

Алла Чутковська
Національний університет
«Києво-Могилянська академія»

КОЛОКАЦІЇ КОЛЬОРОНАЗВ В УКРАЇНСЬКІЙ МОВІ: КОРПУСНИЙ ПІДХІД

У роботі досліджено колокації кольороназв на базі Генерального регіонально анотованого корпусу української мови. Визначено особливості текстового корпусу як джерела для дослідження мови, описано створені корпуси української мови. Розглянуто поняття колокації як у широкому, так і вузькому, власне корпусному, значенні. Простежено кореляцію між частотністю вживання кольороназви та кількістю її значень. Проаналізовано колокації кольороназв, визначено, що вони становлять континуум між вільними та фіксованими сполуками.

Ключові слова: *текстовий корпус, колокація, кольороназва.*

In this work, collocations of color titles were studied on the basis of the General Regionally Annotated Corpus of Ukrainian. The features of the text corpus as a source for the study of the language are described, and the Ukrainian language corpora are analyzed. The concept of collocation in the broad and narrow sense is considered. The correlation between the frequency of use of color names and the number of its values is traced. The collocations of color names are analyzed, it is determined that they constitute the continuum between free and fixed compounds.

Key words: *text corpus, collocation, color name.*

Корпусна лінгвістика є одним із найбільш перспективних напрямів мовознавчої науки. Передумовами її виникнення були розвиток комп'ютерних технологій та наявний у лінгвістиці погляд на текст як на єдиний безпосередньо доступний об'єкт для вивчення мови. Текстовий корпус – це «певним чином організоване електронне зібрання писемних і усних текстів природної мови», якому властиві репрезентативність, автентичність, відібраність, збалансованість та стандартність [2, с. 89]. Перевагами використання корпусу для лінгвістичних досліджень є можливість швидко опрацьовувати великі за обсягом масиви даних, отримувати автоматично визначену

статистичну інформацію. Окрім того, використання електронного корпусу мінімілізує вплив суб'єктивного чинника і допомагає звернути увагу на досі непомічені мовні явища.

Корпус є зручним та корисним інструментом для різноманітних лінгвістичних досліджень. Його можна використовувати для вивчення фонетики, морфології, лексики, граматики, словотвору, синтаксису, семантики. Корпусні дані можуть бути застосовані в діалектології, соціолінгвістиці, прагмалінгвістиці тощо [3]. Головною умовою використання корпусу при цьому є його відповідність дослідницьким завданням, наявність у ньому розмітки певного типу. Сьогодні корпуси активно

застосовують для створення сучасних словників та граматик, для вивчення іноземних мов, перекладу, а також для створення систем автоматичного опрацювання мови.

Наразі створено декілька корпусів української мови. Це Український національний лінгвістичний корпус, розроблений в Українському мовно-інформаційному фонді, корпус української мови, створений в лабораторії комп'ютерної лінгвістики при університеті ім. Т. Шевченка, а також Генеральний регіонально анотований корпус української мови (ГРАК). Кожен із цих корпусів має свої переваги та недоліки. Втім, серед них ГРАК найбільше відповідає вимогам до сучасного корпусу. Він найбільший за обсягом, до нього є вільний доступ через мережу Інтернет, він має стандартне програмне забезпечення для підрахунку статистики, зокрема для автоматичного визначення колокацій. Тому саме Генеральний регіонально анотований корпус було використано для нашого дослідження.

Термін “колокація” не є однозначним, зазвичай його тлумачать у широкому та вузькому аспектах. Широкий підхід визначає колокацію як звичне поєднання певних лексем разом [1, с. 15]. У корпусному підході колокації визначають на основі статистичних даних як слова, що трапляються у текстах разом частіше, ніж за випадковою вірогідністю окремо [5, pp. 115-116]. Колокації є досить неоднорідними, становлять певний континуум між вільними та фіксованими сполуками.

Кольороназви посідають важливе місце у лексичній системі мови. В українській мові більшість слів на позначення основних кольорів є багатозначними. При цьому первинним є значення кольору, а вторинні значення

формується на основі асоціацій чи під впливом позамовної дійсності [4, с. 26]. Також кольороназви мають різну частотність вживання. Найбільш частотними є саме ті кольоративи, що мають найбільшу кількість значень.

Для дослідження обрано назви таких кольорів: “білий”, “чорний”, “червоний”, “жовтий”, “зелений”, “синій”. Визначені корпусом колокати кольороназв є досить різноманітними як за частиномовним вираженням, так і за іншими параметрами. Найбільш частотними є колокації кольоративів із іменниками, службовими частинами мови, прикметниками та дієсловами. Попри те, що не всі ці слова можуть мати безпосередній граматичний зв'язок із кольоративами, частотність, регулярність їхнього вживання у текстах разом дають підстави розглядати їх у межах корпусного підходу як колокації.

Усім кольоративам властиві частотні поєднання з прийменниками, сполучниками, часткою *аж*. Також усі кольороназви часто поєднуються зі словами на позначення інших основних кольорів – *білий, чорний, зелений, червоний, синій*. Щодо поєднань із прикметниками – є слова, що часто виступають колокатами усіх кольороназв. Це прикметники на позначення форми, розміру, матеріалу: *великий, величезний, маленький, малесенький, довгий, короткий, широкий, високий, тонкий, шовковий, оксамитовий, полотняний, вовняний*. Є такі, що часто вжиті поряд лише з певною кольороназвою: із “зелений” – *буйний, свіжий, порослий, молодий, лісовий, квітучий, соковитий, весняний*. Дієслова-колокати найчастіше є універсальними для всіх кольороназв, але є такі, що пов'язані лише з одним кольором (*чорний- зяяти*).

Колокації кольороназв із іменниками є дуже різноманітними, із них найбільше помітно, що колокації є континуумом між вільними та зв'язаними сполуками. Найбільш вільними є ті, у яких зв'язок із певним кольором не визначений семантикою іменника. Наприклад, вільні сполуки з кольоративами утворюють іменники *колір, фарба, барва, тло, фон*, назви одягу (*штани, сукня, футболка, спідниця*), назви тканин (*шовк, оксамит, тканина, сукно*), предметів (*олівець, портфель*) тощо.

Менш вільними вважаємо такі сполуки, в яких поєднання іменника з певним кольоративом зумовлене його семантикою (*біле молоко, зелена трава, синє небо, жовта кульбаба, червоний мак*), та такі, в яких кольоратив актуалізує якість зі своїх додаткових значень. Наприклад, у сполуці *зелений помідор* “зелений” означає “недостиглий”, у сполуці *червоний терор* “червоний” означає “більшовицький, у сполуці *чорна година* актуалізоване значення слова “чорний” як “важкий, пов'язаний із труднощами”.

Не є вільними і сполуки, в яких кольороназва виступає постійним епітетом: *біле лице, чорна рілля, червона калина, синє море, зелений гай*. Близькими до фіксованих є сполуки на позначення сорту, виду чого-небудь: *біле золото, чорний хліб, червоне вино, зелений чай, синій кит*. Фіксованими є сполуки, значення яких не тотожне сумі значень окремих компонентів. Це можуть бути

власні назви (*Біла Церква, Червоне море, Зелений Клин, Синя Борода*), сполуки, в яких кольороназви втратили свою семантику частково (*біла пляма, чорний гумор, червона планета, жовта раса, зелене світло*) або повністю (*зелений змії, синя панчоха, білий віри, жовта пропасниця*). Окрім двослівних, кольороназви можуть входити до складу трикомпонентних колокацій. Утім, виявити їх складніше, адже корпус не має функції автоматичного визначення трислівних колокацій.

Інформація, отримана внаслідок аналізу колокацій, є важливою для розробки систем автоматичного опрацювання мови. Отак, фіксовані та наближені до фіксованих колокації при машинному опрацюванні слід розглядати як цілісне поняття. Також слово-колокат може вказувати на актуалізоване в конкретному вживанні значення багатозначного слова, слугувати для зняття омонімії. Крім цього, дослідження колокацій є корисними для лінгводидактики, перекладу, лексикографії.

Наразі в Україні дослідження колокацій на базі корпусу не є поширеними, адже й сама корпусна лінгвістика розвивається не так інтенсивно, як, наприклад в англійських країнах (хоч деякі здобутки, безперечно, є). Тому є потреба створювати нові корпуси та вдосконалювати вже наявні, проводити на їхній основі дослідження і відкривати українську мову з корпусної точки зору.

Список літератури

1. Бобкова Т. В. Теоретико-методологічні підходи до вивчення колокацій / Т. В. Бобкова // Вісник Київського нац. лінгв. ун-ту. Серія : Філологія. – 2014. – Т. 17. – № 2. – С. 14–22.

2. Демська О. Текстовий корпус: ідея іншої форми / Оріся Демська . – Київ : ВД "Києво-Могилянська академія", 2011 . – 282 с.
3. Лук'янець Г. Г. Основні напрямки сучасних корпусних досліджень мови та перспективи їх подальшого розвитку / Г. Г. Лук'янець // Наукові праці Національного університету харчових технологій. - 2012. - № 44. - С. 127-133. - Режим доступу: http://nbuv.gov.ua/UJRN/Npnukht_2012_44_22.
4. Шеховцова О.А. Формування символічної семантики кольоропозначень: Дис. ... канд. філол. наук: 10.02.15. – Донецьк, 2003. – 212 с.
5. Sinclair J. Corpus, Concordance, Collocation / J. Sinclair — Oxford: Oxford University Press, 1991. — 170 p.