

*Бобкова Т. В.,**докторант**Київського національного лінгвістичного університету*

## КОНЦЕПЦІЯ КОЛОКАЦІЇ: КОРПУСНИЙ ПІДХІД

**Анотація.** В статті встановлено концептуальні джерела й основні положення корпусного підходу дослідження колокацій, обґрунтовано необхідність застосування комплексного підходу на базі корпусу текстів.

**Ключові слова:** колокація, словосполучення, текст, контекстно-орієнтований підхід, корпусний підхід, статистичний підхід.

**Постановка проблеми.** Завдання розробки прикладних систем з автоматичного аналізу тексту й укладання словників потребують опрацювання великих обсягів природно-мовних текстів. Сучасні результати спостережень текстових ресурсів – корпусів текстів – виявляють, що при побудові тексту використовуються не окремі слова, а регулярно відтворювані блоки, або конструкції – колокації. Значені природно-мовні структури цілком видобуваються з пам'яті як заготовлені утворення, подібні до простих слів. Регулярність і частотність появи колокацій у текстах становить значний інтерес для прикладних досліджень і вимагає теоретичного обґрунтування феномена.

Окремі загальнотеоретичні аспекти феномена колокації як вияву сполучуваності були об'єктом досліджень зарубіжних і вітчизняних лінгвістів початку ХХ ст. (Ш. Балі, Дж. Фьорз, В. Порциг, В. Виноградов, Л. Теньєр, Н. Амосова, О. Борисова, Дж. Синклер, Ю. Апрусян, І. Вихованець, К. Городенська, В. Телія, В. Горпинич, А. Загнітко, Н. Клименко, М.П. Кочерган, О. Снитко, Л. Йорданська, І. Мельчук, В. Левицький, С. Різник, Р. Марцинкевічене, Н. Дарчук). Однак до цього часу в лінгвістиці відсутнє чітке визначення поняття «колокація», її істотних ознак і статусу серед словосполучень.

**Мета статті** – дослідити розвиток концепції колокації в сучасній корпусній лінгвістиці. Досягнення поставленої мети передбачає виконання таких завдань: 1) здійснити аспектуалізацію поняття «колокація» з урахуванням широкого й вузького розуміння терміна; 2) дати визначення терміна «колокація» в аспекті корпусного підходу; 3) окреслити історичні й теоретичні передумови формування корпусного підходу до концепції колокації; 4) визначити основні теоретичні положення корпусного підходу до опису колокації; 5) навести аргументи на користь комплексного підходу до дослідження колокації.

**Виклад основного матеріалу дослідження.** У сучасній лінгвістиці питання щодо визначення й статусу колокацій серед словосполучень є дискусійним. Термін «колокація» вважається певною мірою не одновимірним, оскільки покликаний відобразити комплекс лексичних, граматичних і функціональних ознак. Недавнє входження терміна в широкий науковий обіг ще не набуло остаточного закріплення в слов'янській лінгвістичній традиції. Саме тому в сучасній лінгвістичній літературі спостерігається співіснування конкурентних термінів на позначення

поняття «колокація»: стабільне сполучення, стабільний контекст [1, с. 205], стійке сполучення [2, с. 31], невільне сполучення [3, с. 10; 4, с. 120], фразеологізм [5, с. 345], лексико-функціональний вираз, фразема [5, с. 345], неоднослівна лексична одиниця [6, с. 27] або багатослівна конструкція [6, с. 26; 7, с. 158].

Значена неусталеність термінології пояснюється співіснуванням у сучасній лінгвістиці давно сформованих традицій і парадигм корпусної лінгвістики [8, р. 47; 9, с. 398]. З іншого боку, це зумовлено неоднорідністю, багатоаспектністю феномена [3, с. 4; 9, с. 398; 10, р. 89; 11, р. 133] і невизначеним статусом колокацій серед словосполучень. Дискусійний характер визначення спонукав до аспектуалізації аналізованого поняття й виділення широкого – лінгвістично мотивованого значення й вузького – в статистичному аспекті [8, р. 47; 12, р. 26]. На сучасному етапі в широкому сенсі під колокацією розуміється поєднання двох слів або звичне використання певних лексем разом [13, р. 115–116]. Корпусний підхід ґрунтується на вузькому значенні: колокація – це слова, які зустрічаються в тексті частіше поруч, ніж за випадковою вірогідністю окремо [13, р. 115–116].

Залежно від інтерпретації концепції у традиціях певної школи сучасні лінгвісти визначають якісний і кількісний склад колокацій і їхній статус серед інших словосполучень [14, с. 18]. На сучасному етапі виділяють семантико-синтаксичний, контекстно-орієнтований і корпусно-орієнтований (корпусний, статистичний) підходи [5, с. 344; 14, с. 16]. В основу традиційних лінгвістичних досліджень колокацій покладено семантико-синтаксичний підхід, за яким зв'язність визначається семантичною сумісністю елементів словосполучення і/або певною синтаксичною моделлю [3, с. 3; 12, р. 27]. За ступенем фразеологізації, або «жорсткості зв'язку елементів» [4, с. 120] колокації традиційно зараховуються до стійких сполучень [14, с. 16] чи характеризуються проміжним між вільними й фразеологізованими сполуками статусом [1, с. 206; 3, с. 4; 4, с. 120; 6, с. 25; 8, р. 47; 14, с. 16; 15], що дає змогу певні колокації об'єднувати з вільними сполуками, а інші визначати як фразеологізми.

В українській лінгвістиці дослідження колокацій ґрунтуються на засадах семантико-синтаксичного [2, с. 32; 4, с. 174] й статистичного підходів [7, с. 163; 16, с. 184]. При цьому увагу дослідників зосереджено не на розв'язанні «лінгвістичних проблем фразеологізації еквівалентних слову стійких сполучень» [4, с. 120], а на ефективності автоматичного виявлення колокацій у текстах [7, с. 159; 16, с. 184]. «Непередбачуваність» природи колокацій зумовлює їх уключення до реєстру словника, призначеного для кодування тексту [8, р. 42]. Так, у системі АГАТ колокації поруч із безумовними ідіомами заносяться до словника фразеологізмів і при автоматич-

ному синтаксичному аналізу «згортаються» в один вузол [4, с. 121].

Витоки власне корпусного підходу до концепції колокації слід шукати в нерозв'язаних проблемах опису лексичного значення, точніше – синтагматичних відношень у лексиці [1, с. 191; 10, с. 71]. У науковий обіг термін «колокація» вперше був уведений Дж. Р. Фьорзом у межах контекстуальної теорії значення в 1930 рр. [12, р. 9]. На той час під колокаціями розумілось типове оточення – звичні або звичайні позиції даного слова [15]. Основною заслугою представників британського контекстуалізму (J.R. Firth, A. McIntosh, M.A.K. Halliday) було усвідомлення необхідності встановлення значення слова через аналіз вживання форми, тобто контексту [5, с. 344]. І, хоча в межах британського контекстуалізму не було сформульовано чіткого визначення колокації, концептуальні засади на пряму послуговували підґрунтям контекстно-орієнтованого й корпусного підходів. Контекстно-орієнтований підхід передбачає застосування як основного прийому синтагматичного аналізу [10, р. 75] контекстуалізації, результатом якої на лексичному рівні є встановлення колокацій слова.

Контекстуальна теорія значення ґрунтується на визнанні колокації цілісною одиницею сенсу, дослідження якої передбачає виокремлення складових [10, р. 43]. Під колокаціями в цьому разі розуміються характерні стійкі сполучення слів, поява яких поруч зумовлюється регулярним характером взаємного очікування внаслідок дії семантичних факторів. Основним здобутком фьорзіанства вважаються встановлені ознаки формальної зв'язності: колокація – супровід слова, матеріал, у якому воно найчастіше зустрічається, і колігація – взаємозв'язок граматичних категорій у синтаксичній структурі [17, р. 209]. Отже, окреслений вище фьорзіанський підхід працює в межах контекстуальної теорії значення, за якою ідентифікація сенсу слова або виразу базується на формалізації контекстної моделі.

Концептуальною близькістю до британського контекстуалізму визначається теорія синтагматичних полів В. Порцига [1, с. 202; 10, р. 71]. Уведений у науковий обіг Дж. Р. Фьорзом термін «колокація» за поняттям наближується до синтагматичних, або «елементарних полів» В. Порцига, що являють собою «поєднання мінімальної кількості членів (двох одиниць)» [1, с. 205–206]. При цьому значення лексичної одиниці є наслідком її синтагматичних зв'язків, інакше кажучи, певні лексичні одиниці беруть участь у вираженні значення інших одиниць [10, р. 71]. Зв'язність таких сполучень, відмінних від вільних, зумовлюється синтагматичними відношеннями, основу яких становлять позамовні й мовні фактори. Зокрема, останні впливають на утворення елементарних полів, базованих на позамовних зв'язках між відповідними денотатами, що підтверджується можливістю зміни складу синтагматичних полів у процесі розвитку мови.

Теорія В. Порцига була розвинута в українській семасіології Е.Ф. Скороходьком і В.В. Левицьким, які розглядали елементарні поля як різновид стабільних сполучень, відповідних колокаціям [1, с. 202–206]. Застосування квантитативних методів для вивчення параметрів широти, інтенсивності й селективності сполучуваності надало можливість встановити, що «перехід від вільних словосполучень до зв'язаних (стійких) має градуованийий

характер», при цьому високий ступінь зв'язку між компонентами є не випадковим [1, с. 205–206]. Отже, концептуальними джерелами корпусного підходу дослідження колокацій слід визнати розроблену в межах британського контекстуалізму теорію колокацій Дж. Р. Фьорза й теорію елементарних полів В. Порцига.

Однак, незважаючи на усвідомлення необхідності й перспектив дослідження колокацій, традиційні підходи мало що могли запропонувати для розвитку концепції. Зокрема, В.В. Левицьким було висловлено гіпотезу щодо наявності статистично значимого зв'язку між компонентами стабільних сполучень і можливості їх виділення й дослідження за допомогою статистичних методів [1, с. 206]. У такий спосіб недостатньо визнані в семасіології елементарні поля будуть об'єктом лінгвістичного аналізу в квантитативних дослідженнях лексичної сполучуваності. Власне аналіз статистичної зв'язності компонентів словосполучення було покладено в основу формування корпусного підходу. У цьому розумінні появу електронних корпусів другого покоління (1980–1990 рр.) й удосконалення методики статистичного опрацювання великих обсягів текстів слід вважати необхідними історичними передумовами формування нового підходу до концепції колокації. Дослідження в галузі корпусної лінгвістики й автоматичного аналізу тексту призвели до поширення терміна «колокація» у багатьох галузях лінгвістики: лексикології, дериватології, синтаксисі, психолінгвістиці, лексикографії, теорії перекладу й лінгвістиці тексту [10, р. 77] – й уможливили розв'язання основних теоретичних проблем щодо визначення колокацій, їх суттєвих ознак і статусу серед словосполучень.

Подібно до контекстно-орієнтованого корпусний підхід визначається опорою на контекст, який розуміється як колекція текстів (корпус) або одиничні тексти з колекції [6, с. 23; 15]. Відповідно до цього, колокація визначається як не випадкове поєднання двох і більше одиниць, характерне як для мови в цілому, так і текстів певного типу або навіть для (під)вибірки текстів зокрема. Завдяки статистичній інтерпретації, концепція колокації Дж. Фьорза набуває нового виміру в корпусному підході. Отже, за корпусним підходом колокація розуміється як послідовність слів, що зустрічаються разом частіше, ніж можна було б очікувати випадково. Як зазначалось вище, формування концепції колокації пов'язано з проблемами опису лексичного значення. У корпусному підході основи теоретичних новацій закладено роботами Дж. Синклера «з перевизначення одиниці значення в аспекті корпусних даних» [17, р. 214]. На підставі спостереження корпусних даних щодо невідповідності меж значень і лексичних одиниць було розроблено «корпусну модель значення» [17, р. 214], суть якої полягає в тому, що значення не зосереджується в межах однієї лексичної одиниці, а розповсюджується на послідовність слів у словосполученні – колокації. За Дж. Синклером, моделлю вважається поєднання слів, яке двічі зустрілось в аналізованому тексті [13, р. 57]. У такий спосіб, завдяки теоретичним новаціям у корпусній лінгвістиці було відкрито шлях до нового розуміння словосполучення в цілому [10, р. 44] й сформульовано теоретичні принципи нового підходу вивчення колокацій. Застосування Дж. Синклером засад контекстуальної теорії значення в лексикографічній практиці (COBUILD)

завершило виокремлення напряму нео-фьорзіанства в сучасному мовознавстві [18, р. 137].

За корпусним підходом феномен колокації розглядається як основний вияв принципу ідіоматичності, який разом з принципом необмеженого вибору зумовлює побудову тексту на рівні структури й значення [13, р. 110]. При цьому під ідіоматичністю, або фразеологічністю, розуміється наявність у будь-якій мові великої кількості блоків фраз або їх «напівфабрикатів», використовуваних мовцем відповідно до індивідуальних переваг [10, с. 60]. На основі узагальнення спостережень корпусних даних Дж. Синклером встановлено характерні ознаки колокацій як вияву ідіоматичності [13, р. 112–113]: 1) невизначеність обсягу великої кількості фраз у мові (set eyes on); 2–4) можливість внутрішньої лексичної (in some cases/in some instances), синтактичної варіативності (not/hardly/scarcely) і зміни порядку слів (recriminate isn't in his nature/isn't in his nature of an academic); 5–6) зумовленість появи високочастотних слів у колокаціях (hard work, hard evidence), тенденції до сумісної вживаності слів у певних граматичних структурах (set about testing) і в певному семантичному оточенні (happen – accident).

Отже, в межах корпусного підходу поняття колокації переосмислюється й певною мірою спрощується: визначальними для синтагматичної зв'язності визнаються статистичні чинники. Саме тому корпусний підхід доцільно класифікувати як статистичний: колокації визначаються як «статистично стійкі словосполучення» [3, с. 4–5], словосполучення із «статистично значимим зв'язком компонентів» [1, с. 203] або навіть «суто статистичний феномен» [8, с. 47]. Зняття обмежень щодо якісного й кількісного складу забезпечує переваги корпусного підходу в розпізнанні колокацій у тексті, виявленні граматично некоректних структур (I can't stand being alone, She bothered to do it) [8, р. 47] і зрештою у встановленні лінгвістично релевантних колокацій. При цьому показниками сили смислового [15] й синтагматичного зв'язку [3, с. 4] між елементами словосполучення визнаються статистичні міри асоціації (MI, score, tscore і loglikelihood). За корпусним підходом, колокації розглядаються як певний континуум, шкалу без чітких меж між вільними й стійкими сполуками [1, с. 206; 3, с. 10; 8, р. 48].

У такий спосіб корпусний підхід уможливив визначення об'єктивних критеріїв для виявлення, опису колокацій і встановлення певного зв'язку між смислом і структурою мови. На протигагу інтерпретації, базованій на лінгвістичній інтуїції, корпусний підхід наближує до «точного наукового розрахунку значення» лексичної одиниці, подібно до того, як у фізиці можна розрахувати рух [18, р. 137]. Відповідно до концепції, колокації слова набувають певного сенсу не в ізольованому положенні, а лише в контексті: значення лексичних одиниць визначається їх колокатами – релевантними словами. Отже, формування певного значення є результатом колокації – поєднання слів у регулярно повторюваному в корпусі контексті [18, р. 137–138]. Перспектива об'єктивного дослідження колокацій здійснюється лише з появою корпусів із статистичним програмним забезпеченням, здатним виявити, які слова, у який спосіб, наскільки часто поєднуються з аналізованою лексичною одиницею, і що можна очікувати від випадкового розподілу. У цьому розумінні кінцева

мета корпусного дослідження – визначення закономірностей і узагальнень, що «зазвичай викладаються у вступі до словників», збігається з метою лексикографічного аналізу [18, р. 138]. Дослідження колокацій статистичними методами покладено в основу виявлення семантико-синтаксичних ознак лексичних одиниць й укладання сучасних корпусних словників [12, р. 28; 8].

Можливість розв'язання проблем опису лексичного значення – «головного болю лексикографів» [13, р. 112], забезпечили вихід корпусної моделі на перший план зарубіжної лексикографії: на сьогодні колокація відіграє роль основної одиниці значення, ужитку й перекладу. Запропонована Дж. Синклером концепція колокації стає для європейських корпусних лінгвістів моделлю, яка обґрунтовує не тільки використання, а й теорію природної мови [10, р. 62, 75].

Однак слід розуміти, що в межах одного підходу або різних часто конкуруючих шкіл формування всеохопної концепції колокації вкрай утруднено [8, р. 47]. Саме тому застосування лише корпусного підходу для опису колокацій, безперечно, не достатньо: виявлені в корпусі характерні моделі вживання лише окреслюють контури теорії мови [18, р. 144]. Водночас поєднання корпусного підходу з традиційними створює перспективи виходу не тільки за межі статистики, а й деяких звичних уявлень про мову [10, р. 53–57]. Об'єктивне вивчення колокацій має ґрунтуватись на багатоаспектному підході, який передбачає використання корпусного підходу як доповнення до традиційних і в жодному разі не як єдино правильного. Це доведено сучасними дослідженнями колокацій, які поєднують корпусний підхід з синтаксичним [12, 6, 2], семантико-синтаксичним [3, 4], когнітивним [10] і психо-лінгвістичним [11].

У цьому розумінні підґрунтям для укладання словника колокацій має бути відповідний лексикографічний корпус текстів. Так, в основу розробленого нами підкорпусу законодавчих текстів [19] покладено параметри скінченності обсягу, репрезентативності, збалансованості, лінгвістичної і хронологічної однорідності, анотованості й повноти текстів, що становлять ключові вимоги до корпусних об'єктів. Призначений для укладання Словника колокацій українського юридичного дискурсу підкорпус є складовою Корпусу текстів української мови й визначається фіксованим обсягом у 1 157 млн слововживань, що дає змогу заощадити зусилля й детермінувати поріг відображення предметної галузі. Укладання корпусного словника колокацій передбачає формування реєстру за статистичними ознаками: колокація визначається як комбінація слів, що зустрілася в тексті обсягом в 1 млн принаймні двічі [10, р. 104].

Вибір як документального джерела «Зібрання законодавства України» зумовлено максимальною насиченістю колокацій у текстах офіційних документів. Укладений підкорпус є одномовним, однорідним в аспекті функціонального стилю і збалансованим за діапазоном 43 жанрових різновидів. Кожен з жанрів представлено найбільшими за обсягом повними текстами, що забезпечує їхню структурну, лексичну й синтаксичну завершеність. З метою дотримання принципу хронологічної однорідності й максимального представлення сучасного юридичного дискурсу хронологічні межі підкорпусу обмежено 1991 р.

Підкорпус інтегровано в інформаційно-пошукову систему Корпусу текстів української мови й забезпечено набором програм морфологічної, синтаксичної та лексико-семантичної розмітки для виявлення й установлення ознак колокацій. Дотримання експліцитних та імпліцитних принципів корпусної побудови надають можливість класифікувати розроблений підкорпус законодавчих текстів як репрезентативний для укладання словника колокацій українського юридичного дискурсу.

**Висновки.** Здійснене дослідження дає змогу зробити такі висновки: 1. Корпусний підхід базується на вузькому значенні поняття колокації як послідовності слів, що зустрічаються в тексті частіше поруч, ніж за випадковою вірогідністю окремо. 2. Концептуальними джерелами корпусного підходу доцільно визнати контекстуальну теорію значення Дж.Р. Фьорза й теорію елементарних полів В. Порцига. 3. Історичними передумовами формування корпусного підходу слід вважати появу електронних корпусів й удосконалення статистичного опрацювання великих масивів текстів. 4. За корпусним підходом колокації розглядаються як певний континуум між вільними й стійкими сполуками. 5. Основними ознаками колокацій слід вважати такі: регулярність вживання, статистично значиму зв'язність компонентів, можливість лексичної та синтаксичної варіативності, функціонування в певних граматичних структурах і певному семантичному оточенні. 6. Об'єктивне дослідження феномена колокації вимагає застосування багатоаспектного підходу з урахуванням мовних, статистичних чинників.

#### Література:

1. Левицкий В.В. Семасиология / В.В. Левицкий. – Винница : Нова Книга, 2006. – 508 с.
2. Шкурко В.В. Лексикографічний агент екстракції колокацій у природномовному тексті / В.В. Шкурко // Вісник Київського нац. ун-ту ім. Т. Шевченка. Серія «Літературознавство. Мовознавство. Фольклористика». – 2012. – № 28. – С. 31–35.
3. Хохлова М.В. Исследование лексико-синтаксической сочетаемости в русском языке с помощью статистических методов (на базе корпусов текстов) : автореф. дис. ... канд. филол. наук : спец. 10.02.21 «Прикладная и математическая лингвистика» / М.В. Хохлова. – СПб., 2010. – 26 с.
4. Дарчук Н. Комп'ютерне анування українського тексту : результати і перспективи : [монографія] / Н. Дарчук. – К. : Освіта України, 2013. – 544 с.
5. Хохлова М.В. Экспериментальная проверка методов выделения коллокаций / М.В. Хохлова // Slavica Helsingiensia 34. Инструментарий русистики : корпусные подходы. – 2008. – № 34. – С. 343–357.
6. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика : [учеб. пособ.] / [Большакова Е.И., Клышинский Э.С., Ландэ Д.В. и др.]. – М. : МИЭМ, 2011. – 272 с.
7. Романюк А. Розпізнавання багатослівних конструкцій / А. Романюк, Г. Кваснюк, М. Романишин // Вісник Нац. ун-ту «Львівська політехніка». Комп'ютерні системи проектування. Теорія і практика. – 2011. – № 711. – С. 158–165.

8. Fontenelle Th. What on earth are collocations? / Th. Fontenelle // English today : the International Review of the English Language. – Cambridge : Cambridge University Press, 1994. – Vol. 10 (40), No 4. – P. 42–48.
9. Ягунова Е.В. Исследование контекстной предсказуемости единиц текста с помощью корпусных ресурсов / Е. В. Ягунова // Труды междунар. конф. «Корпусная лингвистика–2008» ; 6–10 окт. 2008 г. – СПб., 2008. – С. 396–403.
10. Marcinkevičienė R. Lietuvių kalbos kolokacijos : [monografija] / R. Marcinkevičienė. – Kaunas : Vytauto Didžiojo universiteto leidykla, 2010. – 212 p.
11. Durrant Ph. Are high-frequency collocations psychologically real? / Ph. Durrant, A. Doherty // Corpus Linguistics and Linguistic Theory. – Berlin–New York : De Gruyter, 2010 – Vol. 6, No 2. – P. 125–155.
12. Seretan V. Syntax-Based Collocation Extraction / Violeta Seretan. – Berlin : Springer Science & Business Media, 2011. – 232 p.
13. Sinclair J. Corpus, Concordance, Collocation (describing English Language) / J. Sinclair. – Oxford : Oxford University Press, 1991. – 200 p.
14. Гладка В.А. Структурно-синтаксичний підхід у вивченні колокацій (на матеріалі французької мови) / В.А. Гладка // Наукові записки нац. ун-ту «Острозька академія». Серія «Філологічна». – 2013. – Вип. 39. – С. 16–20.
15. Ягунова Е.В. От коллокаций к конструкциям / Е.В. Ягунова, Л.М. Пивоварова // Acta linguistica petropolitana. Труды Института лингвистических исследований РАН. – СПб., 2011.
16. Зацеркляний М.М. Об'єктно-орієнтований тезаурус і словник колокацій для бази знань криміналістичних інформаційних систем / М.М. Зацеркляний, Д.Ю. Узлов // Системи обробки інформації. – 2013. – Вип. 2. – С. 183–186.
17. Tognini-Bonelli E. Corpus Classroom Currency / E. Tognini-Bonelli // Darbai ir Dienos. – 2000. – No 24. – P. 205–243.
18. Teubert W. Linguistique de Corpus: un Alternative / W. Teubert // Semen. Critical Discourse Analysis I. – Besançon : PUFC, 2009. – T. 27. – P. 130–152.
19. Законодавчі тексти [Електронний ресурс].– Режим доступу : <http://www.mova.info/corpus2.aspx>.

#### Бобкова Т. В. Концепция коллокации: корпусный подход

**Аннотация.** В статье определены концептуальные источники и основные положения корпусного подхода к изучению коллокаций, обоснована необходимость использования комплексного подхода на основе корпуса текстов.

**Ключевые слова:** коллокация, словосочетание, текст, контекстно-ориентированный подход, корпусный подход, статистический подход.

#### Bobkova T. Conception of collocation: corpus-based approach

**Summary.** The article is to outline the corpus-based approach toward collocation studies. The conceptual basis and main principles of corpus-based approach are determined. A necessity of combining approach use in collocation study is proved.

**Key words:** collocation, phrase, text, context-based approach, corpus-based approach, statistical approach.