

ЧАСТКОВА СИСТЕМАТИЗАЦІЯ ТЕРМІНІВ МАТЕМАТИЧНОЇ ЛІНГВІСТИКИ

© Басалкевич О., 2006

Проведено систематизацію деяких термінів математичної лінгвістики з метою створення відповідних термінологічних розгалужень, що дозволяє спростити подання інформації на навчальних WEB-сторінках.

Systematization of some mathematical linguistic terms on purpose to create relevant term tree-forms that enables us to simplify the presentation of information on tutorial WEB-sites is conducted.

Загальна проблема та її науково-практична цінність. Науково-технічний прогрес ХХ століття започаткував тотальне формалізування переважної більшості наукових знань. Якщо ж машинне представлення моделей точних наук є процесом логічно доцільним та легкопроводжуваним, то результат такого ж відображення гуманітарних знань та їхніх систематизованих теорій досі досягається зі значною похибкою. Зокрема, спроба машинно формалізувати природні мови ще не сягнула досконалих форм: відсутні абсолютні перекладацькі інтерпретативні системи, та питанням номер один залишається досягнення якісно вищого рівня розпізнавання мовної інформації. Процес відповідних досліджень є безперервним, і його результати є проміжними у творенні моделі Штучного Інтелекту.

У даному контексті логічно потрібно сформувати гнучку термінологічну систему, що відображатиме нові поняття та теорії. Основною вимогою до такої системи є можливість її подальшого розвитку та вдосконалення. На основі поняттєвих зв'язків необхідно встановити термінологічну систему, що зокрема дасть змогу краще подавати відповідні знання та їх сприймати. Враховуючи актуальність інтерактивного навчання, доцільно адаптувати створену модель для представлення у середовищі Internet.

Аналіз досліджень і публікацій. На основі проаналізованих джерел філологічного [1; 3; 4; 9] та кібернетичного [2; 5; 6; 7] спрямування представлено класифікацію деяких термінів математичної лінгвістики, оскільки у вказаній літературі цей підхід нереалізовано чи реалізовано частково. При огляді відповідного Internet-джерела [8] виникла ідея дещо візуалізувати термінологічне представлення.

Установлення поняттєвих зв'язків та відповідна класифікація термінів математичної лінгвістики. Без сумніву, вихідною точкою зародження та розвитку сучасної математичної лінгвістики є теорія загальної лінгвістики (франц. linguistique, від лат. lingua – мова [1]) – науки про мову. Саме цей термін, а не його український варіант «мовознавство» адаптовано у назві розглядуваної дисципліни. Наступним кроком стало створення швейцарським (так званим Індоевропейським [9]) вченим Фердинандом де Сосюром (1857–1913) структурної лінгвістики – галузі мовознавства, що вивчає формальні моделі природних мов та методологію побудови таких моделей. Вчений запропонував чітке формулювання нових уявлень про мову, нових (точних) методів дослідження мови та вперше чітко визначив прості та загальні закони, що лежать в основі багатьох мовних явищ. За об'єктом досліджень у структурній лінгвістиці виділяють: 1) моделі, імітуючі мовну діяльність людини (об'єктом досліджень є конкретні мовні процеси та явища); 2) моделі, імітуючі дослідницьку діяльність лінгвіста (об'єкт дослідження – процедури, що

дозволяють вченому виявити те чи інше мовне явище, мета – об'єктивно обґрунтувати використовувані в лінгвістиці поняття); 3) метатеорію (від гр. $\mu\epsilon\tau\alpha$ – префікс, що означає проміжний стан, рух у просторі або часі, зміну і переміщення [1]) (об'єкт досліджень – готові лінгвістичні описи, мета – забезпечення системи критеріїв та теоретичних доказів, за допомогою яких здійснюється вибір найоптимальнішої моделі зі сукупності існуючих для імітації одного й того ж явища моделей. У свою чергу моделі, імітуючі мовну діяльність людини, поділяють на моделі аналізу, синтезу та породжуючі моделі [2]. Таким чином, сформувався підґрунтя для розвитку математичної лінгвістики.

На основі вищевказаних поняттєвих зв'язків утворена схема термінів на рис. 1. (А).

Як дисципліна математична лінгвістика (МЛ) з'явилася у 2-й половині 50-х років ХХ ст. Відповідно до основних завдань МЛ відносять до мовознавства або до математики.

МЛ – математична дисципліна, якщо вона розробляє формальний апарат для опису побудови природних та деяких штучних мов.

МЛ – галузь мовознавства, якщо вона застосовує будь-який математичний апарат для лінгвістичних досліджень.

МЛ поділяється на «кількісну» квантитативну лінгвістику на основі застосування розділів «кількісної» математики (математична статистика, теорія ймовірності, теорія інформації, математичний аналіз) та «якісну» комбінаторну на основі застосування «якісного» математичного апарату (теорія множин, математична логіка, теорія алгоритмів).

МЛ оперує: 1) теорією графів (дерев – для опису синтаксичної структури речень, дерев з додатковим відношенням лінійного порядку – для встановлення порядку слів у реченні, семантичною мережею – на ближчих до змісту рівнях); 2) аналітичними моделями мови (АММ) – моделями, в яких вважається заданим деякий набір текстів чи інших даних, що інтерпретуються як емпіричні дані про мову, на основі яких встановлюються ті чи інші закономірності побудови мови. АММ можуть мати за основу статистичні методи, але частіше послуговуються поняттями логіки, теорії множин, алгебри, рідше топології. Застосуванням АММ є лінгвістичне дешифрування текстів; 3) математичними моделями мови (МММ) – формальним апаратом, що використовується для чіткого й однозначного опису структури та породження природних і деяких штучних мов. МММ використовують основні лінгвістичні поняття та певні математичні засоби і операції (теоретико-множинні, алгебраїчні, логіко-математичні, теоретико-ймовірнісні та статистичні). МММ застосовуються для опису понять та відношень теоретичної лінгвістики (формулювання понять «відмінок», «рід», «частини мови»); 4) теорією формальних граматики.

Термінологічна схема на рис.1. (В) відображає вищеописаний поняттєвий зв'язок.

Формальна граMATика (ФГ) (від гр. $\gamma\rho\alpha\mu\mu\alpha\tau\iota\chi\eta$ - письмена, словесність) – система правил, що дає змогу строго та явно описувати деякі граматичні закономірності природної мови, а також задавати синтаксис формальної мови.

Засновники ФГ – американські вчені Н. Хомський та З.Херріс (40–50 рр. ХХ ст.).

ФГ послуговується термінами «контекстно-вільна» та «контекстно-залежна» граматики, «породжувальна» та «розпізнавальна» граматики. Крім того, виділяють стохастичну граматику та двовимірну граматику (як приклади часткових випадків) [2].

Контекстно-вільна граMATика (КВГ) – різновид ФГ – це засіб формально подавати синтаксис та семантику природних і штучних мов. Сукупність лексичних об'єктів у КВГ зображається двома класами символів: термінальними (елементарні одиничні об'єкти) та нетермінальними. КВГ описується правилами підстановок, так званими продукціями.

Контекстно-залежна граMATика (КЗГ) – різновид ФГ, на відміну від КВГ представлена правилами виду $\xi A \eta \rightarrow \xi F \eta$, де термінальний символ А замінюється послідовністю термінальних та нетермінальних символів лише при умові, що він знаходиться у контексті $\xi \dots \eta$ двох послідовностей ξ і η нетермінальних та термінальних символів. КЗГ продукує реальні мовні ланцюги.

Розпізнавальна граMATика (РГ) – система правил, що дозволяє визначити чи утворює даний ланцюг (послідовність символів) речення визначеної мови (фіксованої множини ланцюгів).

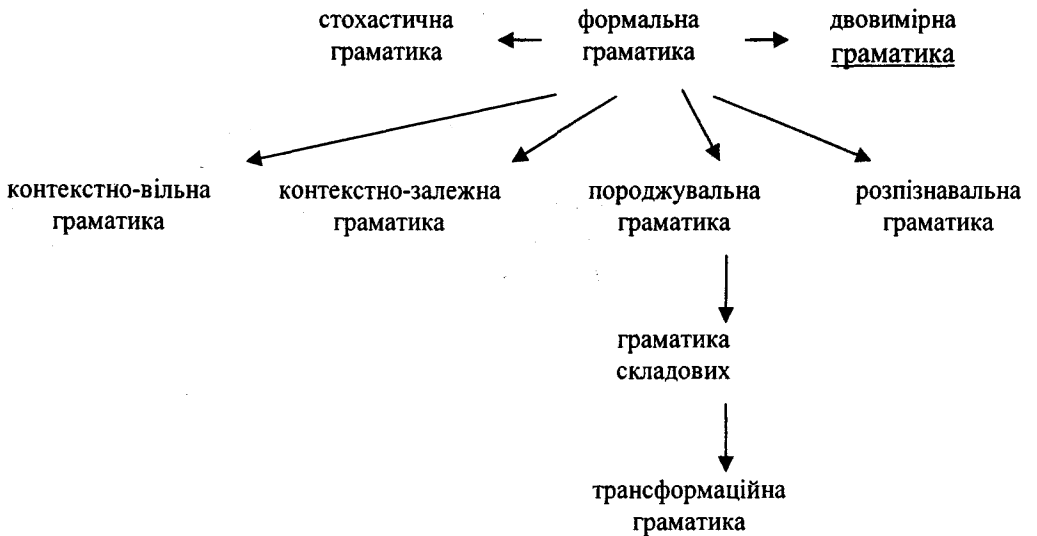
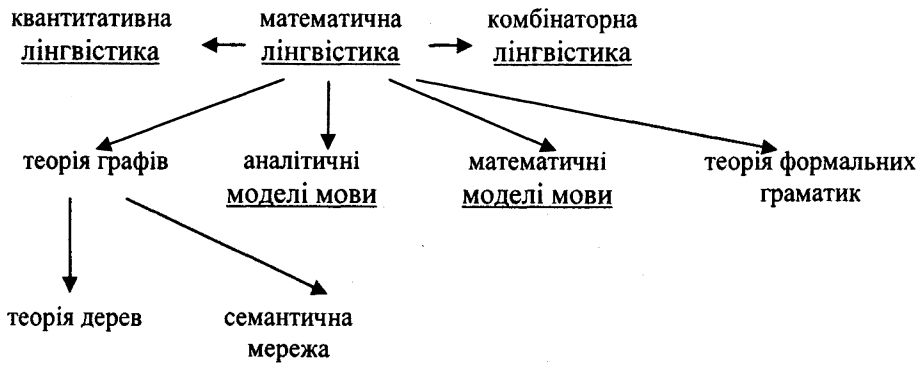
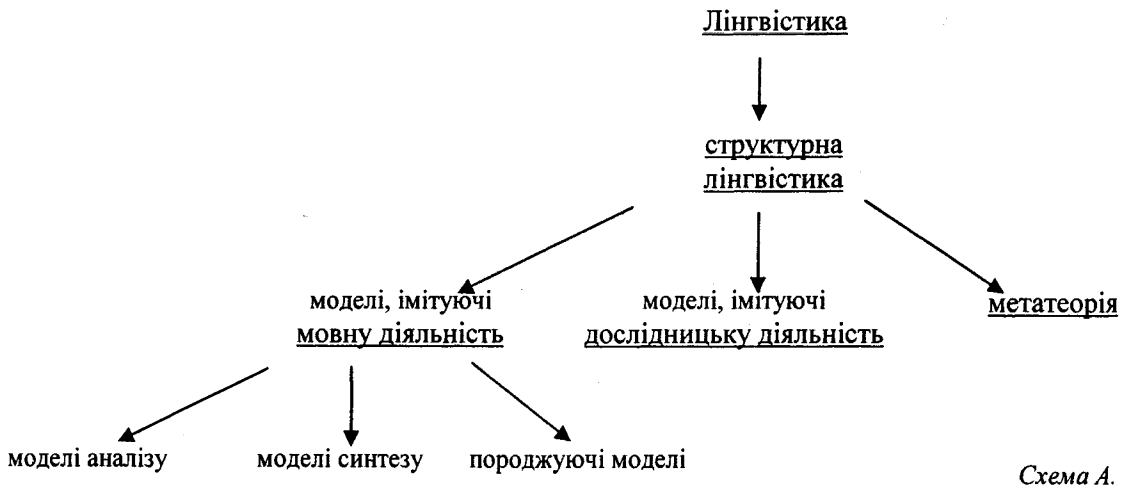


Рис. 1. Схема термінологічного розгалуження

Породжувальна граматика (ПГ) – система правил, що дозволяє будувати кінцеву послідовність символів (речень) та приписувати кожній з них деяку структурну характеристику.

Засновники ПГ – лінгвісти Н. Хомський та В. Інґве (50–60 рр. ХХ ст.). ПГ має своїм логічним продовженням граматику складових та трансформаційну граматику.

Граматики безпосередніх складових (ГБС) – різновид ПГ, складається з кінцевого набору символів та правил перекодування ланцюгів символів. На кожному етапі відбувається лише одна заміна символу. Така система є негнучкою при розмежуванні розповідних та запитальних речень та при усуненні синтаксичної омонімії.

Трансформаційна граматики (ТГ) – система правил, що дозволяє породжувати множини речень природної мови з відповідно невеликої кількості найпростіших, чи ядерних, речень шляхом застосування елементарних правил трансформації, за рахунок яких ТГ є більш гнучкою, оскільки дозволяє здійснювати одночасне перекодування декількох символів та їх перестановку.

Стохастична граматики (СГ) (від гр. $\sigma\chi\alpha\sigma\tau\iota\zeta$ – здогадка [1]) – система правил, що дозволяє породжувати множини ланцюгів з одночасним заданням розподілу ймовірностей на цій множині. Застосовується СГ при необхідності врахування статистичного характеру задач розпізнавання.

Двовимірні граматики (ДГ) – узагальнення ФГ на двовимірний випадок, коли породжуються не одновимірні ланцюги, а двовимірні масиви символів. Дані масиви легко аналізуються програмними засобами.

Відповідно до вищевикладеного визначення понять сформовано підсистему на рис. 1. (С).

Об'єднавши створені схеми, отримуємо деяку термінологічну систему математичної лінгвістики, яка легко представляється у середовищі навчальної Web-сторінки. Тут, вузли розгалужень – терміни – є посиланнями на відповідні їм визначення. У найпростішому випадку це впроваджується за допомогою коду HTML:

`термін ` (де файл `description.htm` містить визначення терміну).

Висновки. На основі проаналізованих літературних джерел описано деякий поняттєвий зв'язок та подано визначення основних термінів математичної лінгвістики, що дозволило систематизувати дані терміни та створити зразок схеми термінологічних розгалужень, зручний для представлення у середовищі навчальної Web-сторінки.

Запропонована модель є частковою, але сприятливою для подальшого вдосконалення та поповнення з потенційною можливістю досягнення універсального стану.

1. *Словник іноземних слів. н/р. О. С. Мельничука. – К.: Головна редакція УРЕ, 1974. – 775 с.*
2. *Словарь по кибернетике. н/р. В. С. Михалевича. – К.: Главная редакция УСЭ, 1989. – 752 с.*
3. *Українська радянська енциклопедія. – К.: Головна редакція УРЕ, 1978–1985. – тт. 1–12.*
4. *Кочерган М. П. Вступ до мовознавства. – К.: Видавничий центр «Академія», 2002. – 368 с.*
5. *Арапов Н. В., Герц Н. Н. Математические методы в исторической лингвистике. – М.: Наука, 1974. – 167 с.*
6. *Блехман М. С. Комп'ютерна лінгвістика. Навчальний посібник. – Х., 1993. – С. 25–40.*
7. *Якименко К. М. Системотехніка та технологія лексикографічних систем семантичного типу. Автореф. дис. ... канд. тех. наук. / К., 2006. – 18 с.*
8. *Комп'ютерна лінгвістика. Електронний ресурс. – Львів: ЛНУ ім. І. Франка, 2005. – www.franko.lviv.ua/ujcl/2_2st.htm.*
9. *New Webster's Dictionary and Thesaurus – Danbury, Lexicon publications, 1993. – P. 889*