

СУЧАСНА ДИФЕРЕНЦІАЦІЯ МОВИ

Козак Л. В. Сучасна диференціація мови.

У статті аналізується сучасна диференціація мови, що зумовлено автоматичним аналізом і синтезом наукових текстів, логіко-лінгвістичних моделей у системах керування та виникненням інженерної лінгвістики. Проілюстровано приклади відтворювальних моделей, що використовуються в системах машинного анотування, реферування-перекладу і в машинних діалогових системах штучного інтелекту.

Ключові слова: лінгвістика, кібернетична лінгвістика, автоматичний аналіз, лексичні одиниці, штучна мова, модель, евристичне моделювання, лінгвістичний об'єкт, математична експлікація.

Козак Л. В. Современная дифференциация языка.

В статье анализируется дифференциация языка, которая возникла на стыке языкознания и ряда смежных дисциплин. Автор детально исследует и анализирует новые разделы лингвистики, а также модели, которые используют в системах машинного аннотирования, реферирования-перевода и в машинных диалоговых системах искусственного интеллекта.

Ключевые слова: лингвистика, кибернетическая лингвистика, автоматический анализ, лексические единицы, искусственный язык, модель, эвристическое моделирование, лингвистический объект, математическая экспликация.

Kozak L. V. The modern differentiation of language.

The article analyzes the differentiation of language, due to automatic analysis and synthesis of scientific texts, logical-linguistic models in control systems engineering and

the emergence of linguistics. Substantiates the idea how the new sections of linguistics, which emerged as the result of the interaction of different social needs: the desire of mathematicians to use a language material for the explication of new ideas, the creation of artificial intelligence, heuristic modeling of complex and supercomplex systems. The author gives examples of replicable models that are built under the scheme that includes feedback, which performs the function of a correspondence between engineering and reproduction of linguistic object. These models are used in the machine annotation, summarization-translation and machine dialog systems, artificial intelligence. The most typical engineering-linguistic reproducible models can be called the algorithm of automatic recognition of the meaning of the text. Reproductive linguistic modeling today called engineering to linguistics. There are three types of natural languages that are used in practice in pure and in mixed form. In the simulation, special attention is drawn to the epistemological features of the model and ontological similarity to the original models. Natural speech develops on the basis of associative thinking, therefore meaning the contacts have a logical, emotional and aesthetic nature. The author gives the definition of the model in both linguistic and structural sense.

Key words: linguistic Cybernetics, linguistics, the automatic analysis of lexical units, artificial model, heuristic modeling, linguistic facility, mathematical explication.

Зацікавленість проблемою різноманіття мовних можливостей походить ще з епохи її усвідомлення людиною і не послаблюється впродовж тривалого часу. Останні десятиліття відзначилися інтересом до лінгвістики, якою зараз займаються не лише мовники, але й філософи та математики, соціологи й кібернетики, психіатри і програмісти.

З одного боку, цей інтерес виник у зв'язку з тим, що проблема людини стає однією з центральних для нашої цивілізації, а дослідження мови – це один з найефективніших способів вивчення індивідуальної та колективної поведінки, а також мислення людини. З іншого, – інтерес до мови й лінгвістики стимулюється інтересом до складних систем, зокрема автоматичного аналізу і синтезу наукових текстів, логіко-лінгвістичних моделей у системах керування, інженерної лінгвістики і теорії мови тощо. Звідси, відповідно, актуальність дослідження полягає в тому, що сучасна диференціація мови на лексичному рівні виявляється в неоднорідності її складу з огляду на сферу вживання лексичних одиниць.

Мета статті полягає у з'ясуванні напрямків розвитку сучасної лінгвістики.

Нові розділи лінгвістики, які останнім часом називають то математичними, то структурно-прикладними, то інженерними або ж кібернетичною лінгвістикою, виникли в результаті взаємодії різних наукових стимулів та соціальних потреб. Серед них найбільш активними виявилися такі фактори:

1. Потреби самої лінгвістики, що пов'язувалися з їх задоволенням за допомогою методів лінгвістичного моделювання досліджувати глибинні лінгвістичні об'єкти та зв'язки.

2. Прагнення математиків використовувати мовний матеріал для експлікації нових ідей (математичні ідеї теорії алгоритмів, графів та ін.).

3. Контакт із семіотичними і системними дослідженнями, спрямованими на створення штучного інтелекту, який імітував би мовномисленнєву діяльність людини, а також евристичне моделювання складних і надскладних систем.

4. Потреби інформаційно-обчислювальної індустрії в автоматизації управління промисловістю, транспортом та іншими галузями.

5. Оперативне оброблення науково-технічних і ділових документів, написаних європейськими та іншими мовами.

6. Необхідність інженерно-статистичної оптимізації у викладанні рідної мови та іноземних.

Основним завданням нової лінгвістики є вивчення інженерно-лінгвістичних моделей, які являють собою штучно створені формальні системи, що будуються за принципово іншими схемами: *лінгвістичний об'єкт* → *гіпотеза* → *аналог лінгвістичного об'єкта*. У цій схемі обернений зв'язок між гіпотезою та штучним аналогом, з одного боку, і натурним об'єктом – з іншого, або зовсім відсутній, або здійснюється шляхом перевірки відповідності об'єкта і моделі (об'єкта й гіпотези).

Відтворювальні моделі будуються за чотиричленною схемою: *лінгвістичний об'єкт* → *гіпотеза* → *аналог лінгвістичного об'єкта* → *інженерне відтворення об'єкта*. Ця схема передбачає зворотний

зв'язок, функцію якого виконує відповідність між інженерним відтворенням і самим лінгвістичним об'єктом.

Як найбільш типові інженерно-лінгвістичні відтворювальні моделі (ІЛВМ) можна назвати алгоритми автоматичного розпізнавання смислу тексту, що використовуються в експериментальних або промислових системах машинного анотування, реферування-перекладу і в машинних діалогових системах штучного інтелекту.

Відтворювальне інженерно-лінгвістичне моделювання є найрозвиненішою і найзавершенішою формою мікромоделювання мови і механізмів текстотворення. Той розділ мовознавства, який займається теорією і практикою відтворювального лінгвістичного моделювання, можна назвати інженерною лінгвістикою (ІЛ), про яку у своїх працях писав ще Л. Щерба [4, с. 117–118].

Інженерна лінгвістика виникла і розвивається при поєднанні філологічного аналізу мови, її математичної формалізації і машинного програмування.

Мова – універсальний і потужний засіб спілкування й утілення мислення – становить надскладну і водночас ефективно функціонувальну семіотичну систему. Саме тому представники технічних наук шукають у лінгвістиці зразки для аналізу та проектування складних і надскладних систем. Використання лінгвістичної методики у природничих та технічних науках стане результатом процесу математизації і технізації мови, який диктується потребами управління в «інформатизаційній індустрії». Сьогодні все частіше використовують обчислювальну техніку для оброблення документів.

У зв'язку з обробленням документів машинним способом особливо важливим і актуальним стало завдання переведення документів на мову машини. Машина потребує принципово нового підходу до мови, бо їй необхідна мова формалізована, що являє собою яку-небудь знакову систему, за допомогою якої можна точно і конкретно передати будь-які відомості. У структурі тієї чи тієї мови можна виокремити чотири основні рівні: набір основних символів, слова, вирази (поняття) і речення.

Основні символи мови – це неподільні знаки, за допомогою яких будуються слова. Слова – це найменші смислові одиниці мови, що утворюються шляхом відповідного розташування символів. Із слів утворюються вирази (поняття), а з виразів – речення. В українській мові набором основних символів є алфавіт, який складається з 33 літер і знаків пунктуації.

Функціонує три типи природних мов, що відрізняються кількістю основних символів, семантикою та синтаксисом:

1. Перший тип мов, у яких кожне слово позначається спеціальним символом, наприклад, китайська мова, де загальна кількість символів складає близько 100 000.

2. Другий тип, у яких окремі склади позначаються своїми символами, загальна кількість символів при цьому складає близько 30 000.

3. Третій тип – мови, що мають певні алфавіти мов, із яких утворюються склади і слова, загальна кількість літер складає близько 30.

На практиці використовуються мови змішаних типів. Наприклад, математична мова передбачає мову третього і першого типів. Перший тип – мова формул, де окремі символи позначають цілі слова (наприклад, + – додавання, X – множення). Якщо машинну мову розглядати щодо названих трьох типів, то її необхідно віднести до мови першого типу, де визначені символи – коди операцій – означають виконання певних дій. При порівнянні природної і штучної мов спостерігаємо такі паралелі:

природна мова

слово

поняття

речення

штучна мова

ідентифікатор

набір символів (вираз)

конструкція

Природна мова щодо штучної характеризується надлишковою інформацією. Кожне слово, а інколи й поняття, можна виразити одним символом. Наприклад, слово срібло – Ag; вираз відношення довжини кола до діаметра – π .

Природна мова стихійно розвивалася на основі асоціативного мислення. Тому кожен лінгвістичний знак являє собою перерахування багатьох асоціацій, які формуються (визначаються) смыслом, змістом. Смыслові контакти мають логічну, емоційну й естетичну природу. Асоціативні можливості людського мозку нічим не лімітовані. Тому для кожного лінгвістичного знака наявні нічим не обмежені можливості стати носієм нових асоціацій. Ця властивість знака є джерелом метафоричності природної мови.

Штучно створені формалізовані мови математики будуються на основі логічного аналізу, при якому виокремлюється деяка мінімальна кількість первинних понять, а всі останні визначаються через первинні. Прикладом такої мови є символічна мова математичної логіки. Словник відповідної мови складається з вихідних, нерозкладних символів, кожному з яких однозначно співвіднесений деякий математичний об'єкт або операція. Із вихідних символів складаються кінцеві послідовності, що називаються формулами.

Символи й формули трактують як однозначні математичні знаки. Повністю відкидаються суб'єктивні інтерпретації математичних знаків. У математиці не допускаються також асоціації естетичного й емоційного характеру або зближення знаків за їх формою.

Природна мова сприймається і використовується людиною найчастіше інтуїтивно. Будь-яка математична мова функціонує у вигляді логічному, кожен елемент має усвідомлену значущість. Отже, з одного боку, лінгвістичний знак є семантично багатшим від математичного знака. З іншого, – однозначність математичної мови дає змогу недвозначно й вичерпно формулювати задачі, розв'язувати їх алгоритмічним шляхом, тобто шляхом точного виконання в певному порядку деякої системи операцій.

У зв'язку з тим, що математичний знак однозначний, а лінгвістичний має складну логічну, експресивну і навіть естетичну структуру, то математична інтерпретація поняття пов'язана з розчленуванням лінгвістичного об'єкта й виокремленням у ньому

одного або кількох логіко-сміслових чи формальних компонентів. Останні логічні та формальні елементи лінгвістичного об'єкта, а також експресивно-естетичні відтінки не розглядаються.

Останнім часом набули актуальності методи машинного програмування й автоматичного аналізу тексту.

Лінгвістичні явища найчастіше неможливо прямо спостерігати із-за їх складності, нерозчленованості та відсутності прямої вираженості. Для прикладу візьмемо процес сприймання тексту і його перероблення в мозку людини. Не можна безпосередньо спостерігати семну організацію слова або морфеми, оскільки семи не мають такого вираження, яке безпосередньо сприймалося б органами чуттів людини. У такому разі дослідник змушений використовувати, наприклад, метод моделювання, суть якого полягає в тому, що досліджується не сам об'єкт, а його аналог або модель.

Моделювання є основним прийомом, яким користується інженерна лінгвістика при побудові алгоритмів аналізу та синтезу тексту. У мовознавстві поняття «модель» не має певних обмежень ні з позиції тих властивостей оригіналу, які вивчаються за допомогою моделі, ні у плані фізичної природи аналога [1].

Моделі використовуються для вивчення й опису внутрішньої будови оригіналу (структурної моделі), її поведінки (функціональні моделі) та будови й поведінки оригіналу одночасно (структурно-функціональні моделі). Аналогами лінгвістичного оригіналу є:

- реальні об'єкти (людина може виконувати роль аналога ЕОМ, моделюючи поведінку автомата);
- штучно створені матеріальні об'єкти (електронно-акустичні синтезатори мовних сигналів є моделлю утворення звуків конкретної мови);
- ідеальні математичні або логіко-математичні конструкції (формули, грам-креслення), що використовуються для опису принципів функціонування або представлення структури якоїсь мовної системи.

У процесі породження наукового тексту автор вербалізує абстрактну когнітивну модель, наявну в його свідомості. Якщо автор трансформує модель (образ) у слово, то реципієнт – слово в модель. Причому, як зазначають автори монографії «Когнітивно-комунікативні аспекти наукового тексту», девербалізація створення адекватного образу залежить від автора, наскільки він зміг передбачити реакцію читача і відповідно вербалізувати свій інтелектуальний продукт [2, с. 217].

При моделюванні особлива увага звертається на такі питання: гносеологічні особливості моделі й онтологічна схожість моделі з оригіналом. Гносеологічні аспекти моделювання полягають у тому, що модель повинна бути аналогом, а не копією натурального об'єкта, що вивчається. Модель повинна мати загальний характер, що дасть змогу використовувати її для опису різноманітних натурних об'єктів. Наприклад, машинну модель українсько-російського словника доцільно побудувати так, щоб у майбутньому її можна було використовувати для моделювання співвідношення лексики інших пар мов: англійсько-української, польсько-української тощо.

Модель повинна відображати й відтворювати ті реальні та істотні риси натурального об'єкта, які є найважливішими для експерименту, її структура повинна бути добре відома досліднику. Модель не повинна містити суперечностей. Наприклад, при складанні частотного словника, у якому лексичні одиниці розташовані в порядку зменшення частоти їх використання, моделюється важливе для лінгвостатистики розподілення ймовірності вживання слів і словоформ у досліджуваній мові.

Отже, використання математичних методів і моделювання в мовознавстві має на меті замінити інтуїтивно сформульоване лінгвістичне завдання одним або кількома простішими, логічно сформульованими, які мають алгоритмічне розв'язання. Членування складної лінгвістичної проблеми на простіші алгоритмізовані задачі називається математичною експлікацією лінгвістичного об'єкта.

Література

1. Васильев В. И. Искусственный интеллект: Проблема обучения распознаванию образов / В. И. Васильев, А. И. Шевченко. – Донецк : ДонГИИИ, 1997. – 223 с.
2. Дядюра Г. М. Когнітивно-комунікативні аспекти наукового тексту : [монографія] / Дядюра Г. М., Кухарева-Рожко В. І., Ракшанова Г. Ф. – Черкаси : Чабаненко Ю. А., 2013. – 226 с.
3. Шмелев Д. Н. Слово и образ / Д. Н. Шмелев. – М. : Наука, 1964. – 120 с.
4. Щерба Л. В. Избранные работы по русскому языку / Л. В. Щерба. – М. : Учпедгиз, 1957. – 188 с.

Стаття надійшла до редакції 17.09.2016 р.