

*Корольова А. В.,
доктор філологічних наук, професор,
проректор з наукової роботи і міжнародного співробітництва
Київського національного лінгвістичного університету*

МУЛЬТИМЕДІЙНИЙ ПІДКОРПУС ТЕКСТІВ ІЗ ЄВРОПЕЙСТИКИ: КРИТЕРІЇ ДЛЯ ТЕМАТИЧНОЇ РОЗМІТКИ

Анотація. У статті розроблено базові критерії для тематичної розмітки мультимедійного підкорпусу текстів із європейстики (англ. European studies) – дисципліни, на меті якої є популяризація інформації, відображеної в засобах масової інформації, про різні сфери діяльності європейських інституцій: про державну політику ЄС, європейську історію, європейське право, економіку та внутрішні взаємини в ЄС від наукових кіл до політичного співтовариства. За допомогою корпусних програм (WordSmith, WordList, MonoConc Pro, CATMA й KORP та ін.) ідентифіковано тексти (англійською, німецькою, французькою мовами), наявні у вільному доступі в мережі Інтернет, за тематикою належності до однієї зі сфер європейстики та укладено їх у мультимедійний підкорпус Мультилінгвального корпусу. Сформульовано оптимальне визначення поняття мультимедійний підкорпус текстів, яке прийнято за базовий критерій для тематичної розмітки. Розкрито мультимодальний характер текстів мультимедійного підкорпусу на основі виявлення сенсорно-асоціативних (невербально-екстравербальних) зв'язків між сконструйованою в текстах інформацією та реаліями (політичними, економічними, правовими) країн ЄС; доведено визначальну роль ключових слів для здійснення тематичної розмітки текстів із європейстики та аргументовано їх статус як концептуальних вербальних маркерів конструювання сенсорно-асоціативних зв'язків у мультимедійному підкорпусі; визначено два типи моделей комунікативного впливу текстів мультимедійного підкорпусу на аудиторію користувачів мультилінгвального корпусу, які теж виконують функцію базового критерію для тематичної розмітки. Обґрунтовано ефективність запропонованих базових критеріїв для виконання тематичної розмітки текстів з європейстики в мультимедійному підкорпусі.

Ключові слова: корпусна лінгвістика, європейстика, мультимедійний підкорпус текстів, тематична розмітка, базові критерії, асоціативно-сенсорні зв'язки, ключові слова-концепти.

Постановка проблеми у загальному вигляді та її зв'язок з важливими науковими та практичними завданнями. Євроінтеграційні прагнення України були Конституційно визначені у 2019 році як зовнішньополітичний орієнтир, і на законодавчому рівні держава утвердила реформування всіх сфер життєдіяльності, зокрема й освітньо-наукової сфери, задля виконання завдань із набуття повноправного членства в ЄС [1].

У Київському національному лінгвістичному університеті на третьому рівні здобуття вищої освіти до переліку освітніх компонентів професійної підготовки аспіранта була включена дисципліна «Мультилінгвальний корпус та його ресурси для

дослідження європейстики» (далі – Мультилінгвальний корпус), яка спрямована на формування комплексу знань здобувачів про спектри досліджень із європейської інтеграції, шляхи консолідованої співпраці між науковцями, освітянами та іншими соціальними групами і спільнотами. Окрім того, одним із завдань навчальної дисципліни є популяризація інформації про державну політику ЄС, європейську історію, європейське право, економіку та внутрішні взаємини в ЄС. Безумовно, що різні сфери діяльності європейських інституцій відображаються в засобах масової інформації, найпопулярнішим з яких серед широкої громадськості є медійний дискурс або, як останнім часом його визначають, – дискурс мультимедіа [2].

Дослідження медійного дискурсу та його текстів у широкому розумінні є актуальним і перспективним питанням із позицій корпусної лінгвістики, зокрема у контексті опрацювання можливостей корпусного інструментарію, по – перше, для акумуляції й статистичного контент-аналізу різномовних полімодальних текстів із європейстики, укладених у мультимедійний підкорпус, а по – друге, для використання цього корпусного ресурсу з метою популяризації євроінтеграційних процесів серед різних соціальних груп і спільнот в Україні.

Мета статті: розробити базові критерії для тематичної розмітки мультимедійного підкорпусу текстів із європейстики як ресурсу мультилінгвального корпусу.

Завдання. Для коректної ідентифікації текстів за тематикою їх належності до однієї зі сфер європейстики та їх укладання в мультимедійний підкорпус необхідно виконати такі завдання:

- 1) сформулювати оптимальне визначення поняття мультимедійний підкорпус текстів з європейстики, яке прийнято за базовий критерій для тематичної розмітки;
- 2) розкрити мультимодальний характер текстів мультимедійного підкорпусу шляхом виявлення сенсорно-асоціативних зв'язків між сконструйованою в них інформацією та реаліями (політичними, економічними, правовими) країн ЄС;
- 3) довести припущення про визначальну роль ключових слів для здійснення тематичної розмітки текстів із європейстики та аргументувати їх статус як концептуальних вербальних маркерів конструювання сенсорно-асоціативно-перцептивних зв'язків у мультимедійному підкорпусі;
- 4) типологізувати моделі комунікативного впливу текстів мультимедійного підкорпусу на аудиторію користувачів Мультилінгвального корпусу, які виконують функцію базового критерію для фіналізації тематичної розмітки;
- 5) обґрунтувати ефективність запропонованих базових критеріїв для виконання тематичної розмітки текстів із європейстики у мультимедійному підкорпусі.

Виклад основного матеріалу дослідження з повним обґрунтуванням отриманих наукових результатів. Найбільшим підкорпусом Мультилінгвального корпусу є мультимедійний, під яким розуміємо динамічну колекцію оцифрованих текстів (англомовних, німецькомовних, французькомовних), переважно газетних, присвячених висвітленню актуальних проблем європейської політики, економіки, правової сфери та взагалі сучасних реалій країн ЄС.

Сфера Європеїстики (англ. *European studies*) окрім наукових студій є також значним ресурсом медійної інформації, де обговорюються важливі питання діяльності всіх європейських інституцій (Європейського парламенту, Європейської ради, Ради Європейського Союзу, Європейської комісії, Суду Європейського Союзу, Європейської Рахункової палати, Європейського центрального банку та ін). Ця сфера вже містить тематичну гіпер-розмітку текстів, яка є базовим критерієм для їх укладання у підкорпус [3, с. 50].

Вербальними експліцитними маркерами для внутрішньої таксономії текстів у підкорпусі є ключові слова (детальніше про це див. нижче), що концентрують основний зміст тексту і викликають сенсорне сприйняття висвітлюваної у ньому проблеми й одночасно стимулюють асоціативну реакцію в потенційного реципієнта на медійно сконструйований зміст. З урахуванням ключових слів із вершиною *Europe* (та колокатами-дериватами на зразок *European Commission*) кожен газетний текст у своїй мультимодальній (вербальній, невербальній та екстравербальній) організації (*Teder-Sallejarvi W.A., Di Russo F., Mc-Donald J.J., Hillyard S.A.*) належить до однієї з тематичних груп мультимедійного підкорпусу текстів [див. докладніше: 4], і залежно від смислових домінант чинить комунікативний вплив на потенційного адресата тексту і користувача корпусом.

Мультимодальні форми конструювання інформації в усіх газетних текстах підкорпусу відтворюють певну тематику, пов'язану з європеїстикою, за допомогою цифрових кодів, що регламентують правила комбінації й алгоритмізації різних видів модальності й перебудовують сенсорні й асоціативні механізми її сприйняття адресатом. Припускаємо, що мультимедійний підкорпус як цілісний корпусний ресурс оцифрованих газетних текстів запускає процес психологічної адаптації користувача (як цим ресурсом, так і інструментами Мультилінгвального корпусу) до різних комбінацій у них мультимодальних форм створення інформації та її динаміки, яка продовжується в інших текстах відповідної тематики. Це сприяє, з одного боку, диференціації тематичних сфер європеїстики, а з іншого, їх інтеграції у сенсорних і асоціативних взаємозв'язках. І завдяки новим можливостям корпусних технологій створюється комплексне і цілісне сприйняття мультимодальних продуктів: від окремих газетних текстів до їх тематично розмічених колекцій у структурі мультимедійного підкорпусу.

Про значущість психоментальних принципів медіаопосередкування йдеться у колективній праці "Oxford handbook of cyberpsychology", 2018 (A. Attrill-Smith, C. Fullwood, M. Keer, D.J. Kuss), де автори припускають, що людина у продуктах мислення (у широкому сенсі таким є мультимедійний підкорпус, а у вузькому – ключові слова-концепти конструювання інформації певної тематики в його текстах) концентрує досвід, пов'язаний із самопізнанням [5]. Тому вся сукупність

способів відтворення реальності (політичної, економічної, правової тощо) спирається на мультимодальний підхід до опису її об'єктів за допомогою комбінацій ключових слів-концептів із різними зображеннями, формулами, фотографіями та навіть комп'ютерними технологіями.

Така артикуляція різних знаків модальності загалом дозволяє окреслити смислові рамки конкретного виду інформаційно-текстової реальності та задати нові топоси її розуміння.

Специфіка цифрового мультимедійного підкорпусу текстів полягає у розширенні спектру управління різними регістрами свідомості й насамперед сенсорно-асоціативними її механізмами за рахунок екстремальної динамізації різних семіотичних систем, ускладнення їх структури та широкої доступності. Мультимодальні концептуальні зв'язки всередині кожного тексту як семіотичного комплексу засвідчують функціонування нового покоління інформаційно-комунікаційних засобів впливу на адресата. Відповідно, і мультимедійний підкорпус текстів як комплекс ресурсів й інструментів управління сенсорним та асоціативно-перцептивним досвідом людини формує нового потенційного адресата-користувача з особливими властивостями бачення, розуміння, пізнання та психологічного реагування.

Процеси, що відбуваються в суспільстві, найбільш динамічно відображає лексико-семантична система кожної мови, оскільки вона відкрита до змін. Усі соціальні зрушення показує динаміка ключових слів, що в певний період перебувають у центрі уваги носіїв мови. Функціонування таких слів у газетному тексті – результат актуалізації у соціальній та асоціативній свідомості представників лінгвосоціуму реалій об'єктивної дійсності. Ключові слова в медіатексті є одиницями з особливим концептуальним статусом, які мають здатність, з одного боку, відображати погляди, цінності суспільства певного періоду, а з іншого – формувати та трансформувати їх.

Статистично визначити часові рамки ключового слова дозволяють корпусні інструменти, наприклад, установити таймінг перебування його в центрі уваги носіїв мови, а головне – умовно розрахувати його життєвий цикл: появу, нарощування частоти вживання, досягнення піку активності, набуття особливої значущості для адресата [6]. Ключові слова, як правило, характеризуються високою частотністю та асоціюються з певною тематичною сферою, при цьому вони можуть виходити за рамки конкретної тематики та вживатись як міжтекстуальні атрактори.

Безпосередньо в мультимедійному підкорпусі текстів з європеїстики *Keywords* виконують додатково функцію інформаційно-пошукової системи. Вони запускають процес пошуку текстів, уже розмічених тематично за її секторами / сферами і, в такий спосіб, допомагають користувачеві здійснити подальшу селекцію текстів за вузькою тематикою (залежно від завдань пошуку). Для виконання цього завдання запит користувача вводиться в поле пошуку. Більш докладно цей алгоритм описано в попередній статті [див. 3, с. 50].

Ключові слова в текстах підкорпусу є своєрідним центром, який пов'язує їх зміст із певною тематикою європеїстики та актуалізує сенсорно-асоціативні зв'язки зі свідомістю адресата. Функція ключового слова реалізується в перебігу сприйняття як однієї фрази (сегмента) тексту, так і цілих фрагментів, де ключові слова, пов'язані між собою змістом, утворюють логічно послідовний тематичний ланцюжок відображення

інформації. Наприклад, до ключових слів мультимедійного підкорпусу текстів із європеїстики політичної тематики можна віднести маркери, пов'язані з політичними ініціативами та конкретними політичними подіями (War in Ukraine); політичні гасла та лозунгові слова; імена політиків, назви політичних партій тощо.

Ключові слова, що формують домінуючі смисли текстів, зазвичай носять концептуальний характер. Концепти виконують функцію категоризації, що дозволяє пов'язати смисл зі словом. Категоризація здійснюється двома операціями: одна полягає в тому, щоб розмежувати об'єкти за категоріями, інша – визначити зв'язки між категоріями. Визначення тематичних кордонів між категоріями також ураховується під час подальшого механічного сортування текстів мультимедійного підкорпусу. Ключове слово-концепт у цілісній єдності тексту корелює з іншими концептами. Якщо слово-концепт належить до базових політичних констант, то вже у вихідній позиції воно орієнтує на апеляцію до європейських політичних інституцій та їх завдань.

З цього випливає висновок про те, що ключові слова як маркери-сміслові домінуючі тексту та циклічні атрактори (ланцюжок концептів) набувають статусу концептуальних структур, що мають поняттєву, ціннісну, сенсорну, асоціативну та перцептивну зони й відіграють важливу синергетичну роль у мультимедійному підкорпусі текстів з європеїстики. Можемо припустити, що за кожним таким ключовим словом стоїть концепт як ментально-психічний конструкт. Функційно, виконуючи роль циклічних атракторів, ключові слова-концепти формують синергетичний та багатоплановий (мультимодальний) і водночас цілісний смисл кожного тексту і мультимедійного підкорпусу загалом.

Виходячи з цих міркувань, зазначимо, що мультимедійний підкорпус Мультилінгвального корпусу текстів з європеїстики в його тематичній розмітці за секторами / рубриками / сферами (європейське політичне життя, європейська економіка, європейська освіта, безпека тощо), виконаної за допомогою корпусного пошуку (із залученням програм WordSmith, WordList, MonoConc Pro, CATMA) в газетних текстах ключових слів-концептів і наявності сенсорно-асоціативних зв'язків сприйняття їх смислу та передбачуваної реакції на інформацію тексту загалом сприятиме комплексному формуванню уявлень його користувачів, переважно – українців зі знанням англійської та / або німецької, та / або французької мов про європейські інститути з їхніми завданнями, цінностями, функціями та повсякденними реаліями країн ЄС.

В основі медіатизації свідомості сучасної людини лежить різка зміна форм керування мультимодальними кодами інформації. Сприйняття й репрезентація розглядаються в єдиному континуумі взаємодії по відношенню до здатності усвідомлення фізичного / реального й віртуального світу. Вони виражаються в тому, що надходження інформації за одним із сенсорних каналів або зміна його стану викликає асоціативну реакцію у відповідь з боку всіх аналізаторів. Цей феномен є одним із проявів сенсорно-асоціативно-перцептивної здатності людини до інтерпретації комплексу всіх модальностей [7]. Медіальні ефекти (що викликають вільні асоціації, експресії ментальних станів, синестезію почуттів та топологічні проєкції) досягаються завдяки цілісному реагуванню аналізаторів на різномодальні види інформації.

Основний акцент мультимодального комунікативного впливу може визначатися в підкорпусі текстів шляхом обчислення ступеня спільності інформації (політичного змісту, економічного, правового тощо), створеної на основі схожості слів-концептів, різномодальних образів, механізмів інтерсенсорної інтеграції, функцій асоціативних ефектів у сприйнятті моделей медіа.

Базовим критерієм (який теж виконує роль тематичної розмітки текстів) для фінального структурування зібраних текстів у мультимедійний підкорпус є ступінь комунікативного впливу на аудиторію. З цією метою проводиться контент-аналіз впливу мультимодальних ефектів на мислення та свідомість потенційного адресата-користувача, оскільки образи виступають як синестетичні утворення – продукти психічного досвіду людини. У зв'язку з цим важливим етапом (хоча й фінальним) для механічної обробки відсортованих за ключовими словами текстів є визначення моделей творення мультимодальної інформації та подальшого прогнозу їхнього комунікативного впливу на адресата.

Попередньо виділено два рівні мультимодальної організації інформації у тексті. Початкова модель чи власне мультимодель формує здатність в адресата поєднувати кілька концептуальних сфер репрезентації інформації (здаймо моделі концептуальної інтеграції Ж. Фоконьє та М. Тернера [8]). Мультимодель надає можливість поєднання минулих і поточних подій для розуміння та інтерпретації ключових слів у їх комбінації із зображеннями, жестами, мімікою (наприклад, політиків) та іншими невербальними й екстралінгвальними засобами.

Наступний рівень організації мультимодальної інформації становить так звана метамодель, яка ґрунтується на відтворенні інтенційності ментального досвіду автора інформації, відображеного в цілісному смислі тексту, прогнозує здатність адресата до репрезентаційного опосередкування й побудови динамічних асоціативно-перцептивних зв'язків у тексті шляхом швидкого перемикавання з одного виду абстракцій на інший тощо. Обидві моделі допомагають фіналізувати тематичну розмітку текстів із європеїстики в мультимедійному підкорпусі.

Висновки. Завдання мультимедійного підкорпусу текстів із європеїстики полягає в генеруванні, накопиченні та систематизації інформації в цій галузі та забезпеченні його користувачів корпусним ресурсом й пошуковою системою для зручного й швидкого опрацювання потрібної інформації про діяльність європейських інституцій та щоденні практики життя в країнах ЄС.

Завдяки мультимодальній організації колекція текстів із європеїстики, тематично розмічена в підкорпусі Мультилінгвального корпусу, сприймається як певний тип мультимедіа. Ефект мультимедіа виникає завдяки тому, що всі види вражень, переживань і емоцій людини у процесі роботи з підкорпусом викликаються шляхом комбінації в його текстах знаків різних семіотичних систем (різних модальностей). Як наслідок, адресат-користувач реагує не лише на смисл окремого тексту, а й на його сенсорно-асоціативно-перцептивні зв'язки з цілою колекцією текстів.

Перспективи подальших досліджень полягають в апробації ефективності розроблених критеріїв для розширення тематики зібраних колекцій текстів під час проведення тренінгів, які плануються в квітні-травні в КНЛУ.

Література:

1. Європейська інтеграція України в умовах війни росії проти України: роль академічних та громадських спільнот: Збірник матеріалів Міжнародної науково-практичної конференції (м. Рівне, 09 листопада 2023 р.). Київ, 2023. 227 с.
2. Холод О. Дискурс мультимедіа як елемент мультикультури. *East European Journal of Psycholinguistics*. 2024. Issue 1. P. 90–97.
3. Корольова А. В. Мультилінгвальний корпус і його програмне забезпечення для дослідження європеїстики. *Вісник Київського національного лінгвістичного університету. Серія «Філологія»*. 2023. Том 26. № 1. С. 49–62.
4. Teder-Sallejarvi W. A., Di Russo F., Mc-Donald J. J., Hillyard S. A. Effects of spatial congruity on audio-visual multimodal integration. *Journal of cognitive neuroscience*. 2005. Vol. 17. No. 9. P. 1396–409.
5. Attrill-Smith A., Fullwood C., Keep M., Kuss D. J. (Ed.). *Oxford handbook of cyberpsychology*. Oxford : Oxford University Press, 2018. 784 p.
6. Adele D. Intersensory facilitation of reaction time: evaluation of counter and diffusion coactivation models. *Journal of Mathematical Psychology*. 1995. Vol. 39. Issue 2. P. 197–215.
7. Noda K., Arie H., Suga Y., Ogata T. Intersensory causality modeling using deep neural networks // *Systems, Man, and Cybernetics (SMC)*. 2013 IEEE International Conference on 13–16 October. Manchester, 2013. P. 1995–2000.
8. Fauconnier G., Turner M. *The Way We Think: Conceptual Blending and The Mind's Hidden Complexities*. New York : Basic Books, 2002. 464 p.

Співфінансується Європейським Союзом. Проте висловлені думки належать лише авторці і не обов'язково збігаються з поглядами Європейського Союзу чи Європейського виконавчого агентства з питань освіти та культури. Ні Європейський Союз, ні орган, що надає гранти, не відповідають за наведені в статті погляди.

Korolyova A. Multimedia subcorpus of texts on European studies: criteria for thematic markup

Summary. The article elaborates basic criteria for thematic markup of the multimedia subcorpus of texts on European studies, a discipline aimed at popularizing information reflected in the media about various areas of activity of European institutions: EU public policy, European history, European law, economy and internal relations in the EU from academia to the political community. With the help of corpus programs (WordSmith, WordList, MonoConc Pro, CATMA and KORP, etc.), the texts (in English, German, French) freely available on the Internet were identified by their belonging to one of the fields of European studies and included in the multimedia subcorpus of the Multilingual Corpus.

The optimal definition of the concept of multimedia subcorpus of texts is formulated, which is taken as the basic criterion for thematic markup. The multimodal nature of the texts of the multimedia subcorpus is shown on the basis of identifying sensory-associative (nonverbal-extraverbal) relations between the information constructed in the texts and the realities (political, economic, legal) of the EU countries; the decisive role of keywords for the thematic markup of texts on European studies is proved and their status as conceptual verbal markers for the construction of sensory-associative relations in the multimedia subcorpus is argued; two types of models of communicative influence of the texts of the multimedia subcorpus on the audience of users of the multilingual corpus are identified, which also function as a basic criterion for thematic markup. The effectiveness of the proposed basic criteria for the thematic markup of texts on European studies in the multimedia subcorpus is substantiated.

Key words: corpus linguistics, European studies, multimedia subcorpus of texts, thematic markup, basic criteria, associative-sensory relations, keywords-concepts.