

В. Висоцька, Т. Шестакевич, Ю. Щербина*

Національний університет “Львівська політехніка”,
кафедра інформаційних систем та мереж,

*Львівський національний університет імені Івана Франка,
кафедра дискретного аналізу та інтелектуальних систем

УТВОРЕННЯ РЕЧЕНЬ АНГЛІЙСЬКОЮ ТА НІМЕЦЬКОЮ МОВАМИ ЗА ДОПОМОГОЮ ПОРОДЖУВАЛЬНИХ ГРАМАТИК

© Висоцька В., Шестакевич Т., Щербина Ю., 2012

Описано застосування породжувальних граматики у лінгвістичному моделюванні. Подані в роботі моделі синтаксису речення застосовують для автоматизації процесів аналізу та синтезу природномовних текстів.

Ключові слова: породжувальні граматики, структурна схема речення, комп'ютерна лінгвістична система.

This paper describes the generative grammar application in linguistic modelling. Presented in work of syntax sentence models is applied to automate the processes of analysis and synthesis of texts in natural language.

Key words: Generative grammar, structured scheme sentences, computer linguistic system.

Вступ. Загальна постановка проблеми

Швидкий темп розвитку Інтернету сприяв створенню різноманітних лінгвістичних ресурсів. Потреба в автоматизації процесів аналізу та синтезу природномовних текстів зумовила появу відповідних лінгвістичних моделей процесів їх опрацювання. Сьогодні важливо забезпечити автоматичне опрацювання природномовної інформації в інформаційних засобах як системах машинного перекладу, інформаційно-пошукових, семантичного, статистичного, оптичного та акустичного аналізу і синтезу мови, автоматизованого редагування, екстракції знань з природномовних текстів, реферування, анотування, індексування текстів, навчально-дидактичних, менеджменту лінгвістичних корпусів, інструментальних засобів укладання словників різних типів тощо. Фахівці інтенсивно шукають нових моделей та способів їх опису, методів автоматичного опрацювання природномовної інформації. Одним із таких способів є розроблення загальних принципів побудови лексикографічних систем синтаксичного типу та побудови за цими принципами зазначених систем для конкретних мов. У зв'язку з цим в умовах постійного зростання інформаційних потоків тема статті є актуальною. Необхідним є розвиток багатьох мовознавчих дисциплін для потреб інформаційних наук. Інтеграційні процеси в більшості галузей життя сучасного світу привертають особливу увагу до розроблення та створення автоматизованих систем опрацювання багатомовної інформації.

Зв'язок висвітленої проблеми із важливими науковими та практичними завданнями

Розглянувши відомі способи та підходи до вирішення проблеми автоматичного опрацювання природномовної інформації [2–13, 15, 18–21, 23–24] та виділивши недоліки й переваги існуючих підходів та результатів [6, 24], отримані у галузі синтаксичних аспектів комп'ютерної лінгвістики, сформулюємо загальні концептуальні принципи моделювання словозмінних процесів при утворенні текстових масивів на прикладі англійських та німецьких речень. Далі з використанням запропонованих синтаксичних моделей та словозмінних класифікацій лексичного складу англійських та німецьких речень розробимо лексикографічні правила синтаксичного типу для автоматизованого опрацювання цих речень. Застосовуючи запропоновану методику, можна досягти вищих показників

надійності порівняно з відомими аналогами та високої ефективності у прикладних застосуваннях при побудові нових інформаційних технологій лексикографування та дослідження словозмінних ефектів природних мов. Робота має практичну цінність, оскільки запропоновані моделі та правила дають змогу ефективно організувати процес створення лексикографічних систем синтаксичного типу.

Аналіз останніх досліджень та публікацій

Традиційно аналіз природномовних текстів складається з трьох послідовних процесів: морфологічного, синтаксичного та семантичного аналізу. Для кожного з цих етапів було створено відповідні моделі та алгоритми [6, 24]. Так, теорія породжувальних граматики, початок якої заклав американський лінгвіст Н. Хомські [7–10, 18–21], є ефективним інструментом лінгвістичного моделювання на синтаксичному рівні мови. Він використав прийом формального аналізу граматичної структури фраз, який дає змогу виділити синтаксичні структури (складові), що є основною схемою фрази незалежно від її значення. Ідеї Н. Хомські розвинув радянський лінгвіст А.В. Гладкий, який застосовував поняття дерев залежності та систем складових для моделювання синтаксичного рівня мови [2, 8–10, 13, 23]. Він запропонував спосіб моделювання синтаксису за допомогою синтаксичних груп, що виділяють складові словосполучень як одиниці побудови дерева залежностей, – таке подання дало змогу об'єднати переваги методу безпосередніх складових і дерев залежностей [4, 5, 23]. Напрацювання Н. Хомські та А.В. Гладкого застосовано до розроблення таких засобів опрацювання природної мови, як інформаційно-пошукові системи, системи машинного перекладу, анутовання текстів, морфологічний, синтаксичний та семантичний аналіз текстів, навчально-дидактичні системи, лінгвістичне забезпечення спеціалізованих програмних систем тощо [12]. На сучасному етапі потреба в розробленні загальних та спеціалізованих лінгвістичних систем змушує прикладну та комп'ютерну лінгвістику виходити далеко за власні межі – в галузь інформаційних технологій [8–11, 15, 18–21]. Розроблення дієвих моделей мовлення для забезпечення комп'ютерних лінгвістичних систем дасть змогу виконати такі завдання прикладної лінгвістики, як аналіз та синтез усних і писемних текстів, описування та індексування документів, перекладання текстів, створення лексикографічних баз даних тощо. Перевагами моделювання за допомогою породжувальних граматики є можливість успішно описувати синтаксичний рівень мови (правила утворення речень зі словоформ) [8–10].

Формальна породжувальна граMATика G – це четвірка $G = (V, T, S, P)$, де V – скінченна непорожня множина, *алфавіт (словник)*; T – її підмножина, елементи якої є *термінальними (основними) символами, терміналами*; S – *початковий символ* ($S \in V$); P – скінченна множина *продукцій (правил перетворення)* вигляду $\xi \rightarrow \eta$, де ξ та η – ланцюжки над V . Множину $V \setminus T$ позначають N , її елементи є *нетермінальними (допоміжними) символами, не терміналами* [8–10]. Граматики класифікують за типами продукцій, на які накладено певні обмеження (табл. 1) [18–21].

Таблиця 1

Класифікація граматики за типами продукцій

Тип	Опис
Довільна	Тут ξ – довільний ланцюжок, що містить хоча б один нетермінальний символ, η – довільний ланцюжок над V .
Контекстно-залежна	В множині продукцій P є продукція вигляду $\gamma\xi\delta \rightarrow \gamma\eta\delta$ $ \xi \leq \eta $ (але не у формі $\xi \rightarrow \eta$), то ξ можна замінити на η лише в оточенні ланцюжків $\gamma\dots\delta$, тобто у відповідному контексті.
Контекстно-вільна	Нетермінал A у лівій частині продукції $A \rightarrow \eta$ може бути замінений ланцюжком η у довільному оточенні щоразу, коли він зустрічається, тобто незалежно від контексту.
Регулярна	Можуть бути лише продукції $A \rightarrow aB$, $A \rightarrow a$, $S \rightarrow \lambda$, де A, B – нетермінали, a – термінал, λ – порожній ланцюжок.

Формулювання мети

Лінгвістичне забезпечення використовують в усіх без винятку комп'ютерних системах. Удосконалення спілкування типу машина–людина є важливим актуальним завданням, вирішити яке можна удосконаленням процесу синтезу текстів на синтаксичному рівні. З цією метою розглянемо процес лінгвістичного моделювання синтезу речень англійською та німецькою мовами за допомогою породжувальних граматик. Для цього здійснимо належний опис словозмінних систем: на основі синтаксичного аналізу визначимо перелік відповідних слів, а також сформуємо систему правил, за допомогою яких можна одержати будь-які правильні форми речень, не отримавши при цьому жодного неправильного. Подання синтаксичної структури в термінах безпосередніх складових широко використовується в лінгвістиці, багато досліджувалося в різних аспектах і завоювало право на існування в теоретичному та експериментальному аспектах (автоматичний переклад або реферування тощо). Тому той факт, що граматики безпосередніх складових при породженні термінальних ланцюжків, наприклад, фраз природної мови, одночасно дають структуру їх безпосередніх складових (виведення фрази фактично і є структурою безпосередніх складових), робить їх особливо привабливими з лінгвістичного погляду. Довільна граMATика (без вимоги заміни відразу лише одного символу) не володіє властивістю заміняти фрази структури їх безпосередніми складовими. Оскільки в такій граMATиці кожного разу замінюється не один символ, а ціла група символів, у виведенні неможливо однозначно вказати для кожного символу його предка, і тому виведення не може бути перетворене в структуру безпосередніх складових. У межах статті покажемо способи застосування апарату породжувальних граматик до моделювання синтаксису речень англійською та німецькою. Для цього розберемо синтаксичну структуру речень, продемонструємо особливості процесу синтезу речень зазначених мов. Розглянемо вплив норм та правил мови на хід побудови граматик [6, 24].

Аналіз отриманих наукових результатів

Тлумачитимемо термінальні символи як словоформи деякої природної мови, нетермінальні символи – синтаксичні категорії, а термінальні ланцюжки, що виводяться, – правильні речення цієї мови [6–11, 18–21, 24]. Тоді виведення речення природно інтерпретується як його синтаксична структура, подана в термінах безпосередніх складових [6, 24]. Пояснимо сказане прикладами.

Приклад 1. Розглянемо процес виведення ланцюжків слів в англійській мові, де речення мають строго визначений, особливий порядок слів. Наприклад, для простого розповідного речення англійської мови основними синтаксичними категоріями є іменна та дієслівна групи [1, 25].

Іменна група виражається *займенником* або має таку структурну схему (тут і надалі у квадратних дужках зазначено необов'язкові елементи, у фігурних дужках – елементи, які можуть повторюватись):

$$[\text{Визначник}] \{ \{ \text{Прикметникова група} \} \} \text{Іменник} [\text{Прийменникова група}]. \quad (1)$$

Визначник – це граMATичний клас (не частина мови), до якого належать слова, що уточнюють іменну групу з погляду визначеності, кількості тощо. *Визначником* у англійській мові слугують артиклі, присвійні та вказівні займенники, кількісні прикметники та іменники у присвійному відмінку.

Прикметникова група має таку структурну схему:

$$\{ \{ \text{Прислівник} \} \} \text{Прикметник} [\text{Прийменникова група}]. \quad (2)$$

Прийменникова група має таку структурну схему:

$$\text{Прийменник} \text{Іменна група}. \quad (3)$$

Основними граMATичними характеристиками складових іменної групи в англійській мові є рід, число, відмінок іменника, відмінок займенника, ступінь порівняння, особа [1, 25].

Дієслівна група має таку структурну схему:

$$\text{Дієслівний вираз} [\text{Прислівникова група}]. \quad (4)$$

Дієслівний вираз складається з лексичних, допоміжних та модальних дієслів (тут і надалі Д. – скорочення для дієслова):

$$[\text{Модальне Д.}] [\text{Д. допоміжне}] [\text{Д. допоміжне}] [\text{Д. допоміжне}] \text{Д. лексичне} . \quad (5)$$

Прислівникова група має таку структурну схему:

$$\{\text{Прислівник}\} \{[\text{Прийменникова група}]\} . \quad (6)$$

Основними граматичними характеристиками складових дієслівної групи в англійській мові є особа, число, час, вид, спосіб, стан. Граматичними категоріями іменної та дієслівної груп є особа та число, бо саме за ними в реченні відбувається узгодження.

Розглянемо породжувальну граматику для моделювання синтаксису речення англійською мовою описаної структурної схеми. Алфавітом є іменникова група, дієслівна група та їх складові (нетермінальні символи), а також лексичний запас англійської мови (відповідні термінальні символи). За вимогами та правилами англійської мови в реченні на першому місці розташовується іменна група, після неї – дієслівна (це правило є основним). Усі можливі перетворення термінальних символів у нетермінальні складають множину правил. Під час виведення отримують незліченну кількість термінальних ланцюжків (речень англійською відповідної структурної схеми), тому така граMATика буде необмеженою і через свою складність застосування не матиме. Для введення контекстно-залежної граматики змодельємо процес побудови простого розповідного речення англійською мовою. Введемо такі обмеження на структуру речення. Розглядатимемо речення з іменною групою, вираженою не займенником, визначником-артиклем, спрощеною прикметниковою групою та без прийменникової групи; дієслівна група виражена лише лексичним дієсловом (вид – простий, спосіб – вказівний, стан – активний), у прислівниковій групі опущено прийменникову групу. Тобто розглядатимемо речення спрощеної структурної схеми. Спрощена іменна група матиме таку структурну схему:

$$\text{Визначник} \{ \text{Прикметник} \} \text{Іменник} , \quad (7)$$

а спрощена дієслівна група:

$$\text{Дієслівний вираз} \{ \text{Прислівник} \} . \quad (8)$$

Тепер основними граматичними характеристиками іменної групи є число, відмінок іменника, ступінь порівняння, особа. Прикметники об'єднуються сполучником. Основними граматичними характеристиками дієслівної групи будуть особа, число, час. Прислівники можуть об'єднуватися сполучником, допускається моделювання прислівника прислівником. Уведемо позначення складових спрощених іменної та дієслівної груп та їх граматичних категорій (у дужках – еквівалент англійською).

Складові іменної групи (після «/» тут і надалі вказано використовувані позначення)/ \tilde{N} : визначник/ D , прикметник/ A , іменник/ N . Граматичні категорії іменної групи та її складових: число (Number)/ NR : одина (Singular)/ sg , множина (Plural)/ pl ; відмінок (Case)/ CS : загальний (Common Case)/ cc , присвійний (Possessive Case)/ pc ; ступінь порівняння (Comparison Degree)/ CD : основна форма (Positive Degree)/ pd , вищий ступінь (Comparative Degree)/ cd , найвищий ступінь (Superlative Degree)/ sd ; особа (Person)/ PR : 1-ша (First)/ 1 , 2-га (Second)/ 2 , 3-тя (Third)/ 3 . Відповідно, скороченими позначеннями іменної групи буде $\tilde{N}_{NR,PR}$, а їх складових – $N_{NR,CS}$, A_{CD} .

Зауваження. Категорія особи – це граматична категорія особового займенника, який опущено у спрощеній формі речення. Проте цю категорію необхідно визначати для іменної групи з метою узгодження вживання дієслівної групи.

Складові дієслівної групи/ \tilde{R} : дієслово/ R , прислівник/ B . Граматичні категорії складових дієслівної групи: число (Number)/ NR : одина (Singular)/ sg , множина (Plural)/ pl ; особа (Person)/ PR : 1-ша (First)/ 1 , 2-га (Second)/ 2 , 3-тя (Third)/ 3 ; час (Time)/ TM : теперішній (Present)/ pr , минулий (Past)/ ps , майбутній (Future)/ ft . Скороченими позначеннями дієслівної групи буде $\tilde{R}_{NR,PR}$, а їх складових – $R_{NR,PR,TM}$, B .

Спосіб задання контекстно-залежної граматики, що виводить речення розробленої структурної схеми англійською мовою, наведемо на прикладі речення *The Ukrainian and Polish fans cheered emotionally and very actively.* (Українські та польські вболівальники вболівали емоційно та дуже активно.) Розглянемо граматику $G_1=(V, T, S, P)$. Алфавіт (позначення синтаксичних категорій

подамо без індексів – для зручності) $V=(S, \tilde{N}, \tilde{R}, D, A, N, R, Q, B, \text{Ukrainian}, \text{Polish}, \text{fan}, \text{cheer}, \text{the}, \text{and}, \text{emotionally}, \text{actively}, \text{very})$, $T=(\text{Ukrainian}, \text{Polish}, \text{fan}, \text{cheer}, \text{the}, \text{and}, \text{emotionally}, \text{actively}, \text{very})$, S – початковий символ. Множину правил P подамо у вигляді табл. 2.

Таблиця 2

Правила формування речення англійською мовою для Прикладу 1

№	Група	Правила
I.	Структура	$S \rightarrow \# \tilde{N}_{NR,PR} \tilde{R}_{NR,PR} \#$.
II.	Іменна група	1) $\tilde{N}_{NR,PR} \rightarrow DA_{CD} N_{NR,CS}$; 2) $A_{CD} \rightarrow A_{CD} QA_{CD}$.
III.	Дієслівна група	1) $\tilde{R}_{NR,PR} \rightarrow R_{NR,PR,TM} B$; 2) $B \rightarrow BQB$; 3) $B \rightarrow BB$; 4) $BB \rightarrow B^{\text{ступінь}} B$, тобто сполучуваність прислівника залежить від контексту. Змістовний сенс такого обмеження полягає у тому, що згідно з виділеними моделями сполучуваності прислівників, прислівники міри та ступеня, в разі сполучення їх з іншим прислівником, виступають здебільшого як препозитивний елемент (<i>so well, almost far</i>) [16].
IV.	Синтаксичні категорії слівформ	1) $A_{CD} \rightarrow \text{Ukrainian}_{CD}$ (український); 2) $A_{CD} \rightarrow \text{Polish}_{CD}$ (польський); 3) $N_{NR,CS} \rightarrow \text{fan}_{NR,CS}$ (вболівальник); 4) $R_{NR,PR,TM} \rightarrow \text{cheer}_{NR,PR,TM}$ (вболівати); 5) $D \rightarrow \text{the}$ (означений артикль); 6) $Q \rightarrow \text{and}$ (і); 7) $B \rightarrow \text{emotionally}$ (емоційно); 8) $B \rightarrow \text{actively}$ (активно); 9) $B^{\text{ступінь}} \rightarrow \text{very}$ (дуже).

Графічно виведення в граматиці G_1 для генерування англійського речення – на рис. 1 (тут і надалі $\#$ – символ межі речення, який вважається термінальним).

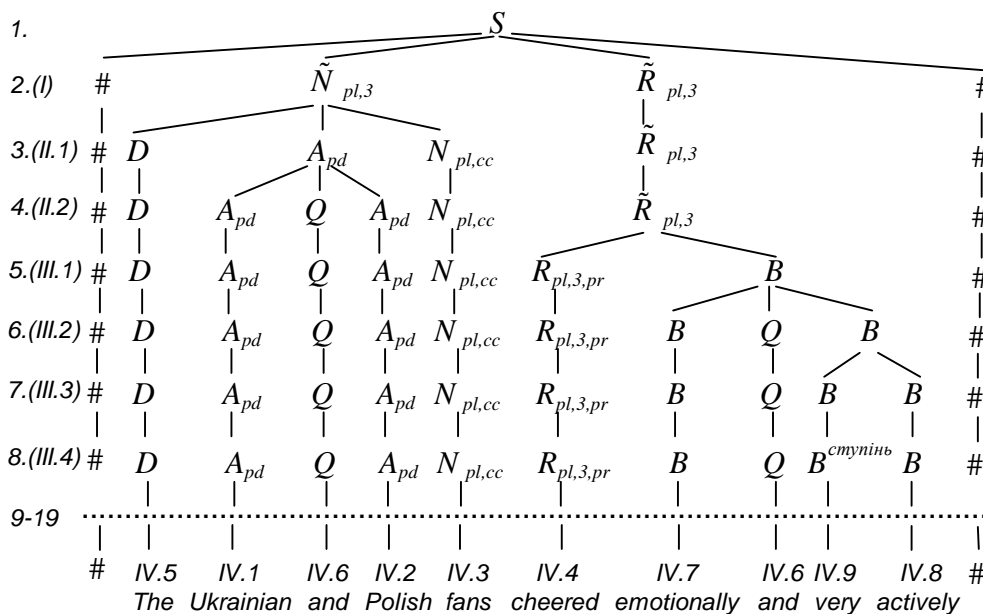


Рис. 1. Виведення в контекстно-залежній граматиці G_1 для генерування речення англійською

Виведення речення заданої структурної схеми подано нижче.

1. S
2. (I) $\# \tilde{N}_{pl,3} \tilde{R}_{pl,3} \#$
3. (II.1) $\# DA_{pd} N_{pl,cc} \tilde{R}_{pl,3} \#$

4. (II.2) # $D A_{pd} Q A_{pd} N_{pl,cc} \tilde{R}_{pl,3}$ #
5. (III.1) # $D A_{pd} Q A_{pd} N_{pl,cc} R_{pl,3,ps} B$ #
6. (III.2) # $D A_{pd} Q A_{pd} N_{pl,cc} R_{pl,3,ps} BQB$ #
7. (III.3) # $D A_{pd} Q A_{pd} N_{pl,cc} R_{pl,3,ps} BQBB$ #
8. (III.4) # $D A_{pd} Q A_{pd} N_{pl,cc} R_{pl,3,ps} BQB^{cmyuninb} B$ #
9. (IV.5) # *The* $A_{pd} Q A_{pd} N_{pl,cc} R_{pl,3,ps} BQB^{cmyuninb} B$ #
10. (IV.1) # *The Ukrainian* $Q A_{pd} N_{pl,cc} R_{pl,3,ps} BQB^{cmyuninb} B$ #
11. (IV.2) # *The Ukrainian Q Polish* $N_{pl,cc} R_{pl,3,ps} BQB^{cmyuninb} B$ #
12. (IV.6) # *The Ukrainian and Polish* $N_{pl,cc} R_{pl,3,ps} BQB^{cmyuninb} B$ #
13. (IV.6) # *The Ukrainian and Polish* $N_{pl,cc} R_{pl,3,ps} B$ and $B^{cmyuninb} B$ #
14. (IV.3) # *The Ukrainian and Polish fans* $R_{pl,3,ps} B$ and $B^{cmyuninb} B$ #
15. (IV.4) # *The Ukrainian and Polish fans cheered* B and $B^{cmyuninb} B$ #
16. (IV.7) # *The Ukrainian and Polish fans cheered emotionally* and $B^{cmyuninb} B$ #
17. (IV.8) # *The Ukrainian and Polish fans cheered emotionally* and $B^{cmyuninb}$ actively #
18. (IV.9) # *The Ukrainian and Polish fans cheered emotionally* and very actively #

Кожен крок виведення полягає або в розгортанні одного з символів попереднього ланцюжка (так, при переході від ланцюжка 2 до ланцюжка 3 символ $\tilde{R}_{pl,3}$ розгортається в три символи – $D, A_{pd}, N_{pl,cc}$), або в заміні його іншим (наприклад, при переході від ланцюжка 7 до ланцюжка 8 символ B замінюється на $B^{cmyuninb}$), інші ж символи переписуються без зміни. Граматика G_1 здатна породжувати й інші фрази (які не обов'язково мають бути змістовними), наприклад: # *The Ukrainian and Ukrainian and Ukrainian and Ukrainian fans cheered actively and very actively* # тощо. Більше того, граматика G_1 породжує нескінченно багато різних фраз, бо до її складу входять циклічні правила (наприклад, II.2). Особливість такого правила полягає в тому, що результат його застосування містить входження його лівої частини, так що воно завжди може бути застосоване до свого власного результату, що і приводить до нескінченної кількості фраз, і множина всіх цих фраз нескінченна.

Приклад 2. Розглянемо процес виведення речення німецькою мовою. Для простого розповідного речення німецькою мовою основними синтаксичними категоріями є іменна та дієслівна групи [14]. Як і у попередньому прикладі, алфавітом є іменникова група, дієслівна група та їхні складові (нетермінальні символи), а також лексичний запас німецької мови (відповідні термінальні символи). Усі можливі перетворення термінальних символів у нетермінальні є множиною правил. Під час виведення отримаємо незліченну кількість термінальних ланцюжків. Така граматика буде необмеженою, і через свою складність застосування не матиме. Щоб отримати контекстно-залежну граматику, введемо певні обмеження. Система порядку слів у німецькій мові не є абсолютно зв'язною, фіксоване розташування слів стосується лише позиції дієслівної групи. Розташування іменної групи та інших синтаксичних форм є відносно нефіксованим [22], а тому допускає і більше структурних видозмін схеми речення. З огляду на це, введемо структурні схеми іменної та дієслівної груп одразу з певними обмеженнями (спрощеннями). Наприклад, **іменна група** має таку структуру:

$$[Визначник] \{ \{ \text{Прикметникова група} \} \} \text{Іменник} [\text{Прийменникова група}] \text{ або } \text{Займенник} . \quad (9)$$

Визначник – це граматичний клас (строго кажучи, не частина мови), до якого належать слова, що уточнюють іменну групу з погляду визначеності, кількості тощо. *Визначником* у німецькій мові можуть бути артиклі, присвійні та вказівні займенники, кількісні прикметники та іменники у присвійному відмінку. *Прикметникова група* має таку структуру:

$$\{ \{ \text{Прислівник} \} \} \text{Прикметник} [\text{Прийменникова група}] . \quad (10)$$

Прийменникова група має таку структуру:

$$\text{Прийменник Іменна група.} \quad (11)$$

Іменна група володіє сукупністю граматичних ознак її елементів, тобто основними граматичними характеристиками іменної групи в німецькій мові є рід, число, відмінок іменника, відмінок займенника, ступінь порівняння, особа [14]. Узгоджуються дієслівна та іменникова групи за особою та числом. Стійкою ознакою синтаксичної будови німецького речення є його рамкова структура [17].

Дієслівна група має таку структуру:

$$\text{Дієслівний вираз [Прислівникова група].} \quad (12)$$

Розглядатимемо наступну спрощену структурну схему іменної групи:

$$\{\{\text{Прикметник}\}\}\text{Іменник}\{\{\text{Іменна група}\}\}\text{або Займенник,} \quad (13)$$

а спрощена структурна схема дієслівної групи:

$$\text{Дієслово}\{\{\text{Іменна група}\}\}. \quad (14)$$

Дієслівна група в німецькій мові є більш місткою порівняно з дієслівною групою української мови внаслідок формального введення до неї модальних членів, часток, заперечень та охоплення групою дієслова всіх частин речення, за винятком першого місця [14]. Нормативність вербальної рамки є стійкою ознакою синтаксичної структури німецького речення. Існує три види вербальних рамок: аналітичні форми дієслова (допоміжне дієслово + основне дієслово, основне дієслово + інфінітив іншого дієслова, дієслово + відокремлюваний префікс), основне дієслово + іменник, основне дієслово + прислівник [17]. Дієслівний вираз складається з лексичних, допоміжних та модальних дієслів:

$$\{\{\text{Модальне Д.}\}\}\{\{\text{Д. допоміжне}\}\}\{\{\text{Заперечення}\}\}\{\{\text{Основне Д.}\}\}\{\{\text{Інфінітив Д.}\}\}\{\{\text{Префікс}\}\}. \quad (15)$$

Прислівникова група має таку структуру:

$$\{\{\text{Прислівник}\}\}\{\{\text{Прийменникова група}\}\}. \quad (16)$$

Уведемо позначення спрощених граматичних категорій для іменної та дієслівної груп.

Складові іменної групи (після «/» тут і надалі вказано використовувани позначення)/ \tilde{N} : займенник (das Pronomen)/ N^{pron} , прикметник (Adjektiv)/A, іменник (Substantive)/N. Граматичні категорії іменної групи та її складових: число (Numerus)/NR: однина (Einzahl, Singular)/sg, множина (Plural)/pl; рід (Genus)/GN: чоловічий (Maskulinum)/mg, середній (Neutrum)/ng, жіночий (Femininum)/fg; відмінок (Kasus)/KS: називний (der Nominativ)/nk, родовий (der Genitiv)/gk, давальний (der Dativ)/dk, знахідний (der Akkusativ)/ak; особа (Person)/PR: 1-ша (erste Person)/1, 2-га (zweite Person)/2, 3-тя (dritte Person)/3. Відповідно, скороченим позначенням іменної групи буде $\tilde{N}_{GN,NR,KS,PR}$, а її складових – $A_{GN,NR,KS}$, $\tilde{N}_{GN,NR,KS,PR}$, $N_{GN,NR,KS,PR}^{pron}$, тобто

$$P = \{ \tilde{N}_{GN,NR,KS,PR} \rightarrow A_{GN,NR,KS} \tilde{N}_{GN,NR,KS,PR}, \tilde{N}_{GN,NR,KS,PR} \rightarrow N_{GN,NR,KS,PR}, \tilde{N}_{GN,NR,KS,PR} \rightarrow N_{GN,NR,KS,PR}^{pron} \}.$$

Складові дієслівної групи/ \tilde{R} : дієслово (Verben)/R, іменник (Substantive)/N. Граматичні категорії дієслівної групи: особа (Person)/PR: 1-ша (erste Person)/1, 2-га (zweite Person)/2, 3-тя (dritte Person)/3, число (Numerus)/NR: однина (Einzahl, Singular)/sg, множина (Plural)/pl; час (Verbformen)/VF: простий дійсний (Präsens)/pr, простий минулий (Präteritum)/pm, складний минулий (Perfekt)/pt, складний минулий (Plusquamperfekt)/pl, майбутній (Futur I)/ft, складний майбутній (Futur II)/fr. Скороченим позначенням дієслівної групи буде $\tilde{R}_{NR,VF,PR}$, а її складової – $R_{NR,VF,P}$, тобто

$$P = \{ \tilde{R}_{NR,VF,PR} \rightarrow \tilde{R}_{NR,VF,PR} \tilde{N}_{GN',NR',KS,PR'}, \tilde{R}_{NR,VF,PR} \rightarrow R_{NR,VF,PR}, \tilde{N}_{GN,NR,KS,PR} \rightarrow A_{GN,NR,KS} \tilde{N}_{GN,NR,KS,PR}, \tilde{N}_{GN,NR,KS,PR} \rightarrow N_{GN,NR,KS,PR}, \tilde{N}_{GN,NR,KS,PR} \rightarrow N_{GN,NR,KS,PR}^{pron} \}.$$

Спосіб задання граматики, що виводить речення розробленої структурної схеми німецькою мовою, наведемо на прикладі речення *In seinem bedeutendsten Werk, in der Gedichtsammlung "Kobsar", zeigt er, Taras Schewtschenko, die bunte Welt des ukrainischen Dorfes in ihrem einmaligen Reiz* [14]. (У своїй найважливішій роботі, в антології "Кобзар", Тарас Шевченко показує барвистий світ українського села в його неповторній привабливості). Для спрощення скоротимо фразу до вигляду: *In seinem bedeutendsten Werk zeigt er die bunte Welt des ukrainischen Dorfes in ihrem einmaligen Reiz*. (У своїй найважливішій роботі він показує барвистий світ українського села в його

неповторній привабливості). Розглянемо граматику $G_2=(V, T, S, P)$. Алфавіт (позначення синтаксичних категорій подамо без індексів – для зручності) $V=(S, \tilde{N}, \tilde{R}, A, N, R, sein, bedeutend, Werk, zeigt, er, bunt, welt, ukrainisch, Dorf, ihr, einmalig, Reiz), T=(sein, bedeutend, Werk, zeigt, er, bunt, welt, ukrainisch, Dorf, ihr, einmalig, Reiz), S$ – початковий символ. Множину правил P подамо у вигляді табл. 3.

Таблиця 3

Правила формування речення німецькою мовою для Прикладу 2

№	Група	Правила
I.	Структура	1) $S \rightarrow \# \tilde{N}_{GN,NR,nk,PR} \tilde{R}_{NR,pr,PR} \#$; 2) $S \rightarrow \# \tilde{N}_{GN,NR,dk,PR} \tilde{R}_{NR,pr,PR} \#$.
II.	Іменна група	1) $\tilde{N}_{GN,NR,KS,3} \rightarrow \tilde{N}_{GN,NR,KS,3} \tilde{N}_{GN',NR',gk,PR}$; 2) $\tilde{N}_{GN,NR,KS,3} \rightarrow A_{GN,NR,KS} \tilde{N}_{GN,NR,KS,3}$; 3) $\tilde{N}_{GN,NR,KS,3} \rightarrow N_{GN,NR,KS}$; 4) $K_1 \tilde{N}_{GN,NR,KS,PR} K_2 \rightarrow K_1 N_{GN,NR,KS,PR}^{pron} K_2$, де символ K_1 відмінний від $A_{GN,NR,KS}$, а K_2 – від символу $N_{GN,NR,KS}$, наприклад, неможливість виведення ланцюжка <i>lächelt Sie (усміхнений ти)</i> або <i>ich Freude (я радості)</i> .
III.	Дієслівна група	1) $\tilde{R}_{NR,pr,PR} \rightarrow \tilde{R}_{NR,pr,PR} \tilde{N}_{GN',NR',ak,PR'}$; 2) $\tilde{R}_{NR,pr,PR} \rightarrow \tilde{R}_{NR,pr,PR} \tilde{N}_{GN',NR',nk,PR'}$; 3) $\tilde{R}_{NR,pr,PR} \rightarrow \tilde{R}_{NR,pr,PR} \tilde{N}_{GN',NR',dk,PR'}$; 4) $\tilde{R}_{NR,pr,PR} \rightarrow R_{NR,pr,PR}$.
IV.	Синтаксичні категорії словоформ	1) $N_{mg,NR,KS} \rightarrow Reiz_{NR,KS}$ (привабливість); 7) $A_{GN,NR,KS} \rightarrow ihr_{GN,NR,KS}$ (їхній); 2) $N_{fg,NR,KS} \rightarrow Welt_{NR,KS}$ (світ); 8) $A_{GN,NR,KS} \rightarrow bedeutend_{GN,NR,KS}$ (важливий); 3) $N_{ng,NR,KS} \rightarrow Werk_{NR,KS}$ (робота); 9) $A_{GN,NR,KS} \rightarrow bunt_{GN,NR,KS}$ (барвистий); 4) $N_{ng,NR,KS} \rightarrow Dorf_{NR,KS}$ (село); 10) $A_{GN,NR,KS} \rightarrow ukrainisch_{GN,NR,KS}$ (український); 5) $N_{mg,sg,KS,1}^{pron} \rightarrow er_{KS}$ (він); 11) $A_{GN,NR,KS} \rightarrow einmalig_{GN,NR,KS}$ (неповторний); 6) $A_{GN,NR,KS} \rightarrow sein_{GN,NR,KS}$ (свій); 12) $R_{NR,pr,PR} \rightarrow zeigen_{NR,pr,PR}$ (показати).

Поділимо речення на 6 складових для перевірки схеми виведення (табл. 4):

1	2	3	4	5	6
Dativ	Präsens	Nominativ	Akkusativ	Genitiv	Dativ
Substantive	Verben	Pronomen	Substantive	Substantive	Substantive
In seinem bedeutendsten Werk	zeigt	er	die bunte Welt	des ukrainischen Dorfes	in ihrem einmaligen Reiz

Таблиця 4

Виведення ланцюжків речення німецькою мовою для Прикладу 2

№	Ланцюжок	Правило	Номер
1	In seinem bedeutendsten Werk zeigt	$S \rightarrow \# \tilde{N}_{GN,NR,dk,PR} \tilde{R}_{NR,pr,PR} \#$	(I.2)
2	zeigt er	$\tilde{R}_{NR,pr,PR} \rightarrow \tilde{R}_{NR,pr,PR} \tilde{N}_{GN',NR',nk,PR'}$	(III.2)
3	zeigt die bunte Welt	$\tilde{R}_{NR,pr,PR} \rightarrow \tilde{R}_{NR,pr,PR} \tilde{N}_{GN',NR',ak,PR'}$	(III.1)
4	die bunte Welt des ukrainischen Dorfes	$\tilde{N}_{GN,NR,PR,3} \rightarrow \tilde{N}_{GN,NR,PR,3} \tilde{N}_{GN',NR',gk,PR}$	(II.1)
5	zeigt in ihrem einmaligen Reiz	$\tilde{R}_{NR,pr,PR} \rightarrow \tilde{R}_{NR,pr,PR} \tilde{N}_{GN',NR',dk,PR'}$	(III.3)

Графічно виведення в граматиці G_2 для генерування німецького речення подано на рис. 2.

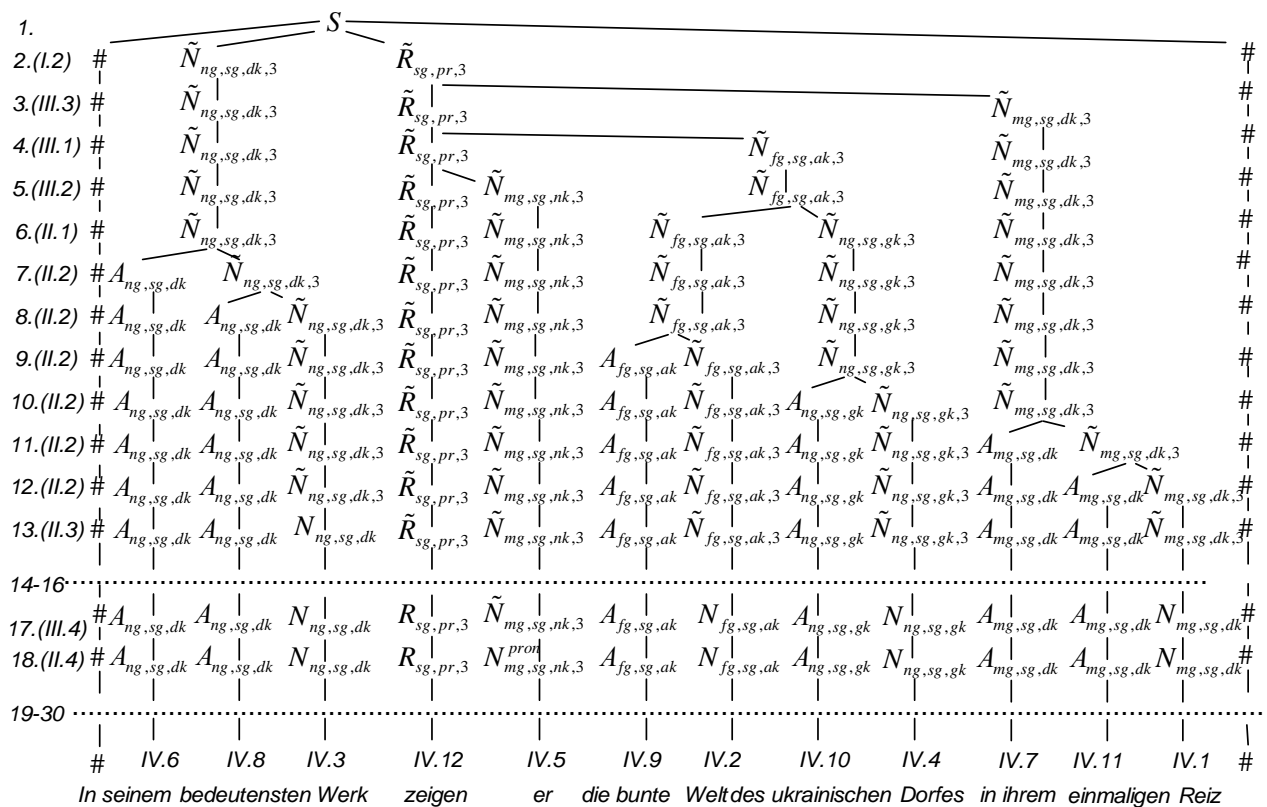


Рис. 2. Виведення в контекстно залежній граматиці G_2 для генерування речення німецькою

Кожен крок виведення полягає в розгортанні одного з символів попереднього ланцюжка (при переході від 2 до 3 ланцюжка $\tilde{R}_{sg,pr,3}$ розгортається в $\tilde{R}_{sg,pr,3}$ та $\tilde{N}_{mg,sg,dk,3}$) або в заміні його іншим (при переході від 16 до 17 ланцюжка $\tilde{R}_{sg,pr,3}$ замінюється на $R_{sg,pr,3}$), а інші символи переписуються без зміни. Граматика G_2 здатна породжувати й інші фрази, не обов'язково змістовні, наприклад: # *In seinem seinem seinem bedeutendsten bedeutendsten Werk zeigt er die bunte bunte Welt des ukrainischen Dorfes in ihrem ihrem ihrem ihrem einmaligen Reiz* #.

Висновки і перспективи подальших наукових розвідок

Дослідження застосування математичних методів для аналізу та синтезу текстів природною мовою беруть свій початок ще з середини 50-х років минулого століття, коли на розроблення математичних алгоритмів та комп'ютерних програм опрацювання природномовних текстів було спрямовано значні зусилля науковців. Апарат породжувальних граматики, запропонований Н. Хомські, моделює процеси на синтаксичному рівні мови – виділені структурні елементи речення дають змогу описувати синтаксичні конструкції незалежно від їх змісту. У статті показано особливості процесу синтезу речень різних мов із застосуванням породжувальних граматики, розглянуто вплив норм та правил мови на процес побудови граматики. Застосування породжувальних граматики має широкі можливості в розробленні та створенні автоматизованих систем опрацювання багатомовної інформації та лінгвістичного забезпечення комп'ютерних лінгвістичних систем.

1. Английская грамматика в доступном изложении // Режим доступа: <http://real-english.ru/crash/lesson3.htm>. 2. Анісімов А.В. Алгоритмічна модель асоціативно-семантичного контекстного аналізу текстів природною мовою / А.В. Анісімов, О.О. Марченко, А.О. Никоненко // Пробл. програмув. – 2008. – № 2, 3. – С. 379–384. 3. Анісімов А.В. Комп'ютерна лінгвістика для всіх: міфи, алгоритми, язык / А.В. Анісімов. – К.: Думка, 1991. – 208 с. 4. Апресян Ю.Д. Идеи и методы современной структурной лингвистики / Ю.Д. Апресян. – М.: Просвещение, 1966. – 305 с.

5. Апресян Ю.Д. Непосредственно составляющих метод / Ю.Д. Апресян // Лингвистический энциклопедический словарь под ред. В.Н. Ярцевой. – М.: Советская энциклопедия, 1990. – Режим доступа: <http://tapemark.narod.ru/les/332a.html>. 6. Висоцька В.А. Генерування речень українською за допомогою породжувальних граматики / В.А. Висоцька, Т.В. Шестакевич // ISDMIT'2012, Євпаторія. – 27–31 травня 2012. – С. 48–50. 7. Гакман О.В. Генеративно-трансформаційна лінгвістика Н. Хомського як вираження його лінгвістичної філософії / О.В. Гакман // Мультиверсум. Філософський альманах. – К.: Центр духовної культури, 2005. – № 45. – С. 98–114. 8. Гладкий А.В. Синтаксические структуры естественного языка в автоматизированных системах общения / А.В. Гладкий. – М.: Наука, 1985. – 144 с. 9. Гладкий А.В. Элементы математической лингвистики / А.В. Гладкий, И.А. Мельчук. – М.: Наука, 1969. – 192 с. 10. Гладкий А.В. Формальные грамматики и языки / А.В. Гладкий. – М.: Наука, 1973. – 368 с. 11. Дарчук Н.П. Комп'ютерна лінгвістика (автоматичне опрацювання тексту): підручник / Н.П. Дарчук. – К.: ВПЦ “Київський університет”, 2008. – 351 с. 12. Любченко Т.П. Лексикографічні системи граматичного типу та їх застосування в засобах автоматизованого опрацювання мови: автореф. дис. канд. техн. наук: спец. 10.02.21 / Т.П. Любченко. – Київ, 2011. – 19 с. 13. Марченко О.О. Алгоритми семантичного аналізу природномовних текстів: автореф. дис. ... канд. фіз.-мат. наук: спец. 01.05.01 // О.О. Марченко. – К., 2005. – 15 с. 14. Носков С.А. Самоучитель немецкого языка. Deutsch für sie / С.А. Носков. – К.: Наука, 1999. – 400 с. 15. Партико З.В. Прикладна і комп'ютерна лінгвістика. Вступ до спеціальності: навч. посіб. / З.В. Партико. – Л.: Афіша, 2008. – 224 с. 16. Торосян О.М. Функціональні характеристики прислівників міри та ступеня в сучасній англійській мові: автореф. дис. ... канд. філол. наук / О.М. Торосян. – Режим доступу: <http://disser.com.ua/contents/6712.html>. 17. Туришева О.О. Порушення рамкової конструкції в сучасній німецькій мові: функціональний аспект, нормативний статус: автореф. дис. ... канд. філол. наук: спец. 10.02.04 / Туришева О.О. – Одеса, 2012. – 20 с. 18. Хомський Н. О некоторых формальных свойствах грамматик / Н. Хомский // Кибернетический сборник. – М.: Мир, 1962. – № 5. – С. 279–311. 19. Хомский Н. Формальный анализ естественных языков / Н. Хомский, Дж. Миллер // Кибернетический сборник. – М.: Мир, 1965. – № 1. – С. 231–290. 20. Хомский Н. Язык и мышление / Н. Хомский // Публикации ОСиПЛ. Серия монографий. – М.: Издательство Московского университета, 1972. – № 2. – 122 с. 21. Хомский Н. Синтаксические структуры / Н. Хомский // Сборник «Новое в лингвистике». – М.: ИЛ, 1962. – № 2. – С. 412–527. 22. Чепурна З.В. Трансформація порядку слів у простому реченні при перекладі з німецької мови українською / З.В. Чепурна // Наукові записки, серія «Філологічні науки (мовознавство)»: у 5 ч. – Кіровоград: РВВ КДПУ ім. В. Винниченка, 2010. – Вип. 89 (1). – С. 232–236. 23. Шаров С.А. Средства компьютерного представления лингвистической информации / С.А. Шаров. – Режим доступа : <http://www.ksu.ru/eng/science/ittc/vol00>. 24. Шестакевич Т.В. Застосування породжувальних граматики для генерування речень українською мовою / Т.В. Шестакевич, В.А. Висоцька // Східно-Європейський журнал передових технологій. – Харків, 2012. – № 3/2 (57). – С. 51–53. 25. English Verbs (Part 1) – Basic Terms. – Режим доступу: <http://sites.google.com/site/englishgrammarguide/Home/english-verbs--part-1---basic-terms>.