

16. Рак Л. К. Мовні засоби гіперболізації в українських народних думках // Укр. мова і літ. в шк. – 1961. – № 4. – С. 27–34; 17. Савицький С. С. Словообразованіє нарицательных названий лиц в украинском фольклоре: Автореф. дис... канд. филол. наук: 10.02.02 / Киев. гос. пед. ин-т, 1990; 18. Славянские древности: Этнолингвистический словарь: В 5 т. / Под. ред. Н. И. Толстого. – Т. 1–3. – М., 1995–2004; 19. Халимоненко Г. Лексика дум та історичних пісень (коментар тюрколога) // Укр. мова та літ. в шк. – 1977. – № 9. – С. 21–29; 20. Хроленко А. Т. Курская лингвофольклористика // Лингвофольклористика. – Курск, 2008. – Вып. 14. – С. 4–8; 21. Хроленко А. Т. Лингвокультуrowедение. – Курск, 2001; 22. Хроленко А. Т. Фольклорный мир в алфавитном порядке // Etnolingwistyka. – Nr. 12. – Lublin, 2000. – S. 121–138; 23. Хроленко А. Т. Что такое лингвофольклористика? // Русская речь. – 1974. – № 1. – С. 36–41; 24. Чабаненко В. А. Основи мовної експресії. – К., 1984; 25. Шульжук К. Ф. Обращения в украинских народных песнях: Автореф. дис... канд. филол. наук: 10.02.02 / Киев. гос. пед. ин-т. – К., 1969; 26. Adamowski J., Niebrzegowska S. Sylwetka naukowa Profesora Jerzego Bartmińskiego // W zwierciadle języka i kultury. – Lublin, 1999. – S. 10–20; 27. Bartmiński J. Ludowy styl artysyczny // Encyklopedia kultury polskiej XX wieku. – T. 2. Współczesny język polski. – Wrocław, 1993. – S. 213–222; 28. Słownik stereotypów i symboli ludowych / Red. J. Bartmiński. – T. 1 (1-2). – Lublin, 1996–1999.

КОМП'ЮТЕРНА ЛІНГВІСТИКА

Наталія Чейлитко

Київський національний університет імені Тараса Шевченка

ЛЕКСИКО-ТЕМАТИЧНА МОДЕЛЬ ДИСКУРСУ

Семантичне моделювання дискурсу є однією з наукових проблем, над якою лінгвісти працюють вже протягом кількох десятиліть. Незважаючи на те, що науковцями запропоновано чимало підходів до її розв'язання, проблему, напевно, можна вважати вічною. Адже, як влучно зауважив В. Звєгінцев, семантика є тією туманною сферою, "в яку занурювалося чимало сміливих дослідників – деякі з них так і загубилися в ній, а ті, хто повертався, зазвичай не приносили з собою значних результатів" [2, с. 80].

Безперечно, семантичні моделі мають різну складність залежно від глибини відображення семантичних процесів у дискурсі. Моделі, які відтворюють тематичне розмаїття дискурсу, узагальнено відображають його смисл. Такі моделі покликані дати відповідь на питання про те, **що, коли, кому, ким і з якою метою** повідомляється.

Як показав В. Баєвський, кожний лексико-семантичний варіант (ЛСВ) у тексті становить тему. Тому близькі за значенням ЛСВ у межах певного дискурсу утворюють великі сюжетні теми [1, с. 193]. Відповідно, продуктивним шляхом створення тематичної моделі дискурсу є підхід, який базується на визначенні тематично актуальних ЛСВ та групуванні їх у лексико-тематичні дерева.

Актуальним залишається питання, якою має бути процедура послідовного, несуперечливого визначення тематичної актуальності ЛСВ. Метою статті є висвітлення результатів дослідження, яке полягало у встановленні тематичних ЛСВ дискурсу та створенні на їхній основі лексико-тематичної моделі (ЛТМ) дискурсу. Новою є розроблена методика виявлення тематичних ЛСВ та створення ЛТМ дискурсу із застосуванням методу зон зв'язків словоформ [3].

Зоною зв'язків словоформи (ЗЗС) називається та частина речення, в якій реалізуються всі синтаксичні зв'язки певної словоформи. Метод ЗЗС є графічним, бо передбачає, що ЗЗС можна показати графічно, провівши відрізок через усі елементи речення (ЕР), пов'язані з аналізованим ЕР. Одержаний набір відрізків, що відображають усі зони зв'язків у реченні, уможлиблює обчислення сили синтаксичної спаяності елементів речення. Під **силою синтаксичної спаяності ЕР** розуміємо кількість ЗЗС, які містяться в інтервалі між сусідніми елементами речення. З-поміж показників сили спаяності ЕР для кожного конкретного речення можна визначити найвищий, найнижчий та проміжні між ними показники. Найвищий показник сили спаяності називають **піком найвищої спаяності** в реченні (рис. 1). Кілька послідовно розташованих показників, що мають однакове найвище значення, називають **плато спаяності**. У реальному реченні може бути кілька піків та плато спаяності. Елементи речення, між якими кількість ЗЗС максимальна, утворюють **блок спаяності**.

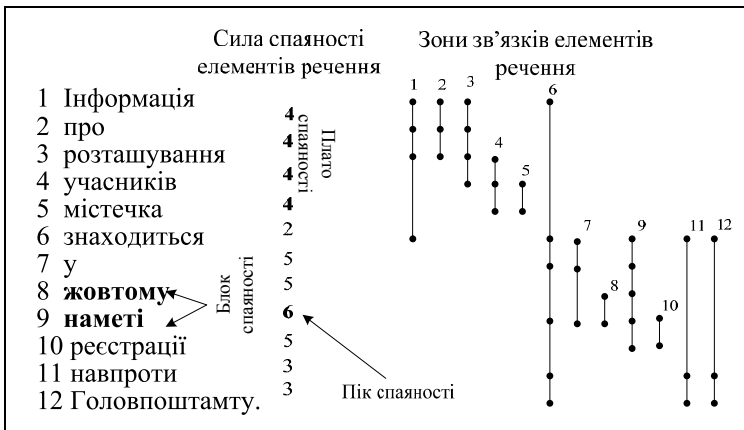


Рис. 1. Піки, плато та блоки спаяності елементів речення

Чим більша кількість ЗЗС припадає на межу між двома елементами речення, тим більш синтаксично спаяними в лінійному потоці мовлення вважаються ці елементи.

ЛСВ, які потрапили до блоків спаяності ЕР, можна вважати тематичними, і тому лексико-тематична модель дискурсу, створена на їх основі, адекватно відобразить тематичну ієрархію проаналізованого дискурсу.

Дискурс можна розуміти як обмежений набір тверджень, що формують і змінюють світ шляхом приписування нових значень мовним знакам. Варто пам'ятати, що значення в дискурсі не є сталим. Жоден дискурс не є замкненим та завершеним: "Ми намагаємося розкинути рибальські тенета так, щоб значення кожного знака опинилося у певних відношеннях з іншими. Але це, зрештою, неможливо, тому що кожна конкретна фіксація значення знака – умовна" [4, с. 49]. Відповідно, будь-яка тематична модель дискурсу, створена один раз і назавжди, є статичною, такою, що не відбиває зміни значень та їх співвідношення в дискурсі. Тому доцільно аналізувати певну послідовність текстів для того, щоб показати динаміку відтворення й трансформації дискурсу. Тоді постає потреба у створенні динамічної тематичної моделі дискурсу, яка могла б оновлюватися – у разі залучення нових текстів (нової сукупності тверджень), що репрезентують певний дискурс. На нашу думку, лексико-тематична модель дискурсу, створена на основі методу ЗЗС, є такою динамічною – відкритою до оновлення – моделлю, оскільки в ній передбачено:

- 1) автоматичний відбір тематичних лексем із текстів на основі позиційного критерію – входження до складу блоків спаяності ЕР;
- 2) автоматизований контроль значення відібраних лексем, який уможлиблює встановлення тих лексико-семантичних варіантів, що є тематично значущими для дискурсу.

Н. Шведова вказала на надзвичайно важливу роль ідеографічного впорядкування лексики з метою виявлення особливостей бачення й відображення світу мовцями. Дослідниця слушно зауважує, що підмножини ЛСВ, які наповнюють нижні рівні ідеографічного дерева, не є простими групами одиниць, відібраними на підставі спільної семантики. На думку науковця, "такі підмножини виконують важливе власне конструктивне й інформативне завдання, відкриваючи перед нами певний "шматочок дійсності" [5, с. 428]. Н. Шведова показала, що "картинки життя" в ідеографічному дереві "вимальовуються" насамперед лексикою, що називає різноманітні реалії буття, – іменниками та дієсловами. Тому ми уклали два лексико-тематичні дерева (ЛТД) тих іменників та дієслів, які увійшли до складу всіх блоків спаяності в проаналізованих текстах. Ми припустили, що лексико-тематична модель (ЛТМ), яка становить сукупність ЛТД повнозначних частин мови, адекватно відобразить "картину життя", що постає з текстів новин, присвячених виборам 2004 р.

Процес створення лексико-тематичної моделі дискурсу складається з таких основних етапів:

1) відбір елементів речення, які увійшли до всіх блоків спаяності ЕР у досліджуваній вибірці текстів;

2) лематизація (приведення до словникових форм) та впорядкування ЕР за частиномовною віднесеністю;

3) визначення для кожної лєми (словникової форми) її значення на основі вивчення контекстів її вживання – відповідно, надалі у фокусі уваги перебувають не лєми, а лексико-семантичні варіанти;

4) присвоєння кожному ЛСВ числового індексу, який показує, скільки разів аналізований ЛСВ увійшов до складу блоку спаяності;

5) упорядкування ЛСВ одного лексико-граматичного класу слів у лексико-тематичне дерево (рис. 2).

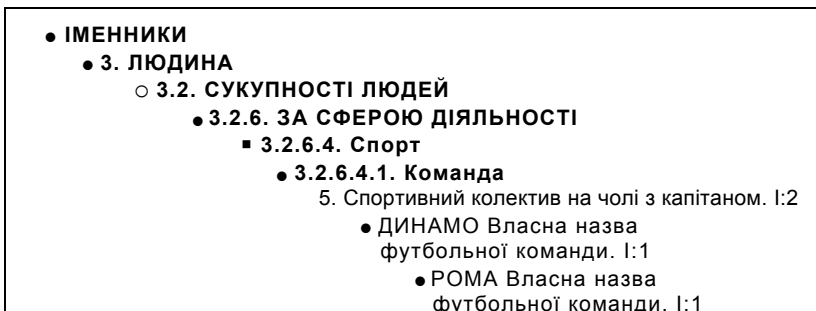


Рис. 2. Фрагмент лексико-тематичного дерева іменників

Таким чином, лексико-тематична модель дискурсу становить сукупність лексико-тематичних дерев повнозначних частин мови.

Зауважимо, що спеціально обчислені показники %_Sum_ЛСВ (кількість ЛСВ, які потрапили до піддерева певного вузла, у відсотках) та %_Sum_I (сума індексів усіх ЛСВ, які потрапили до піддерева цього вузла, у відсотках) уможливили здійснення кількісної характеристики ЛТД.

Розглянемо ЛТМ новин, присвячених виборам 2004 року в Україні. Аналіз за частиномовною належністю ЛСВ, що увійшли до блоків спаяності, виявив таку відсоткову співвіднесеність: 39,6 % усіх ЛСВ становлять власне іменники, що свідчить про основну роль іменників у тематичному розгортанні дискурсу. Тому в цій статті обмежимося описом ЛТД іменників.

Загальна кількість ЛСВ іменників, які потрапили до **лексико-тематичного дерева іменників**, становить 1740 одиниць. ЛТД іменників має шість рівнів глибини та п'ять основних розгалужень.

Таблиця 1. Основні вузли лексико-тематичного дерева іменників

	%_Sum_ЛСВ	%_Sum_I
Іменники	100	100
1. Буття	8,45	10,4
2. Абстрактні поняття	6,03	4,29
3. Людина	42,53	41,56
4. Суспільство	41,32	42,99
5. Природа	1,55	0,55

Відповідно, п'ятьома основними вузлами дерева є вузли другого рівня, які найбільш загально окреслюють явища дійсності, названі іменниками в проаналізованому дискурсі: "Буття", "Абстрактні відношення", "Людина", "Суспільство", "Природа" (табл. 1). Найбільш наповненими лексико-семантичними варіантами є вузли "Людина" – 42,53 % від загальної кількості ЛСВ, "Суспільство" – 41,32 % ЛСВ. Ці ж два вузли мають найбільші показники індексу, тобто частоту потрапляння ЛСВ до блоків спаяності – 41,56 % для "Людини" і 42,99 % для "Суспільства". Цей факт яскраво свідчить про основну тематичну спрямованість дискурсу – суспільне життя людини.

У піддереві вузла "Буття" превалюють ЛСВ на позначення часово-просторових характеристик зображуваних подій та їх кількісні виміри.

У піддереві вузла "Абстрактні поняття" особливу увагу привертає вузол "Оцінка", наповнений незначною кількістю ЛСВ із низькими індексами, тобто він не становить тематичної домінанти дискурсу. Проте саме цей вузол показує, в який спосіб оцінювалася політична ситуація під час виборів 2004 року. Так, найвищий індекс серед означених ЛСВ має *криза 'різка зміна звичайного стану речей, злам, загострення становища'* (I=15). Негативну оцінку містять ЛСВ *худо 'щось погане, неприємне, зло, лихо'* (I=1) та *негарзд 'стан, коли все погано, не дуже добре, не так, як треба'* (I=1); метафоричні номінації *чума 'надзвичайно небезпечне, згубне соціальне явище'* (I=1); *болото 'усе те, що характеризується брудом, застоєм, відсутністю живої діяльності, ініціативи'* (I=1), які вказують на особливості концептуально-метафоричного простору проаналізованого дискурсу: стан суспільства порівнюється, з одного боку, з хворобою (чума), а з другого – з брудом (болото).

Серед ЛСВ, які потрапили до піддереву вузла "Людина", близько 50 % припадає на назви осіб, половина з яких – назви осіб за сферою діяльності. Це засвідчує намір авторів проаналізованих текстів подати вибори та події навколо них як масові – такі, що активували

численні групи населення, різноманітні й за сферою діяльності (політика, військова справа, промисловість і підприємництво, фінансова сфера, медицина, спорт, наука, освіта, мистецтво, релігія), і за родинними зв'язками (від бабусь до онуків), і за громадянством (українці, білоруси, грузини, поляки, канадці та навіть новозеландці), тобто спричинили не лише всенародний, а й усесвітній розголос. Високий показник %_Sum_I вузла з назвами осіб за сферою діяльності "Політика" (третина всіх назв осіб за сферою діяльності) вказує на тематичну домінанту дискурсу – суспільно-політичне життя держави. Тому в дискурсі переважають назви політичних діячів.

Піддерево вузла "Людина як соціальна істота" засвідчило, що в проаналізованому дискурсі суспільно-політичне буття виявляє себе насамперед через активну діяльність людей або їхніх угруповань. Провідне місце посідає діяльність, пов'язана з виборами: ЛСВ *вибір 'той, кого вибрано, те, що вибране'* (I=18); *голос 'право висловлювати свою думку під час вирішення державних або громадських справ'* (I=30); *голосування 'обирання кого-небудь, вирішення чого-небудь поданням голосів'* (I=36).

Піддерева вузлів "Людина як фізична істота", "Людина як психічна істота", "Людина як розумна істота" показують різноаспектні прояви людини в проаналізованому дискурсі. Так, ЛСВ, які наповнюють вузли "Почуття", "Емоційний стан", "Ставлення", дають уявлення про психологічне тло подій, відображених у текстах, – настрої та світовідчуття людей. Переважають ЛСВ на позначення бурхливих людських переживань: *вогонь 'душевне піднесення, натхнення'* (I=1); *ейфорія 'сприйняття, оцінка чого-небудь у надто або невиправдано оптимістичних тонах'* (I=1); *захоплення 'велике внутрішнє піднесення, збудження, порив, запал'* (I=1); *істерика 'напад істерії'* (I=2); *обурення 'сильне невдоволення, роздратування'* (I=1), *підйом 'ентузіазм'* (I=1); *піднесення 'запал, натхнення, душевне піднесення, ентузіазм'* (I=1); *пристрассть 'сильне, бурхливе, нестримне у своєму виявленні почуття'* (I=2); *раж 'сильне збудження, несамовитість, шаленство'* (I=1); *хвиля 'наплив яких-небудь почуттів, думок, що визначають настрої людини'* (I=1).

У піддереві "Суспільство" найбільші показники %_Sum_I припадають на ЛСВ, пов'язані з поняттям держави (вузол "Держава" – 12,87 % від %_Sum_I по цілому дереву, більше третини від сумарного індексу по піддереву "Суспільство"). Це не випадково з огляду на те, що вибори є одним із заходів, який має державне значення.

У проаналізованому дискурсі представлена військова тема (вузол "Військова справа"), оскільки ЛТД іменників містить багато ЛСВ на позначення явищ дійсності, пов'язаних із цією темою. Особливу увагу

привертає значна кількість назв протистоянь, зокрема: **страйк** 'форма боротьби, яка полягає в колективній відмові робітників і службовців від праці з висуненням підприємцям або урядові економічних і політичних вимог' (I=25); **сутичка** 'невеликий короткочасний бій' (I=8). А також назви військових дій, зокрема: **блокування** 'здійснення блокади' (I=5); **напад** 'силова дія, спрямована на кого-, що-небудь з метою захоплення, пограбування, знищення і т. ін.' (I=6); **облога** 'оточення військами укріпленого пункту (міста, фортеці і т. ін.) з метою оволодіння ним' (I=3); **опір** 'протидія кому-, чому-небудь, боротьба з кимось, чимось' (I=7). Хоча вибори 2004 року відбулися переважно як мирна подія, проте наведені дані свідчать про те, що вони осмислювалися й називалися в термінах військового протистояння.

Значна кількість ЛСВ та високі сумарні показники індексів у вузлах "Заклади та організації" (2,1 %), "Населені пункти та їх частини" (6,86 %), "Будівлі, споруди" (4,47 %) вказують на прагнення авторів текстів якнайповніше подати відомості про місце відображених у них подій.

Піддерево вузла "Природа" заповнене невеликою кількістю ЛСВ (27 одиниць) і, відповідно, має невисокі показники індексів. Причина полягає в тому, що тема природи є периферійною в проаналізованих текстах.

Таким чином, ми показали, що одним із шляхів створення тематичної моделі дискурсу є підхід, який базується на виявленні в тексті тематично актуальних ЛСВ та групуванні їх в лексико-тематичні групи.

Тематичну актуальність ЛСВ визначено на основі методу ЗЗС: відібрано всі ЛСВ, які входять до блоків спаяності EP, а також кожному ЛСВ приписано індекс, який відображає кількість уходжень ЛСВ до блоку спаяності. ЛСВ із найбільшим індексом укажуть на тематичне ядро дискурсу, ЛСВ з невисоким індексом – на додаткові теми.

Упорядковані в лексико-тематичну модель ЛСВ іменників та дієслів адекватно відобразили тематичну спрямованість дискурсу політичних новин, присвячених виборам 2004 року.

Описана методика створення лексико-тематичної моделі має великий науковий потенціал, оскільки уможливіло аналіз різноманітних дискурсів, визначення домінантних та периферійних тем.

1. *Баевский В. С.* Лингвистические, математические, семиотические и компьютерные модели в истории и теории литературы. – М., 2001; 2. *Звегинцев В. А.* Предложение и его отношение к языку и речи. – М., 2001; 3. *Перебийнос В. И.* Еще раз о структуре предложения // Чествуя филолога: К семидесятилетию Ф. А. Литвина. – Орел, 2002; 4. *Филлипс Л.* Дискурс-анализ: Теория и метод. – Х., 2004; 5. *Шведова Н. Ю.* Русский язык: Избранные работы. – М., 2005.