

## ТИПИ ЛІНГВІСТИЧНИХ МОДЕЛЕЙ ТА ЇХ ЗАСТОСУВАННЯ ДЛЯ РОЗВ'ЯЗАННЯ ЛІНГВІСТИЧНИХ ЗАДАЧ

© Лукач М.О., 2013

**Окреслено розуміння понять «модель» та «моделювання» у лінгвістиці. Розглянуто основні типи лінгвістичних моделей та їх застосування для розв'язання різноманітних лінгвістичних задач.**

**Ключові слова:** моделювання, модель, породжуюча модель, дослідницька модель, аналітична модель, синтетична модель.

**This paper outlines understanding of the notions 'model' and 'modeling' in linguistics. The main linguistic models and their application to solving different linguistic tasks are described.**

**Key words:** modeling, model, generative model, research model, analytical model, synthetic model.

### Вступ. Загальна постановка проблеми

За останнє десятиліття спостерігаємо бурхливий розвиток прикладної лінгвістики у зв'язку з постійним зростанням потреб у застосуванні механізмів природної мови в інформаційно-комп'ютерних та людино-машинних системах [3, с. 3].

Особливого значення набули питання моделювання природної мови та мовлення. Це спонукає науковців до розроблення різноманітних лінгвістичних моделей, що могли б розв'язати практичні завдання лінгвістики, а саме: інформаційний пошук, машинний переклад, розуміння природної мови, екстракція та локалізація знань з текстів тощо.

### Аналіз останніх досліджень та публікацій

Однією з важливих лінгвістичних моделей є розроблена І. А. Мельчуком модель «Смисл □ Текст». Хоча цю багаторівневу модель перетворення смислу на текст і навпаки створено ще в другій пол. ХХ ст., проте вона і нині продовжує досліджуватись і застосовується для задач машинного перекладу. Крім того, розмітка синтаксичного підкорпусу Національного корпусу російської мови виконується лінгвістичним процесором ЕТАП-3, що базується на принципах моделі «Смисл □ Текст» [8].

Також слід згадати про граматичну модель HPSG (Head-driven phrase structure grammar) – це породжуюча граматична теорія, яку розробили К. Поллард та І. Саг на початку 1980-х рр. [9, с. 44]. З цього часу написано різні парсери на основі HPSG формалізму. Граматики HPSG різних мов у наш час розробляють у межах проекту (DELPH-IN).

### 1. Поняття «модель» та «моделювання» у лінгвістиці

В лінгвістичній літературі термін «модель» вперше вжив американець З. Херріс у 1944 р., характеризуючи відмінність між методологічними прийомами двох лінгвістів С. Х'юмена та Е. Сепіра, а потім у 1951 р. він ще раз вжив цей термін для визначення результатів описової термінології Е. Сепіра. Проте в більш специфічному значенні (а саме для застосування в граматиці) термін «модель» вжив Ч. Хокетт у 1954 р., а також Н. Хомський у 1956 р. Останні визначають **модель** як узагальнену та формалізовану структуру або процес тих чи інших фактів мови. А. Етгінджер у 1957 р. вже говорить про моделі не тільки як про відтворення дійсності, але і як про вплив на цю дійсність. Ще один лінгвіст В. Інґве поширює поняття моделі з конкретних мов на механізм мови загалом [5, с. 14].

Поняття «модель» широко використовується у різних галузях знань і відповідно відрізняються його визначення залежно від конкретного застосування. Розглянемо деякі підходи до розуміння цього терміна у мовознавчій науці.

**Модель** (франц. *modèle*, від лат. *modulus* – міра) в лінгвістиці – 1. Штучно створений лінгвістом реальний або уявний пристрій, що відтворює, імітує своєю поведінкою (зазвичай у спрощеному вигляді) поведінку якогось іншого («справжнього») пристрою (оригіналу) в лінгвістичних цілях. 2. Зразок, що слугує стандартом (еталоном) для масового відтворення; те саме, що і «тип», «схема», «парадигма», «структура» і т. д. (наприклад, «модель дієвідміни або відміни», «словотвірна модель», «модель речення» і т. д.) [4, с. 304].

В іншому словнику [2, с. 238] подане подібне визначення. **Модель** (англ. *pattern, model*) – 1. (схема) закономірне розташування послідовностей сегментних і надсегментних елементів в складних мовних одиницях, що характеризується властивостями даної мовної структури та регулярно відтворюється в мовленні (наприклад, модель речення, словотвірна модель). 2. властива даній мові парадигматична схема елементів (складових) її структури; компактно символічне зображення цієї схеми (або будь-яких її частин чи елементів), що є основою моделювання мови як одного із методів її вивчення (наприклад, породжуюча модель).

**Моделювання** – науковий прийом, що полягає у схематичному відтворенні об'єкта, який не піддається безпосередньому спостереженню або є дуже складним [2, с. 238]. Необхідність у моделюванні виникає у всіх тих наукових областях, де об'єкт науки недоступний для безпосереднього спостереження. У таких випадках він зазвичай уподібнюється якійсь «чорній скриньці», про яку відомо лише, які вихідні матеріали вона отримує «на вході» та які кінцеві продукти видає «на виході». Завдання, отже, полягає у тому, щоб дізнатися вміст «чорної скриньки» – той прихований від дослідника механізм, що здійснює перероблення початкових матеріалів на кінцеві продукти. Оскільки неможливо розібрати «чорну скриньку», не порушивши її функціонування, то залишається єдиний шлях до пізнання об'єкта: слід на основі зіставлення вихідних та кінцевих даних побудувати образ об'єкта, тобто висунути гіпотезу про його можливу будову та реалізувати її у вигляді логічної машини, яка може переробляти певний матеріал точно так само, як це робить сама «чорна скринька». Якщо наш логічний пристрій працює так, то він є апроксимацією або моделлю об'єкта, і можна вважати, що закладений в нього механізм у всіх суттєвих деталях збігається з механізмом, що міститься в «чорній скриньці». В результаті можна зрозуміти механізм, недоступний нашому спостереженню, уподібнюючи його до механізму, будова якого нам добре відома. Отже, як зазначає Ю. Д. Апресян, зміст **моделювання** полягає в тому, щоб замість прихованих від нас властивостей об'єкта вивчити ті властивості моделі, що вже задані в явному вигляді і поширити на об'єкт всі ті закони, які виведені для моделі [1, с. 78–79].

Інший російський лінгвіст [6, с. 8] визначає поняття «**моделювання мови**» як метод, за яким спочатку на основі певних найзагальніших рис конкретних мов формулюються деякі гіпотези про будову мови як абстрактну семіотичну систему, а потім встановлюється відношення між результатами цих гіпотез та фактами реальних мов, що описуються конкретними лінгвістичними дисциплінами.

Виділяють такі основні властивості моделей [1, с. 79–99]:

1. Моделювати можна лише такі явища, суттєві властивості яких вичерпуються їх структурними (функціональними) характеристиками і ніяк не пов'язані з їх фізичною природою. Моделлю об'єкта, для якого істотними є тільки його функціональні властивості, слід вважати будь-який пристрій, функціонально схожий на нього. Інакше кажучи, поведінка моделі повинна бути схожа на поведінку об'єкта; матеріал, в якому вона реалізована, може відрізнитися (і практично майже завжди відрізняється) від матеріалу, з якого побудований об'єкт.

Всі інші властивості моделі пов'язані з її основною властивістю, тобто з тим, що вона є функціональною апроксимацією об'єкта.

2. Модель завжди є деякою ідеалізацією об'єкта. Щоб зрозуміти складні явища, треба почати з вивчення найпростіших і загальних випадків, навіть якщо вони ніколи не зустрічаються в чистому вигляді, і від них просуватися до більш складних і спеціальних випадків. Найважливішим

прикладом ідеалізації лінгвістичного об'єкта є припущення, що кількість речень в мові нескінченна і, що довжина речень в принципі нічим не обмежена (тобто можливі речення, довжина яких буде більша за будь-яку наперед задану). Очевидно, що фактична кількість написаних або вимовлених речень будь-якою мовою велика, але скінченна; однак, щоб пояснити здатність мовця будувати нові, раніше ніколи не вимовлені і не написані речення, необхідно розглянути не цю реально спостережувану ситуацію, а деяку ідеальну ситуацію.

3. Як правило, модель оперує не поняттями про реальні об'єкти, а поняттями про ідеальні об'єкти, так звані конструктами, не виведеними безпосередньо і однозначно з даних, а побудованими «вільно» на підставі деяких загальних гіпотез, підказаних сукупністю спостережень і дослідницькою інтуїцією. Будь-яка модель є конструкцією, логічно виведеною з гіпотез за допомогою певного математичного апарату. В ідеалі лінгвістичні конструкти побудовані без безпосереднього звернення до фонетичної або семантичної субстанції тих явищ, для вивчення яких вони створені.

4. Будь-яка модель повинна бути формальною. Модель вважається формальною, якщо в ній однозначно та у явному вигляді задані вихідні об'єкти, що зв'язують їх твердження та правила поведінки з ними (правила утворення або виділення нових об'єктів і тверджень). В ідеалі будь-яка формальна модель є математичною системою, тому в певному сенсі поняття формальності рівнозначно поняттю математичної точності або однозначності. Формальна модель зв'язується з даними за допомогою тієї чи іншої інтерпретації. Дати інтерпретацію моделі – значить вказати правила, імовірнісні або строгі, підстановки об'єктів деякої предметної області, наприклад мови, замість об'єктів (символів) моделі.

5. Будь-яка інтерпретована модель повинна мати властивість пояснювальної сили. Модель володіє цією властивістю, якщо вона: 1) пояснює факти або дані спеціально поставлених експериментів, які не можна пояснити за допомогою старої теорії; 2) передбачає не відому раніше, але принципово можливу поведінку об'єкта, яка пізніше підтверджується даними спостереження або новими експериментами. В обох випадках пояснювальна сила моделі збільшується за вищої міри збігу передбачень з експериментальними даними.

Основним поняттям будь-якої моделі мови можна вважати поняття тексту в певному алфавіті. Текст слід розуміти як реально вимовлений текст («множина правильних послідовностей мови») або як якість його узагальнення, наприклад послідовність класів, що відповідають елементам реального тексту [7, с. 17].

Побудова моделі передбачає такі етапи [1, с. 99]:

1. Фіксування фактів, що потребують пояснення.
2. Висування гіпотез для пояснення фактів.
3. Реалізація гіпотез у вигляді моделей, що не тільки пояснюють вихідні факти, але і передбачають нові факти, які раніше не спостерігалися.
4. Експериментальна перевірка моделі.

## 2. Основні типи лінгвістичних моделей

Типи лінгвістичних моделей розглядають на основі різних класифікаційних ознак. Так, за характером розглянутого в них об'єкта лінгвістичні моделі поділяються на [1, с. 99–100]:

1. Моделі, в яких об'єктом виступають конкретні мовні процеси і явища, які імітують мовленнєву діяльність людини. Вперше їх розробили празькі структуралісти.

2. Моделі, в яких як об'єкт розглядаються процедури, що ведуть до виявлення того чи іншого мовного явища. Ці моделі певною мірою імітують дослідницьку діяльність лінгвіста і називаються дослідницькими моделями. Перший серйозний крок в їх розробленні зробили американські дескриптивісти, що намагались створити універсальні алгоритми відкриття граматик природних мов.

3. Моделі, в яких як об'єкт розглядаються вже готові лінгвістичні описи, а не мовленнєва діяльність людини або дослідницька діяльність лінгвіста. Якщо звичайна граматики є теорією конкретних мовних процесів, то модель третього типу є теорією теорії або метатеорією. Першими моделі третього типу розробляли глосематики.

Такі метамоделі містять систему критеріїв і теоретичних доказів (метамова), за допомогою яких з декількох альтернативних моделей, що моделюють те саме явище, можна вибрати кращу. У математичній лінгвістиці розробляють метамоделі (метатеорії), об'єктами яких є не окремі лінгвістичні поняття, а цілісні моделі мови [4, с. 304].

Залежно від того, який зміст вкладається в поняття «володіння мовою» (чи входить до нього ознака володіння значенням слів, чи ні), моделі мовленнєвої діяльності можна поділити на: 1) несемантичні, або суто синтаксичні, що імітують володіння граматиною, тобто здатність носіїв мови розуміти і будувати граматично правильні, але не обов'язково осмислені фрази; 2) семантичні, що імітують здатність носіїв мови розуміти і будувати осмислені речення. Залежно від того, яка сторона мовленнєвої діяльності – слухання або говоріння – є предметом моделювання, моделі мовленнєвої діяльності поділяють на моделі аналізу і моделі синтезу. Моделлю аналізу називається скінченна кількість правил, здатних проаналізувати нескінченну кількість речень певної мови. Синтаксичні аналітичні моделі отримують на «вході» текст, а на «виході» видають для кожного речення запис його синтаксичної структури. Семантичні аналітичні моделі отримують на «вході» той самий матеріал, а на «виході» видають смисловий запис (зображення сенсу) кожного речення спеціальною семантичною мовою. Моделлю синтезу називається скінченна кількість правил, здатних побудувати нескінченно велику кількість правильних речень. Синтаксичні синтетичні моделі використовують як вихідну інформацію запис синтаксичної структури речень, а на виході видають правильні речення мови. Семантичні синтетичні моделі отримують на «вході» смисловий запис деякого речення спеціальною семантичною мовою і видають на «виході» множину речень природної мови, синонімічних цьому реченню (табл. 1) [1, с. 106–107].

Таблиця 1

#### Основні типи лінгвістичних моделей

Ознаки	Що відомо лінгвісту	Характер вихідної інформації	Характер кінцевої інформації	Мета
Тип моделі				
Дослідницькі	Текст (і множина правильних фраз)	Текст	ГраMATика і словник	Змодельовати діяльність лінгвіста
Аналітичні	ГраMATика і словник	Текст	Відображення структури тексту	Змодельовати розуміння тексту
Синтетичні	ГраMATика і словник	Відображення структури тексту	Текст	Змодельовати створення тексту
Породжуючі	ГраMATика і словник	Алфавіт символів і правила створення і перетворення фраз	Множина правильних фраз і відображення їх структури	Змодельовати вміння відрізнити правильне від неправильного в мові

З поняттям аналітичних і синтетичних моделей пов'язане поняття зворотності моделі. Модель M1 називається зворотною щодо моделі M2, якщо вихідні об'єкти M1 є кінцевими об'єктами M2, а кінцеві об'єкти M1 – вихідними об'єктами M2. Деякі дослідники розглядають синтетичні моделі як зворотні щодо аналітичних. За такого підходу немає потреби в побудові двох самостійних моделей певного явища: аналітичну модель можна отримати простим зверненням до синтетичної, і навпаки. Крім моделей аналізу і синтезу, існують ще так звані породжуючі моделі, в деякому розумінні проміжні між моделями аналізу і синтезу. Породжуючою моделлю називається конструкція, що містить алфавіт символів і скінченну кількість правил утворення (і перетворення) виразів з елементів цього алфавіту, здатна побудувати нескінченну кількість правильних речень даної мови і приписати кожному з них деяку структурну характеристику [1, с. 106–107].

Залежно від того, в якій математичній формі викладена модель, моделі поділяються на числення і алгоритми. Числення – це система дозволів (дозволу), а алгоритм – це послідовність

наказів (команд). Зазвичай числення має вигляд математичної системи, яка охоплює: 1) вихідні (первинні або не визначені) поняття, імена яких утворюють «алфавіт символів»; 2) первинні (не доведені) твердження про зв'язки між цими поняттями (аксіоми); 3) правила виведення нових тверджень (теорем) з уже наявних. Замість аксіом і правил виводу іноді використовуються правила створення і перетворення виразів з елементів алфавіту. У численнях часто використовуються так звані рекурсивні визначення і правила. Рекурсивними називаються визначення і правила, які будуються за два кроки, причому перший крок містить визначення найпростішого окремого випадку, а другий – визначення загального випадку через окремих. Числення дає змогу задати за допомогою кінцевого апарату всі об'єкти деякої множини, зокрема нескінченної (наприклад, всі речення певної мови). Ця властивість числень і повинна використовуватись у лінгвістиці, яка має справу з дуже великою або нескінченною кількістю одиниць. Алгоритмом називається послідовність команд, виконання яких веде до виділення (або побудови) бажаного об'єкта. Алгоритм повинен допускати абсолютно автоматичну реалізацію. Алгоритм, записаний зрозумілою для машини мовою, називається програмою. Будь-яку модель, зокрема числення, необхідно представити у вигляді алгоритму (або забезпечити алгоритмом), щоб реалізувати на машині, тому що машина розуміє тільки мову команд, а не мову дозволів. Дослідницькі, аналітичні та синтетичні моделі оформляються зазвичай у вигляді алгоритмів, а для викладу породжуючих моделей, як правило, використовується форма числень [1, с. 108–110].

Залежно від того, якого роду правила використовуються в моделі, розрізняють імовірнісні (статистичні) та детерміністські (структурні) моделі. Існують також змішані структурно-статистичні моделі. Природні мови здебільшого влаштовані так, що невелика кількість правил охоплює основну множину фактів, але для пояснення небагатьох фактів, що залишилися, здебільшого непродуктивних, потрібна дуже велика кількість правил. Тому інколи вигідніше пояснити певну сукупність фактів не детерміністською моделлю, яка через велику кількість правил може виявитися надмірно громіздкою для виконання деякої цілком визначеної задачі, а ймовірнісною моделлю, якій потрібно менше суто статистичних правил і тому вона є менш громіздкою. Втрата точності правил компенсується в такій моделі її відносною простотою. Найважливішими детерміністськими моделями є моделі бінарних диференційних структур в області фонології і морфології, модель безпосередніх складових, трансформаційна та аплікативна моделі в області синтаксису, модель «семантичних множників» у області семантики [1, с. 112].

### **Формулювання мети статті**

Мета статті – окреслити підходи до визначення поняття «модель» у лінгвістиці, а також здійснити огляд основних типів лінгвістичних моделей та їх застосування для розв'язання різноманітних лінгвістичних задач.

### **Виклад основного матеріалу**

#### **1. Дослідницькі моделі**

Дослідницькі моделі можна поділити на три класи залежно від того, яка інформація слугує в них вихідною. У моделі першого класу як початкова інформація використовується тільки текст, і всі відомості про систему, тобто мову, що породжує цей текст, видобуваються виключно з текстових даних. Це приклад класичних дешифрувальних моделей. У моделях другого класу заданим є не тільки текст, але і множина правильних фраз мови. Це означає, що, розробляючи моделі, лінгвіст вдається до допомоги інформанта, який з приводу кожної запропонованої йому фрази повинен говорити, правильна вона чи ні. Інформантом може бути і сам лінгвіст, якщо він досконало володіє мовою, що вивчається. Нарешті, в моделях третього класу задані не лише текст і множина правильних фраз, але і множина семантичних інваріантів. Це означає, що інформант повинен визначати не тільки правильність кожної запропонованої йому фрази, але й про будь-які дві фрази говорити, чи означають вони те саме, чи ні [1, с. 101].

### 1.1. Дешифрувальні дослідницькі моделі

Перший тип дослідницьких моделей називається моделями лінгвістичного дешифрування, оскільки, крім своєї основної теоретичної функції (обґрунтування лінгвістичних понять і тверджень), вони, в принципі, можуть мати і деяку прикладну – дешифрувальну функцію. Вихідною інформацією для моделей цього типу є текст, про який заздалегідь нічого не відомо. Невідома ні мова (код), використана для «шифрування» тексту, ні генетичні зв'язки цієї мови з уже відомими мовами, ні переклади тексту на відомі мови, ні та область дійсності, яку описує текст. Всі інші відомості, тобто відомості про елементарні одиниці тексту (букви або звуки, морфеми, слова, речення і, нарешті, сенси), класи елементарних одиниць (голосні й приголосні, лексичні та граматичні морфеми, частини мови, типи речень, семантичні поля) і закони поєднання одиниць різних класів (наприклад, синтаксичних зв'язків слів у реченні) повинні бути абсолютно автоматично отримані з тексту [1, с. 120].

Прикладами таких моделей є алгоритми Б. В. Сухотіна, З. Харріса. Окреслимо деякі загальні риси цих алгоритмів. По-перше, в основу всіх таких алгоритмів покладено прості та загальні уявлення про мову, що підтверджуються певними універсальними закономірностями, наприклад: «буква – це стійке поєднання точок»; «морфема – це стійке поєднання фонем»; «словоформа – це стійке поєднання морфем»; «в кожній природній мові є як мінімум два рівні – рівень значущих одиниць (морфем, словоформ, конструкцій) і рівень незначущих одиниць (фонем)»; «в будь-якій мові є лексичні морфеми, причому розподіл лексичних морфем в тексті відрізняється від розподілу граматичних морфем»; «синтаксичним відмінностям відповідають семантичні відмінності»; «близькі за змістом слова стоять в тексті недалеко один від одного» і т. п. По-друге, у всіх алгоритмах такого роду використовується інформація про дистрибуцію елементів та їх числових параметрів. По-третє, зазвичай спочатку визначається множина допустимих рішень, а потім у цій множині за допомогою так званих функцій вигідності знаходять найкраще рішення. Функціями вигідності називають числові функції, які в разі правильних (найкращих) рішень набувають певного (наприклад, мінімального можливого або максимально можливого) значення. Кожна функція вигідності формалізує деяку змістовну гіпотезу про можливі властивості досліджуваного об'єкта. Алгоритм виділення голосних і приголосних Б. В. Сухотіна припускає, що вже виділено алфавіт літер, за допомогою якого записаний цей текст. Його завдання полягає в тому, щоб надати ключ до правил читання тексту, без яких дешифрування не можна вважати закінченим. Допустимим рішенням вважається будь-яке розбиття алфавіту на два класи. Алгоритм формалізує таку змістовну гіпотезу про властивості голосних і приголосних букв: голосні і приголосні в тексті чергуються; не існує текстів, що складаються з одних голосних, як і не існує текстів, що складаються з одних приголосних; за голосними слідує приголосні, а за приголосними – голосні [1, с. 122–123].

Іншим прикладом дешифрувальної дослідницької моделі є морфологічний алгоритм З. Харріса. Алгоритм працює так, що спочатку у тексті, записаному у вигляді фонологічної транскрипції, вибирає деяке речення, наприклад англ. [Hiyzklever] – «він розумний». Потім шукають всі речення, що починаються з тієї ж фонем [h], і підраховують кількість різних фонем, які слідує за [h] в цих реченнях. Такі фонем називаються «наступниками» (successors) [h]. Кількість «наступників» для [h] дорівнює 9. Після того знаходять всі речення, які починаються з тих самих двох фонем, що і в цьому реченні, тобто з [hi], та підраховують кількість «наступників» для цієї пари фонем (воно дорівнює 14) і т. д. У різних місцях речення кількість «наступників» змінюється: то зростає, утворюючи піки, то зменшується. Якщо вважати правильним припущення, що фонологічна різноманітність на стику морфів є більшою, ніж у середині морфів, то піки повинні траплятися на межі між морфами (табл. 2).

Таблиця 2

#### Приклад застосування морфологічного алгоритму З. Харріса

h	i	y	z	k	l	e	v	Ə	r
9	14	29	29	11	7	...	...	...	...

У цьому реченні є два піки – після фонемі [y] і після фонемі [z], саме у цих місцях і слід провести морфологічні границі. Отримана сегментація (he/is/clever) відображає дійсний стан речей [1, с. 127–128].

## 1.2. Експериментальні дослідницькі моделі

Експериментальні дослідницькі моделі ставлять ту саму мету, що і дешифрувальні: у них теж ставиться завдання переходу від тексту до «системи», тобто завдання отримання за текстовими даними відомостей про елементарні одиниці мови, класи елементарних одиниць, закони поєднання елементів різних класів тощо. Відмінність між цими типами моделей полягає у методі отримання вироблюваної інформації та в характері вихідної інформації. Якщо в дешифрувальних моделях лінгвіст обмежує себе чистим спостереженням, фіксуючи поведінку певного об'єкта (фонемі, морфемі, конструкції) у природних умовах тексту і дані спостережень проходять тільки математичну обробку, то в експериментальних дослідницьких моделях він спостерігає матеріал і в штучних умовах, піддаючи його всякого роду перетворенням. Тобто в моделях другого типу він використовує експеримент, неможливий у межах дешифрувальних моделей. Для того, щоб мати можливість проводити експерименти, лінгвістові недостатньо мати у своєму розпорядженні текст, тому що всякий реальний текст має обмежену протяжність і, отже, не містить усіх принципово можливих у цій мові фраз. Крім тексту, лінгвіст повинен мати множину правильних фраз мови, в принципі нескінченну. Інакше кажучи, для побудови експериментальної моделі лінгвіст повинен мати можливість звернутися до інформанта, який з приводу кожної пред'явленої йому фрази повинен говорити, чи належить вона до фраз на його мові, чи ні (можна так сказати чи ні) [1, с. 149].

Прикладом цього типу моделей є модель за безпосередніми складовими, яку розробили Л. Блумфільд, Р. Уелс, З. Харріс як експериментальну дослідницьку модель, але яку згодом почали широко застосовувати в моделюванні мовленнєвої діяльності.

Як бачимо, на відміну від дешифрувальних дослідницьких моделей, експериментальні дослідницькі моделі є гнучкішими і за допомогою них можна отримати більше інформації про одиниці мови з вихідного тексту.

## 2. Моделі мовленнєвої діяльності

Загалом моделі мовленнєвої діяльності поділяються на породжуючі моделі, синтаксичні моделі аналізу та семантичні моделі аналізу.

### 2.1. Породжуючі моделі

Породжуюча модель, яка зазвичай представлена у формі числення, це кінцевий набір правил, здатних задати або породити всі правильні, і тільки правильні, об'єкти деякої множини, зокрема нескінченної, і приписати кожному об'єкту певний аналіз [1, с. 198]. Розглянемо три типи синтаксичних породжуючих моделей:

#### 1. Модель породження за безпосередніми складовим

Прикладом моделі породження за безпосередніми складовими є модель В. Інґве. Модель складається з: 1) граматики; 2) механізму. Граматика – це кінцева нерегульована множина правил розгортання за безпосередніми складовими таких типів: а)  $A \rightarrow B + C$  (символ  $U$  лівій частині правила розгортається в два нові символи, наприклад  $S \rightarrow NP + VP$ , де  $NP$  – група підмета, а  $VP$  – група присудка. б)  $A \rightarrow B$  (символ в лівій частині правила замінюється одним, зазвичай термінальним символом, тобто символом, до якого не застосовні ніякі інші правила граматики безпосередніх складових), наприклад:  $T \rightarrow the, a$  (означений або неозначений артикль); в) у третю групу входять правила типу  $B \rightarrow D + \dots + E$ , призначені для породження конструкцій з так званими розривними складовими, тобто складовими типу «якщо ..., то», між частинами яких можуть міститися інші складові. Другою частиною моделі є механізм – ідеалізована, але фізично реалізована обчислювальна машина, що складається з чотирьох взаємопов'язаних частин: пристрій виведення (друкує по одному елементу, що виводяться), вирішальний пристрій (в який поміщається один символ, що розгортається в цей момент за одним з описаних вище правил), постійна пам'ять (в ній

зберігаються всі правила граматики, тобто правила розгортання за безпосередніми складовими), швидкодіюча пам'ять (в ній зберігаються проміжні результати) [1, с. 199–201].

## 2. Трансформаційна породжуюча модель

Адекватна граматична модель повинна містити кілька рівнів опису, відповідних рівням структури мови, причому кожен рівень опису повинен являти собою специфічний набір правил. Трансформаційна породжуюча граматики містить три рівні опису, кожен з яких задається частково впорядкованою множиною правил певного виду і є одним з компонентів трансформаційної породжуючої граматики. До таких компонентів належать: 1) правила за безпосередніми складовими; 2) трансформаційні правила і 3) морфологічні правила. Вихідними для трансформаційної породжуючої граматики є правила розгортання за безпосередніми складовими. Вони поділяються на контекстовільні (застосовуються завжди) і контекстозалежні (використовуються лише тоді, коли лівий символ правила перебуває в певному контексті). Ці правила являють собою правила підстановки виду  $A \rightarrow B + C$ , на які накладені такі обмеження [1, с. 213–216]:

1. Кожне правило застосовується до ланцюжка певного зовнішнього вигляду незалежно від того, як цей ланцюжок отриманий. Так, ті самі правила розгортання символу NP (групи іменника) застосовні і в тому випадку, коли NP отриманий за правилом  $S \rightarrow NP + VP$ , і в тому випадку, коли NP отриманий за правилом  $VP \rightarrow V + NP$ . Єдиною необхідною умовою застосування цього правила є наявність в деякому ланцюжку безпосередніх складових символу, який стоїть в лівій частині цього правила.

2. Результатом застосування правил безпосередніх складових до вихідного ланцюжка має бути дерево складових, оскільки трансформації застосовуються саме до дерев, а не до ланцюжків, причому дерево складових має бути приписано цьому реченню цілком однозначно. Тому в компоненті безпосередніх складових трансформаційної граматики забороняється розгортати більше від одного символу одночасно.

3. Заборонені будь-які рекурсивні правила, тобто правила виду  $A \rightarrow A$  чи  $A \rightarrow A + B$ . Рекурсивні правила слід перенести в трансформаційний компонент граматики.

4. З компонента безпосередніх складових трансформаційної граматики повинні бути вилучені і правила виду  $A + B \rightarrow B + A$  (перестановки), які можна витлумачити як дві рекурсії. Ці правила також переносяться в трансформаційний компонент граматики.

До переваг трансформаційних моделей зараховують те, що: 1) трансформаційне числення забезпечує породження речень таких типів, які, мабуть, принципово неможливо вивести за допомогою правил за безпосередніми складовими; 2) трансформаційне виведення деяких складних синтаксичних типів з простих володіє тією перевагою, що описує породжуючий процес і його результат природніше; 3) трансформаційне числення дає змогу встановити досить глибокі зв'язки між синтаксичною структурою речення і його значенням, які не фіксуються правилами за безпосередніми складовими [1, с. 223].

## 3. Аплікативна модель

Аплікативна модель принципово відрізняється від розглянутих вище моделей породження в двох аспектах: своєю логічною структурою і формалізованою в ній лінгвістичною концепцією. Своєрідність логічної структури аплікативної моделі полягає в тому, що породжуючий процес явним чином представлений в ній як той, що відбувається на двох істотно різних рівнях – рівні конструктів і рівні спостереження. Породжуючий процес починається з виведення ідеальних об'єктів – конструктивних аналогів слів і речень, які на другому етапі породжуючого процесу перетворюються за допомогою певних правил інтерпретації на реальні слова і речення певної природної мови. Породжуючий механізм, продуктом якого є ідеальні об'єкти, працює незалежно від правил інтерпретації. Ідеальні об'єкти не містять в собі інформації про граматичні (словотвірні й словозмінні) категорії тієї чи іншої конкретної мови, таких, як: рід, число, відмінок, особа, час, вид тощо. Ці та подібні до них категорії виникають лише при інтерпретації моделі. Тому або сам механізм з наявними в ньому ідеальними словами і правилами створення ідеальних речень, або породжена ним множина ідеальних об'єктів може розглядатися як ідеальна мова, що може, зокрема, використовуватись як мова-посередник для типологічних зіставлень [1, с. 223].



На відміну від моделі за безпосередніми складовими і трансформаційної моделі, в яких об'єктами операцій є ланцюжки або дерева, що перетворюються на нові ланцюжки або дерева, в аплікативній породжуючій моделі розглядаються два види лінгвістичних об'єктів – класи слів і комплекси слів, і відповідно до цього в ній передбачено два різні, хоча і пов'язані один з одним породжуючі механізми – пристрій породження класів слів і пристрій породження комплексів слів. Перший з них моделює відношення між одиницями на парадигматичній осі мови, а другий – відношення між одиницями на синтагматичній осі мови. Лінгвістично аплікативна породжуюча модель відрізняється від трансформаційної та моделі за безпосередніми складовими ще в одному аспекті. У ній використовуються дві різні операції: аплікація, що є єдиним правилом створення об'єктів, та трансформація, що є єдиним правилом їх інваріантного перетворення [1, с. 224].

Алфавіт символів, використовуваних в аплікативній породжуючій моделі, складається із символів чотирьох класів ідеальних основ: N (інтерпретується як клас основ непохідних іменників типу будинок, риба, вікно), V (інтерпретується як клас основ непохідних дієслів типу йти, їхати), A (інтерпретується як клас основ непохідних прикметників типу білий, великий) і D (інтерпретується як клас основ непохідних прислівників типу там, тоді, туди). З цих символів за правилами аплікації утворюються комплекси. Елементарними комплексами називаються символи N, V, A, D, а також будь-яка n-ка (пара, трійка, четвірка або п'ятірка) символів, утворена одноразовим застосуванням будь-якої комбінації правил аплікації до символів N, V, A, D, наприклад великий будинок (AN), похмурий день видніється (ANV) [1, с. 225].

Прикладами аплікативної породжуючої моделі є модель Н. Хомського та модель С. М. Шаумяна на базі принципу двоступінчастості.

Отже, порівняно з моделлю породження за безпосередніми складовими, трансформаційна породжуюча модель має три рівні опису, що дає змогу глибше відобразити структуру мови. Аплікативна породжуюча модель, на відміну від двох попередніх моделей, має інше лінгвістичне підґрунтя, тобто вона за допомогою правил аплікації породжує класи слів та комплекси слів з алфавіту символів чотирьох класів ідеальних основ.

## 2.2. Синтаксичні моделі аналізу

Існують чотири основні підходи до розв'язання задачі автоматичного аналізу тексту (переходу від тексту до синтаксичної інформації, що міститься в ньому): послідовний аналіз, аналіз за допомогою передбачень, пошук опорних точок і метод фільтрів.

Основні принципи методики послідовного автоматичного аналізу тексту сформулював В. Інґве у 1954 р. Пристрій, що аналізує текст, повинен зберігати в пам'яті список (словник) типових для певної мови синтагм, записаних у вигляді послідовностей класів слів. Синтагми розглядаються як еталони; синтаксичну структуру деякого речення шукає машина в результаті послідовного зіставлення різних ланцюжків словоформ речення з еталонами, що зберігаються в словнику, причому в ролі терміна для порівняння кожен раз вибирається максимально довга синтагма. Знайденим словосполученням приписуються номери відповідних синтагм. Після виявлення всіх словосполучень цього речення встановлюються відношення між ними (ієрархія словосполучень), і синтаксичний аналіз речення вважається вичерпаним. Найпослідовніший окреслений підхід до вирішення проблеми автоматичного аналізу тексту наведено в алгоритмі І. А. Мельчука. Процедура аналізу тексту розпадається на дві великі частини – попереднє підготування тексту, що здійснюється допоміжними алгоритмами, і власне аналіз, проведений основним алгоритмом. «Входом» для допоміжних алгоритмів є реальний письмовий текст, а «виходом» – так звані гіпотетичні (вихідні) об'єкти тексту, до яких належать гіпотетичні словоформи (розділові знаки або послідовності букв, обмежені пробілами), гіпотетичні сегменти (послідовності словоформ, обмежені знаками пунктуації або сполучниками) і гіпотетичні фрази (послідовності сегментів, обмежені знаками крапки). «Входом» для основного алгоритму є підготовлений текст, розбитий на гіпотетичні словоформи, сегменти і фрази, а «виходом» – синтаксичне дерево залежностей чи якийсь інший подібний до нього формальний об'єкт, що в явній формі зображає синтаксичну структуру кожного з речень. На тому чи іншому етапі в процесі

переходу від спочатку приготованого тексту до його синтаксичного дерева кожен гіпотетичний об'єкт перетворюється на остаточної або приведеній. Остаточна словоформа – це аналог слова або фразеологічної одиниці; остаточної сегмент – це аналог простого речення (тобто просте речення, дієприкметниковий та інші звороти); остаточної фраза відповідає повному реченню [1, с. 234–241].

Основні принципи автоматичного аналізу за допомогою передбачень викладено у працях Ч. Хоккета, а потім формалізовані І. Родесом і А. Еттінгером. Засобом аналізу синтаксичної структури речення є набір синтаксичних передбачень (НСП), що зберігається в пам'яті машини і містить гіпотетичний перелік можливих в деякому реченні синтаксичних структур. Процес аналізу полягає в послідовному порівнянні слів аналізованого речення з усіма передбаченнями, перерахованими в НСП. Коли для слова знайдено передбачення, то йому приписується позначка відповідної синтаксичної функції, а в НСП вносяться необхідні зміни. Аналіз продовжується доти, доки кожній словоформі речення не буде приписана визначена синтаксична функція [1, с. 242–243].

Метод опорних точок являє собою подальший розвиток ідей, покладених в основу аналізу за допомогою передбачень. Пошук синтаксичної структури речення починається з виявлення його опорних точок. Опорною точкою складного речення вважається головне речення; опорною точкою простого речення – його присудок. Саме ці точки містять максимальну кількість граматичної інформації; їх виявлення дає змогу обґрунтовувати гіпотези (передбачення), що стосуються інших елементів речення. Для того, щоб подібний спрямований пошук синтаксичної структури речення міг бути здійснений, в пам'ять машини повинен бути введений автоматичний словник, в якому кожному слову приписаний певний граматичний код, який представляє морфологічні властивості слова та вказує всі його потенційні синтаксичні можливості (керування, узгодження). Тоді синтаксичний пошук зводиться до встановлення того, які з притаманних слову можливостей фактично реалізуються в цьому реченні [1, с. 243–244].

Метод фільтрів подібний до перших двох методів тим, що в ньому пошук синтаксичної структури речення починається з висунення та перевірки деякої множини гіпотез, проте його принципова відмінність від аналізу за допомогою передбачень та методу опорних точок полягає в тому, що на початку роботи множина гіпотез ніяк не обмежена. Основні принципи методу фільтрів сформульовано в роботах І. Лесерфа, їх розвинула низка інших дослідників, серед яких особливої згадки заслуговує Л. Н. Йорданська. Ці принципи можна підсумувати так: завдання полягає в тому, щоб знайти синтаксичну структуру тексту, тобто зіставити кожне речення, що міститься в ній, з деяким деревом залежностей. Виявлення дерева залежностей починається з розгляду всієї множини допустимих рішень, кожне з яких є гіпотезою про можливі синтаксичні функції словоформ речення. Потім гіпотези перевіряються за допомогою спеціальної програми, і ті, які не виправдовуються, відкидаються. Отже, програма працює як своєрідний фільтр, що затримує всі неправильні рішення. На вихід пропускається найкраща в деякому сенсі гіпотеза, яка і вважається рішенням [1, с. 245].

У вищезазначених підходах кожне речення порівнюють з деяким наперед заданим еталоном. Так за допомогою синтаксичних моделей аналізу з тексту отримують інформація про синтаксичну структуру його речень.

### **2.3. Семантичні моделі аналізу**

Моделювання осмисленої мовленнєвої поведінки людини було предметом вивчення багатьох вчених, зокрема у А. К. Жовківського, Н. Н. Леонтьєвої, Ю. С. Мартемянова, В. Ю. Розенцвейга, Ю. К. Щеглова та ін. Семантичні моделі аналізу імітують здатність людини розуміти і створювати «семантично правильні», тобто осмислені, речення. В цьому випадку моделюється не тільки володіння граматиною мови, але й володіння значеннями слів. Володіння значеннями слів проявляється у мовця в здатності по-різному висловити ту саму думку, а у слухача – у здатності зрозуміти семантичну тотожність зовні різних висловлювань. Отже, існує якась не дана нам у прямому спостереженні семантична мова або «мова думки». Створення осмисленого речення можна представити як переклад з семантичної мови, на природну, а розуміння речення – як переклад з природної мови на семантичну. Значення слова в загальному випадку не є елементарною

семантичною одиницею. Воно поділяється на елементарні смисли, які і є одиницями словника семантичної мови. Та чи інша комбінація елементарних смислів утворює значення деякого слова природної мови. Невелика кількість елементарних смислів дає дуже велику кількість можливих комбінацій, які реалізуються словами природної мови. Одиниці семантичної мови – це ті елементи («атоми»), з різних комбінацій яких складаються «молекули» – значення реальних слів природної мови. У загальному випадку вираз семантичною мовою має вигляд дерева (або дерев) залежності і, отже, семантика може описуватися, як синтаксис [1, с. 252-255].

Отже, семантична мова має свої слова (елементарні смисли) і свою граматику (дерева залежностей). Щоб змодельовати осмислену мовленнєву поведінку людини під час сприйняття тексту (аналіз), необхідно мати: 1) семантичний словник певної мови і 2) алгоритми перекладу з цієї мови на семантичну. Щоб змодельовати осмислену мовленнєву поведінку людини під час створення тексту (синтез), необхідно мати: 1) семантичний словник певної мови і 2) алгоритми перекладу з семантичної мови на цю мову [1, с. 262].

Алгоритм перекладу з природної мови на семантичну поділено на три етапи. На першому етапі кожна частина дерева синтаксичних залежностей, що містить слово, яке позначає відношення і всі залежні від нього слова, переписується за допомогою семантичного словника у вигляді семантичного дерева залежностей. На другому етапі семантичні дерева, побудовані аналогічно для кожної з частин синтаксичного дерева, поєднуються за певними правилами в семантичне дерево всього речення. Отже, на другому етапі все речення цілком перекладається на семантичну мову. На третьому етапі отриманий запис зводиться до стандартного вигляду за допомогою семантичних рівностей. Найпростіший приклад семантичної рівності – правило зняття подвійного заперечення, яке рівносильне твердженню [1, с. 263–264].

Отже, семантичні моделі аналізу за допомогою наперед заданої інформації про граматику та значення слів деякої мови моделюють здатність людини розуміти та продукувати осмислені речення.

### Висновки

У роботі здійснено огляд визначень понять «модель» та «модельовання» у лінгвістиці. Зазначено етапи побудови моделі та її базові властивості. Детально розглянуто основні типи лінгвістичних моделей та їх застосування для розв'язання різноманітних лінгвістичних задач.

Отже, модельовання мовних явищ є сьогодні надзвичайно актуальним і потребує подальших досліджень, оскільки ще є багато невирішених питань.

1. Апресян Ю. Д. *Идеи и методы современной структурной лингвистики (краткий очерк)*. – М.: Просвещение, 1966. – 305 с. 2. Ахманова О. С. *Словарь лингвистических терминов*. – 2-е изд., стер. – М.: Сов. энциклопедия, 1969. – 608 с. 3. *Корпусна лінгвістика: монографія* / В. А. Широков, О. В. Бугаков, Т. О. Грязнухіна, О. М. Костишин, М. Ю. Кригін; НАН України, Укр. мов.-інформ. фонд. – К.: Довіра, 2005. – 472 с. 4. *Лингвистический энциклопедический словарь* / гл. ред. В. Н. Ярцева. – М.: Сов. энциклопедия, 1990. – 683 с. 5. Лосев А. Ф. *Введение в общую теорию языковых моделей* / под ред. И. А. Василенко. – 3-е изд. – М.: Эдиториал УРСС, 2010. – 296 с. 6. Ревзин И.И. *Модели языка*. – М.: АН СССР, 1962. – 192 с. 7. Ревзин И. И. *Современная структурная лингвистика (Проблемы и методы)*. – М.: Наука, 1977. – 264 с. 8. *Синтаксически размеченный корпус русского языка: инструкция пользователя*. [Електронний ресурс]. – Режим доступу : — <http://ruscorpora.ru/instruction-syntax.html> 9. Bolshakov Igor, Gelbukh Alexander. *Computational Linguistics: Models, Resources, Applications. Electronic edition IPN-UNAM-FCE, 2004.* – 198 p.