apostrophe” в “Custom Punctuation” щоб програма ідентифікувала слова з апострофом (*Ukrayna'nın*, *Türkiye'ye*, *İsviçre'de*, тощо) як один токен.

Для ідентифікації колокацій застосунок має один інструмент – “N-Gram”, в параметрах пошуку якого ми визначили “Minimum frequency” (мінімальну кількість повторень колокації в кор-

пусі) – 2, “Minimum range” (мінімальну кількість текстів, у яких зустрічається колокація) – 2, “Search Query” (пошуковий запит) – “Regex” (регулярні вирази) та “Case” (урахування регістру) (Рис. 4, 5, 6). В 4 версії AntConc також є можливість додати список слів, які не відобразяться в результатах пошуку, за допомогою функції “Hide words in file”. Слова у списку повинні бути прописані в такій формі, у якій вони зустрічаються в корпусі, відмінкові афікси додавати не потрібно.

Результат програма не видає єдиним переліком колокацій, а відображає окремими списками: 2-grams, 3-grams, тощо. За переліченими вище параметрами AntConc виявив 181 двослівну, 46 трислівних, 16 чотирислівних – усього 243 колокації (рис. 4), але в графі результатів відображається загальна кількість ідентифікованих n-gram з урахуванням прихованих стоп-слів. В таблиці результатів у колонці “Freq” вказано частотність словосполучення, у колонці “Range” – кількість текстів у яких воно зустрічається (рис. 4).

QQL [stem=="türm"] [stem=="Tatar"] [stem=="mllr"] [stem... + 3
323.84 per million tokens • 0.033%

Left context KWIC Right context

- doc#4 unun devam ettiğini hatırlattı.</s></s>Türkiye Cumhurbaşkanı ayrıca **Kırım Tatar Millî Meclisi** Birinci Başkan Yardımcısı Nariman Cela'ın ve Rusya tarafından yasa
- doc#7 arımızla birlikte çalışmaya hazırız.</s></s>No: 125, 29 Haziran 2024, **Kırım Tatar Millî Meclisi** Başkan Yardımcısı Neriman Cela'ın Serbest Bırakılması Hk.</s></s>
- doc#7 inarak, daha sonra tutuklanan ve bilahare hapis cezasına çarptırılan **Kırım Tatar Millî Meclisi** Başkan Yardımcısı Neriman Cela'ın esir takası çerçevesinde serbest

QQL [stem=="Tatar"] [stem=="mllr"] + 4
434.45 per million tokens • 0.043%

Left context KWIC Right context

- doc#4 unun devam ettiğini hatırlattı.</s></s>Türkiye Cumhurbaşkanı ayrıca Kırım **Tatar Millî Meclisi** Birinci Başkan Yardımcısı Nariman Cela'ın ve Rusya tarafından yas
- doc#7 arımızla birlikte çalışmaya hazırız.</s></s>No: 125, 29 Haziran 2024, Kırım **Tatar Millî Meclisi** Başkan Yardımcısı Neriman Cela'ın Serbest Bırakılması Hk.</s></s>
- doc#7 inarak, daha sonra tutuklanan ve bilahare hapis cezasına çarptırılan Kırım **Tatar Millî Meclisi** Başkan Yardımcısı Neriman Cela'ın esir takası çerçevesinde serbet
- doc#30 mize, arabalarımıza sahip çıkmaya devam edeceğiz.</s></s>Nitekim Kırım **Tatar millî** marşı nedir?</s></s>"Ant etmiş milletimin yarasını samaya, nasıl olur da

Рис. 3. Словосполучення ідентифіковане програмою Sketch-Engine як дві різні колокації

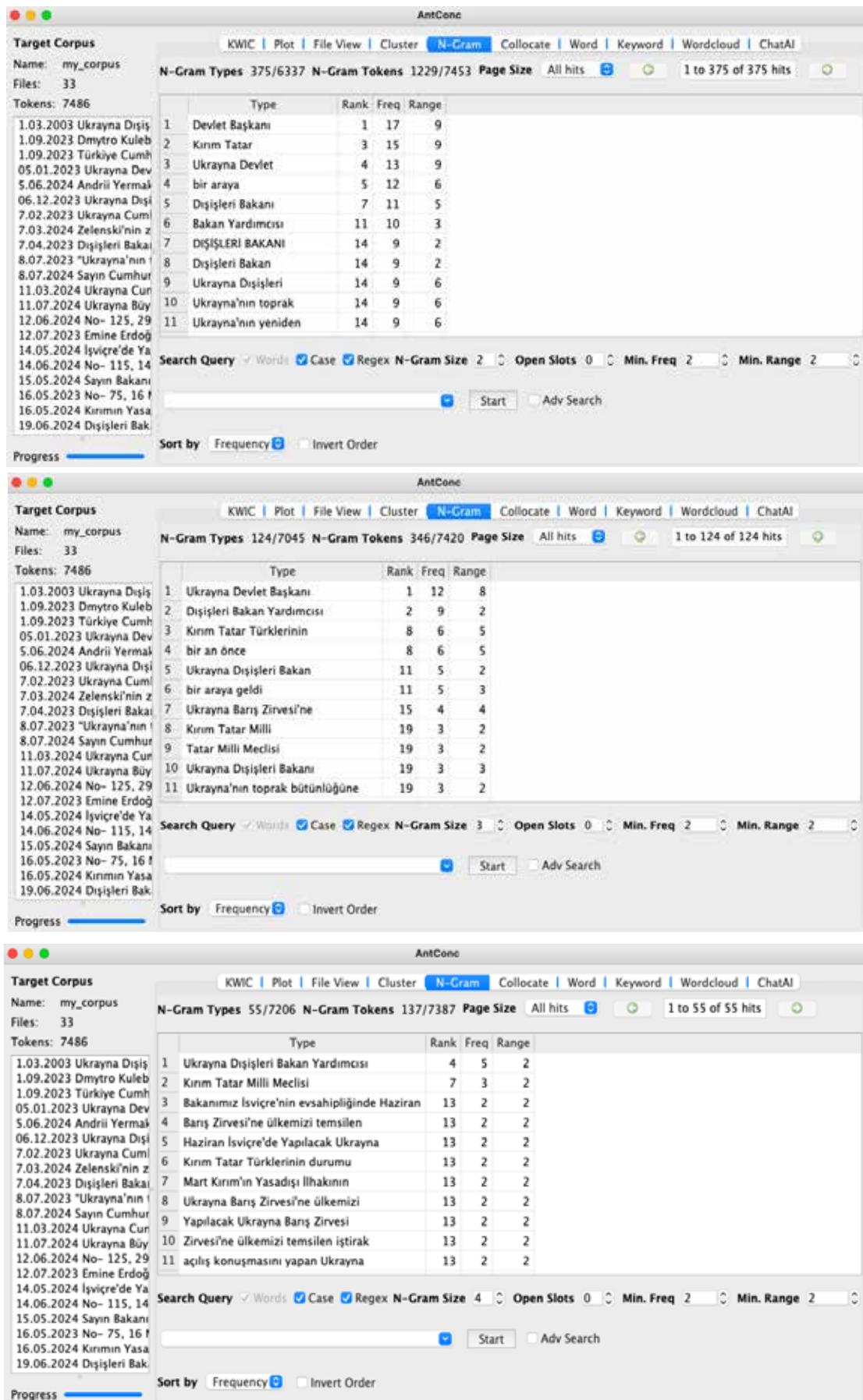


Рис. 4. Результати ідентифікації 4-grams в програмі AntConc

AntConc, так само як і Sketch-Engine, варіації одного й того ж словосполучення ідентифікує як окремі колокації (рис. 5) тому, фіксуючи статистичні дані, потрібно перевіряти кожен випадок вживання за допомогою інструменту KWIC (рис. 5).

Екстракція ССК та валідація результатів. Згідно вище сформульованого нами визначення поняття ССК, з 716 колокацій, зафіксованих корпусним менеджером Sketch-Engine, ми виділили 54, а з 243 колокацій, зафіксованих корпусним менеджером AntConc – 53 ССК. Зведені результати продемонстровані в таблиці 1, де колонка “ч” – частотність, “т” – кількість текстів, у яких вжито ССК (Табл. 1). Бачимо, що жоден корпусний менеджер не ідентифікував як колокації виокремлені нами ССК *siyasi tutuklular/mahkumlar; karşı karşıya kalmak, iş adamı/iş insanı, insansız hava aracı (İHA), Birleşmiş Milletler (BM)*. Статистичні дані подекуди різняться, проте в контексті мети нашого дослідження це не є суттєвим, оскільки вони враховуються лише для визначення актуальності ССК.

В ході роботи було відмічено некоректне написання українських імен та власних назв у досліджуваному матеріалі: *Kiev* [22; 23], *Dmitro* [23], *Andrii* [16], *Volodimir* [35], *Zelenski* [32; 33], попри запроваджену у 2019 році Міністерством закордонних справ України кампанію #CorrectUA

[20]. Також в процесі дослідження було виявлено тексти, які містять значну кількість актуальних ССК: “Andrii-Yermak: “Ukrayna için adil bir barış: Türkiye’nin rolü ve Küresel Barış Zirvesi” [16] та “Dışişleri Bakanı Sayın Hakan-Fidan’ın Ukrayna Dışişleri Bakanı Dmitro-Kuleba ile Ortak Basın Toplantısı” [23] й можуть бути використані в якості навчальних матеріалів для курсу практичного перекладу.

Висновки. Результати дослідження продемонстрували, що стійкі словесні комплекси становлять основу лексичного складу дипломатичного дискурсу, оскільки охоплюють як специфічні поняття, так і типові формулювання. Достовірність отриманих результатів підтверджено застосуванням відразу трьох методів пошуку: самостійного, за допомогою застосування AntConc та з використанням веб-ресурсу Sketch-Engine. Підсумки показали, що обидва корпусні менеджери однаково ефективні в ідентифікації колокацій, але 4 версія AntConc формує втричі коротший список результатів, що полегшує виокремлення стійких словесних комплексів. Окрім досягнення основної мети – екстракції та формування реєстру стійких словесних комплексів турецько-українського дипломатичного дискурсу, в ході дослідження виявлено переваги та недоліки використання корпусних менеджерів AntConc та Sketch-Engine в автоматизованій обробці турецьких текстів,

The image displays two screenshots of the AntConc software interface. The top screenshot shows the 'KWIC' view with a search query that has returned 2 hits. The table below shows the results:

File	Left Context	Hit	Right Context
1 25.09.2023 Dışişleri Bak...	İstisnasız tüm önemli konuları kapsadı.	Karadeniz Tahil Girişimi'nin	yeniden canlandırılmasının bu görüş
2 5.06.2024 Andrii Yermak...	toprak bütünlüğünü desteklemektedir.	Karadeniz Tahil Girişimi'nin 2022-2023	yıllarında başarılı olmasına aktif olara

The bottom screenshot shows the 'Target Corpus' view with 6 hits. The table below shows the results:

File	Left Context	Hit	Right Context
1 25.09.2023 Dışişleri Baka...	m aslında her iki Bakana da olacak. Bildiğimiz üzere	Karadeniz Tahil Girişimi	17 Temmuz itibarıyla sona erdirmiş
2 8.07.2023 "Ukrayna'nın ...	k girişimleri neticesinde İstanbul'da atılan imzalarla	Karadeniz Tahil Girişimi'ni	hayata geçirdik. Türk boğazlarında
3 25.09.2023 Dışişleri Baka...	görüşmeler istisnasız tüm önemli konuları kapsadı.	Karadeniz Tahil Girişimi'nin	yeniden canlandırılmasının bu gör
4 5.06.2024 Andrii Yermak...	Ukrayna'nın toprak bütünlüğünü desteklemektedir.	Karadeniz Tahil Girişimi'nin	2022-2023 yıllarında başarılı olma
5 1.09.2023 Dmytro Kuleb...	ü konularında kapsamlı görüşmeler gerçekleştirdik.	Karadeniz Tahil Girişimini	yeniden tesis etmek için birlikte çalış
6 25.09.2023 Dışişleri Baka...	HAKAN FIDAN- Cevap için çok teşekkür ediyorum.	Karadeniz Tahil Girişimiyle	ilgili soruya cevaben şunları ifade e

Рис. 5. Словосполучення ідентифіковане програмою AntConc як дві різні колокації

Таблиця 1

	ССК	самостійно		Sketch Engine		AntConc	
		ч	т	ч	т	ч	т
1	Dışişleri Bakanı	20	5	20	5	20	5
2	Devlet Başkanı	18	9	18	9	18	9
3	toprak bütünlüğü	14	10	14	10	14	10
4	(Ukrayna / Küresel) Barış Zirvesi	14	5	14	5	14	5
5	ele almak	13	8	10	6	13	8
6	bir araya gelmek	12	6	12	6	12	6
7	sona ermek	12	4	12	4	12	4
8	NATO	11	6	4	4	11	6
9	uluslararası hukuk	11	5	11	5	11	5
10	Bakan Yardımcısı	11	4	10	3	11	4
11	Karadeniz (tahıl) girişimi / tahıl girişimi	10	5	8	5	10	5
12	esir takası/değişimi/mübadelesi	7	5	7	5	7	5
13	Dışişleri Bakanlığı	7	5	7	5	6	4
14	(Karadeniz) tahıl koridoru	7	4	7	4	6	4
15	(küresel/dünya) gıda güvenliği	7	4	7	4	6	3
16	Barış Formülü	6	6	6	6	6	6
17	iş çevresi / iş dünyası	6	4	6	4	6	4
18	(İstanbul) Tahıl (Koridoru) Anlaşması	5	5	5	5	3	3
19	adil (bir) barış	5	5	4	4	3	3
20	uluslararası toplum	5	4	5	4	5	4
21	ağırlamaktan memnuniyet duymak	4	4	4	4	4	4
22	Kırım'ın (yasadışı) ilhaki	4	4	2	2	-	-
23	Kırım Platformu (Zirvesi)	4	3	4	3	4	3
24	BM	4	3	-	-	-	-
25	insansız hava aracı (İHA)	4	3	-	-	-	-
26	Başkan Yardımcısı	4	3	4	3	4	3
27	(Uluslararası) Kırım Platformu	4	3	4	3	4	3
28	meydana gelmek	4	2	4	2	4	2
29	kalıcı (bir) barış	3	3	3	3	3	3
30	ikili ilişki	3	3	2	2	2	2
31	ikili görüşme	3	3	3	3	3	3
32	hayatını kaybedenler	3	3	3	3	2	2
33	dile getirmek	3	3	3	3	3	3
34	(ortak) basın toplantısı	3	3	3	2	3	2
35	Ticaret Bakanı	3	3	3	3	3	3
36	Meclis Başkanı	3	2	3	2	3	2
37	Kırım Tatar Milli Meclisi	3	2	3	2	3	2
38	ayağa kaldırmak	3	2	3	2	3	2
39	netice almak	3	2	3	2	3	2
40	Ukrayna Büyükelçiliği	3	2	3	2	3	2
41	Savunma Bakanı	2	2	2	2	2	2
42	basın açıklaması	2	2	2	2	2	2
43	mutabakat zaptı	2	2	2	2	2	2
44	üst düzey ziyaret	2	2	2	2	-	-
45	açılış konuşması	2	2	2	2	2	2
46	yerli halk	2	2	2	2	2	2
47	iş adamı / iş insanı	2	2	-	-	-	-
48	baş başa	2	2	2	2	2	2
49	karşı karşıya kalmak	2	2	-	-	-	-
50	taziye iletmek	2	2	2	2	2	2
51	başsağlığı dilemek	2	2	-	-	2	2

52	meydan okumak	2	2	2	2	2	2
53	ortaya koymak	2	2	2	2	2	2
54	ortaya çıkarmak	2	2	2	2	2	2
55	altını çizmek	2	2	-	-	2	2
56	kamuoyuna saygıyla duyurulur	2	2	2	2	2	2
57	müteahhit şirketleri/müteahhitlik şirketler	2	2	2	2	2	2
58	insani yardım	2	2	2	2	2	2
59	savaş esirleri	2	2	2	2	2	2
60	siyasi tutuklular/mahkumlar	2	2	-	-	-	-

зокрема проблему з токенизацією в Sketch-Engine та необхідність створення еталонного корпусу для виокремлення ключових слів за допомогою AntConc. Побудований на основі аналізу теоретичних даних алгоритм необхідних дій для ек-

трекції стійких словесних комплексів засвідчив, що популярні корпусні менеджери AntConc та Sketch-Engine на даному етапі розвитку не полегшують суттєво процес пошуку стійких словесних комплексів у текстових корпусах.

Список літератури:

1. Алефіренко М. Теоретичні питання фразеології: монографія. Харків: Вища шк., 1987. 167 с.
2. Баран Я. А., Зимомря М. І. Теоретичні основи фразеології: навч. посіб. Ужгород: Ужгор. держ. ун-т, 1999. 176 с.
3. Баран Я. А. Фразеологія у системі мови : автореф. дис. ... д-ра філол. наук : 10.02.15. Київ, 1998. 391 с.
4. Габер Є. В. Війна на словах: як переклад може змінити результати дипломатичних переговорів. *Європейська правда*. URL: <https://www.eurointegration.com.ua/experts/2021/09/30/7128432/> (дата звернення: 21.09.2024).
5. Демська О. М. Текстовий корпус: поняття і визначення. *Дивослово*. 2011. № 10. С. 35–37.
6. Жуковська В. В. Вступ до корпусної лінгвістики: навч. посіб. Житомир: ЖДУ ім. Ів. Франка, 2013. 142 с.
7. Карпіловська Є. А. Вступ до прикладної лінгвістики: комп'ютерна лінгвістика. Донецьк: Юго-Восток, 2006. 188 с.
8. Колоїз Ж. В. Структурно-функціональні різновиди усталених конструкцій у мові публіцистики Сергія Єфремова: автореф. дис. канд. філол. наук: 10.02.01. Дніпропетровськ, 1996. 19 с.
9. Корнієнко Л. М. Українська фразеологія другої половини ХХ – початку ХХІ століть: тлумачні та перекладні словники: дис. канд. філол. наук: 10.02.01. Київ, 2017. 218 с.
10. Локота О. П. Абревіація як особливий спосіб утворення термінів у турецькому дипломатичному дискурсі. *Проблеми семантики, прагматики та когнітивної лінгвістики*. 2009. № 16. С. 281–286.
11. Локота О. П. Формально-структурний аналіз дипломатичної термінології сучасної турецької мови. *Проблеми семантики слова, речення та тексту*. 2010. № 24. С. 73–79.
12. Скрипник Л. Г. Фразеологія української мови: монографія. Київ: Наук. думка, 1973. 278 с.
13. Спотар-Аяр Г. Ю. Культура усного та писемного мовлення: діловий та академічний дискурс. Київ: Авантпост-Прим., 2017. 162 с.
14. Удовиченко Г. М. Словосполучення у сучасній українській літературній мові: монографія. Київ: Наук. думка, 1968. 227 с.
15. Ужченко В. Д., Ужченко Д. В. Фразеологія сучасної української мови: навч. посіб. Київ: Знання, 2007. 494 с.
16. Andrii Yermak: “Ukrayna için adil bir barış: Türkiye'nin rolü ve Küresel Barış Zirvesi” / Ukrayna'nın Ankara Büyükelçiliği. *Embassy of Ukraine in the Republic of Türkiye*. URL: <https://turkiye.mfa.gov.ua/tr/news/andrij-yermak-spravedlivij-mir-dlya-ukrayini-rol-turechchini-ta-globalnij-samit-miru> (дата звернення: 21.09.2024).
17. AntConc (Windows, MacOS, Linux). *Laurence Anthony's Website*. URL: <https://www.laurenceanthony.net/software/antconc/releases/AntConc431/help.pdf> (date of access: 21.09.2024).
18. collocation. *Cambridge Dictionary / English Dictionary, Translations & Thesaurus*. URL: <https://dictionary.cambridge.org/uk/dictionary/english/collocation> (date of access: 21.09.2024).
19. Corpus Analysis with Antconc / Programming Historian. *Programming Historian / Programming Historian*. URL: <https://programminghistorian.org/en/lessons/corpus-analysis-with-antconc> (date of access: 21.09.2024).

20. #CorrectUA / Міністерство закордонних справ України. *Ministry of Foreign Affairs of Ukraine*. URL: <https://mfa.gov.ua/en/correctua> (date of access: 21.09.2024).
21. Dağ A. E. Uluslararası ilişkiler ve diplomasi sözlüğü: İngilizce-Fransızca-Türkçe. İstanbul: Vadi Yayınları, 2009. 515 p.
22. Dmytro Kuleba Kyiv’de Türkiye Cumhuriyeti Dışişleri Bakanı Hakan Fidan ile görüştü. / Ukrayna’nın Ankara Büyükelçiliği. *Embassy of Ukraine in the Republic of Türkiye*. URL: <https://turkiye.mfa.gov.ua/tr/news/dmytro-kuleba-kyivde-turkiye-cumhuriyeti-disisleri-bakani-hakan-fidan-ile-gorustu> (date of access: 21.09.2024).
23. Dışişleri Bakanı Sayın Hakan Fidan’ın Ukrayna Dışişleri Bakanı Dmytro Kuleba ile Ortak Basın Toplantısı, 25 Ağustos 2023 / T.C. Dışişleri Bakanlığı. *Republic of Türkiye Ministry of Foreign Affairs*. URL: <https://www.mfa.gov.tr/disisleri-bakani-sayin-hakan-fidan-in-ukrayna-disisleri-bakani-dmitro-kuleba-ile-ortak-basin-toplantisi-25-8-2023.tr.mfa> (date of access: 21.09.2024).
24. Finegan E. Language: Its structure and use. 5th ed. Boston, MA: Thomson Wadsworth, 2008. 574 p.
25. N-grams, multiword expressions, lexical bundles / Sketch Engine. *Sketch Engine*. URL: <https://www.sketchengine.eu/guide/n-grams-multiword-expressions/> (date of access: 21.09.2024).
26. Reference corpus for term extraction / Sketch Engine. *Sketch Engine*. URL: <https://www.sketchengine.eu/corpora-and-languages/reference-corpus/> (date of access: 21.09.2024).
27. O’Keeffe A., McCarthy M. J. The Routledge Handbook of Corpus Linguistics. 2nd ed. London: Routledge, 2022. URL: <https://doi.org/10.4324/9780367076399> (date of access: 21.09.2024).
28. Stefanowitsch A. Corpus linguistics: A guide to the methodology. Berlin: Language Science Press, 2020. 510 p.
29. T. C. Cumhurbaşkanlığı. URL: <https://www.tccb.gov.tr> (date of access: 21.09.2024).
30. T.C. Dışişleri Bakanlığı – Kiev Büyükelçiliği. *T.C. Dışişleri Bakanlığı – Kiev Büyükelçiliği*. URL: <https://kiev-be.mfa.gov.tr/Mission/Index> (date of access: 21.09.2024).
31. T.C. Dışişleri Bakanlığı. *Republic of Türkiye Ministry of Foreign Affairs*. URL: <https://www.mfa.gov.tr/default.tr.mfa> (date of access: 21.09.2024).
32. Ukrayna Cumhurbaşkanı Volodimir Zelenski ile Telefon Görüşmesi / Basın Açıklamaları / Faaliyetler. *T. C. Cumhurbaşkanlığı*. URL: <https://tccb.gov.tr/basin-aciklamalari/365/142812/ukrayna-cumhurbaskani-volodimir-zelenski-ile-telefon-gorusmesi> (дата звернення: 21.09.2024).
33. Ukrayna Devlet Başkanı Volodimir Zelenski ile Telefon Görüşmesi / Basın Açıklamaları / Faaliyetler. *T. C. Cumhurbaşkanlığı*. URL: <https://tccb.gov.tr/basin-aciklamalari/365/142427/ukrayna-devlet-baskani-volodimir-zelenski-ile-telefon-gorusmesi> (date of access: 21.09.2024).
34. Ukrayna’nın Ankara Büyükelçiliği. *Embassy of Ukraine in the Republic of Türkiye*. URL: <https://turkiye.mfa.gov.ua/tr> (date of access: 21.09.2024).
35. “Ukrayna’nın toprak bütünlüğüne desteğimizi tüm platformlarda ifade ettik” / Haberler. *T. C. Cumhurbaşkanlığı*. URL: <https://tccb.gov.tr/haberler/410/147638/-ukrayna-nin-toprak-butunlugune-destegimizi-tum-platformlarda-ifade-ettik-> (date of access: 21.09.2024).
36. Ünsal G. Diplomasi dili ve çevirisi. *RumeliDE Dil ve Edebiyat Araştırmaları Dergisi*. 2019. No. 17. P. 22–39.
37. User Guide / Sketch Engine. *Sketch Engine*. URL: <https://www.sketchengine.eu/guide/> (date of access: 21.09.2024).
38. Word sketch – collocations and word combinations / Sketch Engine. *Sketch Engine*. URL: <https://www.sketchengine.eu/guide/word-sketch-collocations-and-word-combinations/> (date of access: 21.09.2024).

Celik K. S. EXTRACTION OF STABLE VERBAL COMPLEXES FROM THE TEXTS OF MODERN TURKISH-UKRAINIAN DIPLOMATIC DISCOURSE

The article is dedicated to the study of Turkish-language texts of modern Turkish-Ukrainian diplomatic discourse in order to identify relevant stable verbal complexes. Considering that there is no common definition of the concept of “stable verbal complex”, the paper illustrates the existing approaches and presents its own understanding of stable verbal complex as semantically coherent stable combinations of words that naturally occur in speech and can perform nominative and communicative or expressive functions. The article then outlines the role of stable verbal complexes in diplomatic discourse and substantiates the importance of compiling and regularly updating their register in view of the continuous change in their relevance. The practical part of the work describes the algorithm of our extraction of relevant stable word complexes of Turkish-Ukrainian diplomatic discourse based on the material of Turkish-language texts published between 01.01.2023 and 11.07.2024 on the official websites of the Ministry of Foreign Affairs of Turkey, the Presidency of the Republic of Turkey, the Embassy of the Republic of Turkey in Ukraine and the Embassy of Ukraine in

the Republic of Turkey. The procedure, carried out with the involvement of Sketch Engine and AntConc 4.3.1 software, consisted of the following stages: formation of a collection of texts; single-handedly search of stable verbal complexes and fixation their frequency in the formed collection; creation of the text corpus and identification of collocations in the corpus with the help of corpus managers; extraction of stable verbal complexes from the obtained list of collocations; validation of the results; forming a list of relevant stable verbal complexes of Turkish-Ukrainian diplomatic discourse. Besides achieving the main goal of the research, the study identified the main advantages and disadvantages of using corpus managers to identify collocations in Turkish texts; revealed some inconsistencies in the transliteration of Ukrainian proper names with the originals; and identified texts that can be considered as educational materials.

Key words: *Turkish-Ukrainian diplomatic discourse, stable verbal complex, stable verbal complexes of Turkish diplomatic discourse, processing Turkish with Sketch Engine, processing Turkish with AntConc.*