

В. А. Висоцька¹, О. М. Наум²¹Національний університет “Львівська політехніка”,
кафедра інформаційних систем та мереж,²Дрогобицький державний педагогічний університет ім. І. Франка,
кафедра інформаційних систем і технологій

ПОРІВНЯННЯ СКЛАДНОСТІ АВТОМАТИЧНОГО ОПРАЦЮВАННЯ АНГЛІЙСЬКИХ ТА УКРАЇНСЬКИХ ТЕКСТІВ З УРАХУВАННЯМ СЕМАНТИКИ ТА СИНТАКСИСУ ПРИРОДНИХ МОВ

© Висоцька В. А., Наум О. М., 2017

Проаналізовано складність автоматичного опрацювання англійських та українських текстів з урахуванням семантики та синтаксису природних мов. Проведено порівняльний лінгвістичний аналіз синтаксису та семантики англійської та української мов для синтезу корпусів природомовних текстів або автоматичного опрацювання відповідних текстових масивів даних. Порівняння англійської та української мов полегшує розроблення лінгвістичних алгоритмів опрацювання природомовних текстів германського та слов'янського типів.

Ключові слова: породжувальні граматики, структурна схема речення, інформаційна лінгвістична система.

The complexity of automatic processing of English and Ukrainian texts with regard to syntax and semantics of natural languages is analyzed. The comparative linguistic analysis of syntax and semantics of English and Ukrainian languages synthesis buildings natural texts or automatic text processing relevant data sets is analyzed. Comparison of English and Ukrainian languages facilitates the development of linguistic processing algorithms natural texts Germanic and Slavic type.

Key words: generative grammar, structured scheme sentences, information linguistic system.

Вступ. Загальна постановка проблеми

Кожна мова унікальна як за структурою, так і за особливостями побудови лінгвістичних одиниць для передавання змістової інформації. Для автоматичного опрацювання змістовних текстових масивів даних цей факт значно ускладнює розроблення алгоритмів аналізу та синтезу природомовних текстів, наприклад, для автоматичного перекладу, синтезу відповідей в робототехніці, автоматичного формування діалогу в режимі машина-людина, автоматичного опрацювання інформаційних ресурсів різних мов від фільтрації/пошуку релевантних даних до формування пропозицій щодо сюжетних подій для ЗМІ. Основна складність розроблення таких алгоритмів полягає в тому, що розробник цих алгоритмів змалечку привчений мислити в межах своєї природної мови і не розуміє глибини мови іншої мовної групи, ніж його рідна.

Так, наприклад, в чому складність вивчення української мови для іноземців, для яких рідна мова з іншої мовної групи, не слов'янської? Або особливість автоматичного перекладу чи опису розпізнаних зображень з урахуванням не лише синтаксису (тут задача майже ідеально розв'язана), але і семантики текстів? Наприклад, перед нами стіл. На столі є склянка і виделка. Що вони роблять? Стакан стоїть, а виделка лежить. Якщо застромити виделку в стільницю, вона стоятиме. Тобто стоять вертикальні предмети, а лежать горизонтальні? Додаємо на стіл тарілку і пательню. Вони ніби горизонтальні, але на столі стоять. Тепер покладемо тарілку в пательню. Там вона лежить, хоча на столі стояла. Може, стоять предмети, готові до використання? Ні, виделка готова

була, коли лежала. Тепер на стіл залізла кішка. Вона може стояти, сидіти і лежати. Якщо в плані стояння і лежання вона вписується в логіку “вертикальний-горизонтальний”, то сидіння – це нова властивість. Сидить вона не на лапках. Тепер на стіл сіла пташка. Вона на столі сидить, але сидить на лапках. Хоча начебто повинна стояти. Але стояти вона не може зовсім. Але якщо з цієї пташки зробити опудало, воно на столі стоятиме. Складається враження, що сидіння – атрибут живого, але чобіт на нозі теж сидить, хоча він не живий і не має тієї частини тіла, на якій кішка сидить. Так і не зрозуміло, за яким принципом та за якою логікою щось/хтось сидить, лежить або стоїть. Тому, напевно, іноземці вважають нашу мову складною і порівнюють жартома з “китайською”. Для людини, яка думає з народження рідною мовою, зрозумілі такі нюанси опису, і зовсім заплутані для іноземця, особливо якщо його рідна мова з іншої мовної групи [1].

Тепер порівняємо структури англійського та українського речення. В англійській мові, на відміну від української, прямий (твердий) порядок слів у реченні. Тобто за підметом іде присудок, а потім додаток. Щодо інших членів речення, то вони, як правило, можуть стояти перед підметом або після додатка [1]. Українською мовою можна сказати:

- | | |
|--------------------------------|--------------------------------|
| 1. Дітям подобається морозиво. | 2. Морозиво дітям подобається. |
| 3. Дітям морозиво подобається. | 4. Подобається дітям морозиво. |
| 5. Морозиво подобається дітям. | 6. Подобається морозиво дітям. |

Англійською можна лише сказати: *підмет – присудок – додаток: Children like ice-cream.* Англійська є аналітичною мовою, а українська мова синтетична [1–2]. В аналітичних мовах керування в реченні здійснюється переважно за рахунок слів, допоміжних слів тощо, тоді як у синтетичних мовах це відбувається завдяки відмінюваності слів (наявності, наприклад, суфіксів).

Аналіз наукових публікацій

Основою основ для формування будь-якого лінгвістичного алгоритму є визначення лінгвістичної одиниці [3–10]. Для цього провадиться аналіз двох напрямів – призначення майбутнього алгоритму та групи мов, для яких цей алгоритм розробляється. Згідно з призначенням майбутнього алгоритму вибирають напрями дослідження та відповідні аналізи: графемний, морфологічний, синтаксичний, семантичний, лексичний тощо [11–21]. Від останнього залежить вибір лексичної одиниці як частини природної мови [22–32]. Будь-яка мова складається зі звуків (в усному мовленні) та літер (на письмі). Своєю чергою, із звуків і літер формуються склади, а зі складів – слова. Зі слів будуються речення, а речення утворюють текст.

Напрями досліджень та аналізу природної мови	Лінгвістичний аналіз				
	Структурний			Лексичний	
	Графемний	Морфемний	Морфологічний	Синтаксичний	Семантичний
Усне мовлення (фонетика)	звук	склад	слово	речення	текст
Письмо (орфографія)	літера				

Рис. 1. Загальна структура природної мови

Для кожного з цих складників існують певні правила та закономірності їх функціонування [22–32]. Так, правила вимови звуків, слів, інтонацію речень вивчає фонетика; правила написання слів, речень – орфографія; словниковий склад мови – лексикографія; граматика займається формою слів (морфологія) та правилами об’єднання слів у речення (синтаксис) [1–32].

Всі мови (досі не відома точна кількість мов, якими розмовляє населення Землі, але приблизно їх декілька тисяч) об’єднано в групи [1]. Англійська мова споріднена з датською, німецькою, норвезькою, шведською та іншими мовами. Ці та ще деякі мови утворюють групу германських мов, які разом з іншими мовами Європи та Азії, наприклад, слов’янськими (білоруською, болгарською, польською, українською, російською, чеською тощо), романськими (іспанською, італійською, румунською, французькою тощо) та іншими утворюють індоевропейську сім’ю мов (табл. 1) [1]. В англійській мові менше літер, зате більше приголосних і голосних звуків, є дифтонги, які відсутні в українській мові (табл. 2) [1–2].

Таблиця 1

Системні особливості англійської мови порівняно з українською [1–2, 21]

Частина мови	Англійська мова	Українська мова
Артикль	Дві форми – означений, неозначений (особливі слова для уточнення іменників)	Відсутній
Іменники	Не мають граматичного роду	Мають граматичний рід
	Поділяються на іменники, які означають людей, і на іменники, що означають предмети і явища	Поділяються за ознакою граматичного роду на іменники чоловічого, жіночого й середнього роду
	Існують два відмінки – загальний і присвійний	Є сім відмінків
	Відношення і зв'язки виражаються за допомогою прийменників	Відношення і зв'язки передаються відмінками
Прикметники	Не узгоджуються і не змінюються за родами, числами, відмінками	Узгоджуються з іменником, до якого вони відносяться, за родами, числами, відмінками.
Числівники	Не узгоджуються за відмінками, родами	Узгоджуються за відмінками, родами
Займенники	Поділяються на сім розрядів	Поділяються на дев'ять розрядів
	Відсутній особовий займенник <i>ти</i> (його функцію виконує займенник <i>you</i>)	Є дві форми особового займенника другої особи: в однині <i>ти</i> , у множині <i>ви</i>
	Особові займенники: <i>he</i> – він замінює іменники, які позначають живі істоти чоловічої статі; <i>she</i> – вона замінює іменники, які означають живі істоти жіночої статі; <i>it</i> – він, вона, воно замінює іменники, які означають неживі предмети і тварин чоловічого, жіночого, середнього роду однини	Особові займенники: <i>він</i> замінює всі іменники чоловічого роду; <i>вона</i> замінює всі іменники жіночого роду; <i>воно</i> замінює всі іменники середнього роду
Дієслова	Вживаються залежно від того, чи відбувається дія взагалі, завжди, постійно, повторно, чи вона відбувається в момент мовлення, чи протягом часу, який ще триває, чи відбувалася до якоїсь іншої дії в минулому тощо.	Вживаються для вираження завершеності чи незавершеності дій, які не завжди залежать від тих чинників, про які йдеться стосовно англійського дієслова
	Широко вживаються з різними прислівниками, які втрачають своє лексичне значення	Такого явища не існує
Інфінітив	Крім простої форми, яка подібна на форму в українській мові, існує ще п'ять складних форм	Існує лише проста форма
Дієприкметник	Є дві форми дієприкметника: теперішнього й минулого часу, має деякі дієприслівникові властивості	Існує лише одна форма
Дієприслівник	Відсутній у “чистому” вигляді	Існує дві форми
Герундій	Існує шість форм	Ця категорія відсутня
Безособові дієслова	Відсутні	Наявні, наприклад, <i>вечоріє</i>
Прислівник, прийменник, сполучник та вигук	Не мають істотних відмінностей	
Речення	Твердий (прямий) порядок слів у реченні: підмет – присудок – інші члени речення	Порядок слів у реченні вільний

Фонетичні та орфографічні особливості англійської мови порівняно з українською [1–2, 21]

Лінгвістична одиниця	Англійська мова	Українська мова
Літери:	26	32
– приголосні	20	22
– голосні	6	10
Звуки:	44	38
– приголосні	24, поділу немає; майже всі вимовляються твердо перед будь-яким голосним	32, можуть бути твердими і м'якими, перед деякими голосними відбувається пом'якшення приголосних (звук [с] у словах <i>сіно</i> і <i>сірий</i>)
– голосні	12, можуть бути довгими й короткими, заміна одного звуку іншим призводить до зміни значення слова	6, поділу немає
Дифтонги (звуки)	8, деякі англійські голосні складаються з двох елементів, що вимовляються в межах одного складу (дифтонги)	Немає
Дзвінкі приголосні	Завжди вимовляються дзвінко в кінці слова та перед глухими приголосними; їх оглушення часто приводить до зміни значення слова	В таких випадках оглушуються, наприклад, <i>віз</i> [віс], <i>Бог</i> [бох], <i>дуб</i> [дуп]

У табл. 3 подано деякі фонетичні й граматичні відмінності між сучасними англійською та українською мовами. Аналіз динаміки потоку англійськомовного/україномовного текстового контенту та побудова етапів опрацювання інформаційних ресурсів цього контенту є важливими та актуальними [3–30]. Якісне опрацювання цих текстових масивів даних із ефективним аналізом синтаксису (для встановлення ключових слів та рубрикації тексту, визначення авторства, ефективного перекладу тощо) та семантики (для виділення змісту та призначення тексту) необхідне для оперативного отримання сучасних даних у світі, який перевантажений цінною інформацією та надлишком інформаційного шуму. Ефективне розроблення і впровадження тематичних інформаційних ресурсів сьогодні неможливе без коректного визначення множини ключових слів для ефективного позиціонування та ранжування цих ресурсів у пошукових системах [3–30]. Крім того, вони мають вживатись у текстовому масиві контенту не аби як, а за певним розподілом, щоб пошукові системи не індексували і не класифікували це як спам. Розроблення методу опрацювання україномовного/англійськомовного текстового контенту для автоматичного виявлення значущих ключових слів та його рубрикації є одним зі стратегічних напрямів розвитку вітчизняного е-бізнесу та ІТ [3–30]. Забезпечення можливості автоматизації процесів опрацювання інформаційних ресурсів для виявлення значущих ключових слів та рубрикації контенту, аналізу синтаксису та семантики тексту сприяє збільшенню обсягів продажу контенту, товару або послуг постійному користувачу, активному залученню потенційних користувачів та розширенню меж цільової аудиторії [3–30]. Зокрема, ці принципи і технології в е-комерції активно застосовують під час створення систем on-line/off-line продажу та аналізу/обміну/збереження контенту, інтернет-магазину, cloud storage/computing [3–30]. Відсутність загального стандартизованого підходу до опрацювання української текстової інформації з метою автоматичного виявлення значущих ключових слів та рубрикації контенту, аналізу синтаксису та семантики тексту, а також проектування функціонала систем підтримки е-комерції призводить до виникнення проблем під час реалізації типової структури таких систем [3–30]. Метою роботи є визначення оптимального методу автоматичного опрацювання множини україномовного текстового контенту для виявлення значущих ключових слів та автоматичної рубрикації контенту, аналізу синтаксису та семантики тексту. Опрацювання множини контенту *C* для виявлення значущих ключових слів зазвичай побудовано на принципі знаходження ключових слів за змістом (термами), що оснований на законі Зіпфа і зводиться до

вибору слів із середньою частотою появи. Оскільки це прямий та найпростіший спосіб, пропонується застосовувати складніші та здійснити відповідне експериментальне дослідження. Експериментальною базою для такого дослідження вибрано 100 наукових публікацій Вісника двох номерів: 783 (<http://science.lp.edu.ua/SISN/SISN-2014>) та 805 (<http://science.lp.edu.ua/sisn/vol-cur-805-2014-2>). Тому розроблено інший алгоритм знаходження множини ключових слів з урахуванням основ тематичних слів, та розміщено у відкритому доступі (<http://victana.lviv.ua/index.php/kliuchovi-slova>). Проаналізовано та використано довідники з [31–32].

Аналіз наукових результатів

На рис. 2, а подано таблицю ключових слів, де атрибут `rub_id` описує рубрику, до якої належить конкретне ключове слово в цьому записі. Атрибут `flag` конкретизує властивості цього ключового слова (до якої частини мови належить).

id	ordering	state	word	rub_id	flag	lang
70597	0	1	підтвердження	4	ij	uk
71207	0	1	планшет	3	efg	uk
71209	0	1	план	4	ef	uk
71728	0	1	побічний	2	VW	uk
72303	0	1	повнота	2	aZ	uk
72555	0	1	погодження	3	ijZ	uk
74132	0	1	поліморфний	2	V	uk
74194	0	1	політехніка	0	ab	uk
74195	0	1	політехнікум	0	ef	uk
74196	0	1	політехнічний	0	V	uk

а

```

Файл  Правка  Вид  Справка
#####
# Групи а б с d o
#
# -- Перша відміна: іменники жіночого та чоловічого та середнього роду
#
# -- Друга відміна: іменники чоловічого роду із закінченням на -ар -ир
#                       наголошені (Мішана група на -ар -ир)
#
# -- Друга відміна: іменники чоловічого роду з чергуванням -і -о
#
# -- Числівники -ять, -сят, -сто
#
#
SFX а у 235
#
# ОДНИНА (множина перенесена в гр. б)
#
# Спочатку перша відміна
#
# тверда група в Називному відмінку однини з закінченням на -а
# однина
SFX а а и [^жщц]а # хата хати (Р.)
SFX а а і [^г'кх]а # хата хаті (Д.М.)
SFX а а у а # хата хату (З.)
SFX а а ою [^жщц]а # хата хатою (О.)

```

б

Рис. 2. Таблиця ключових слів (а) та словник іменників (б)

На рис. 3, а подано правила приведення до основи слова. За атрибутом `flag` визначається тип слова (в цьому прикладі – іменникова група, однина), за атрибутом `mask` – правило ідентифікації закінчення, за атрибутом `find` – закінчення слова в називному відмінку, за атрибутом `gerl` – закінчення слова під час відмінювання. В квадратних дужках – винятки. Наприклад, перший рядок описує конкретний приклад # Іменники із закінченням на -ін з чергуванням -і-о.

Фонетичні та граматичні особливості англійської мови порівняно з українською [1–2, 21]

Лінгвістична одиниця	Трактування означення	Мова	
		Англійська	Українська
1	2	3	4
Активний стан	дію виконує підмет. Див. <i>пассивный стан</i> .	Строгий порядок слів (A boy broke a cup).	Довільний порядок слів (хлопчик розбив чашку або чашку розбив хлопчик).
Альтернативне питання	питання, що пропонує вибір	Строгий порядок слів (Does he speak English or Spanish?).	Довільний порядок слів (Він говорить англійською чи іспанською? або Говорить він англійською чи іспанською?).
Вищий ступінь порівняння	форма прикметника або прислівника, яка вживається для порівняння двох осіб, речей, понять тощо	Три ступені порівняння: звичайний (the Positive Degree), вищий (the Comparative Degree) і найвищий (the Superlative degree): (<i>smaller than</i> – менший ніж, <i>more expensive</i> – дорожчий ніж).	Два ступені порівняння: вищий (смачніший) і найвищий (найсмачніший)
Відносне підрядне речення	речення, що вводиться відносним займенником	Строгий порядок слів (This is the book <i>which I bought yesterday</i>).	Довільний порядок слів (Це книга, яку я купив вчора або Це книга, яку вчора я купив)
Відносні займенники	використовуються для зв'язку головного і підрядного речень.	<i>who</i> (хто, той, що, який, котрий); <i>whom</i> (кого, кому); <i>that</i> (який); <i>which</i> (котрий, який, хто, що); <i>whose</i> (чий, чия, чий, чий).	що, хто, скільки, який, чий, котрий.
Дієприкметник	це форма дієслова, яка означає ознаку предмета за дією або станом і відповідає на питання який? яка? яке? які? (хмарою повиті, врятована планета, зачарований красою).	Неособова форма англійського дієслова, що має властивості дієслова, прислівника та прикметника. В українській мові англійський дієприкметник відповідає дієприслівнику та дієприкметнику.	Активні дієприкметники виражають ознаку предмета за його ж дією (палаюче небо). Пасивні дієприкметники виражають ознаку предмета за дією, яка зумовлена дією іншого предмета над ним (посіяне жито (хтось посіяв)).
Дієприкметник-І теперішнього часу (Present Participle I)	Він має дві форми: Present Participle Simple, що відповідає українському дієприкметнику теперішнього часу. Present Participle Perfect, що відповідає укр. дієприкметнику теперішнього часу та дієприслівнику недоконаного виду.	-ing форма дієслова (reading the book he makes notes – читаючи книгу, він робить позначки). Present Participle Simple в активному стані утворюється за допомогою додавання закінчення -ing до 1 форми дієслова – так само, як і герундій. Укр. він перекладається дієприкметником в активному стані	Активні дієприкметники теперішнього часу утворюються від основи теперішнього часу перехідних і неперехідних дієслів недоконаного виду за допомогою суфіксів -уч(ий), -юч(ий) для дієслів 1-ї дієвідміни і -ач(ий), -яч(ий) для дієслів 2-ї дієвідміни (<i>реве</i> ® <i>режучий, працює</i> ® <i>працюючий</i>).

1	2	3	4
Дієприкметник-2 минулого часу (Past Participle або Participle II)	Він відповідає дієприкметнику минулого часу в укр. мові. Пасивні дієприкметники в укр. мові утворюються від основи інфінітива перехідних дієслів доконаного і недоконаного виду за допомогою суфіксів -т(ий) , -н(ий) , -ен(ий) , -єн(ий) : <i>митий</i> ® <i>митий</i> , <i>засіяний</i> ® <i>засіяний</i> , <i>везти</i> ® <i>везений</i> , <i>засвоїти</i> ® <i>засвоєний</i>	третя форма дієслова break-broke-broken (а broken cup – розбита чашка). Дієприкметник минулого часу має лише пасивну форму і перекладається як дієприкметник минулого стану українською мовою	Активні дієприкметники минулого часу утворюються від основи інфінітива лише перехідних дієслів доконаного виду за допомогою суфікса -л(ий) : <i>замерзнути</i> ® <i>замерзлий</i> , <i>побіліти</i> ® <i>побілілий</i>
Дієслово	окреме слово чи фраза, що описує стан або дію (He loves his children. Children <i>play</i> in the yard). Дієслова в англійській мові, як і в українській, означають дію (<i>to go, to build</i>), стан (<i>to sleep, to rest</i>), почуття (<i>to hear, to like</i>), процеси мислення (<i>to think, to realize</i>). В англійській мові є кілька складних дієслів, які мають дві основи: <i>to whitewash, to browbeat, to machine-gun</i> . Багато англійських дієслів збігаються за формою з іменниками (рідше — з прикметниками)	Модальні, почуттєві, фразові, неправильні. Дієслова бувають прості , похідні , складні й складені . Прості дієслова складаються з однієї непохідної основи: <i>to run, to speak, to go, to try</i> та ін. Похідні дієслова мають суфікси або префікси: <i>to organize, rewrite, to discover, to mispronounce</i> . Складені дієслова складаються з двох частин — дієслівної основи і відокремленого суфікса, які пишуться окремо і можуть роз'єднуватися іншими словами: <i>to stand up, to sit down, to go away, to put on</i> та ін. Складені , <i>please! Put your car on!</i> Складені дієслова дуже поширені в англійській мові. Всі закінчення у таких дієсловах приєднуються до основи. <i>He always wakes up at 7 o'clock. I'm writing down your address</i>	В українській мові дієслово має п'ять типових форм. Ці форми можна розпізнати за характерними закінченнями: 1) неозначена форма (інфінітив); 2) особова форми: (він) <i>пиш-е, писа-в-О, напиш-е, буде + писа-ти, писати-ме, писа-в-О + би, хай + пиш-е</i> ; 3) дієприкметник: <i>пожовті-л-ий, посиві-л-ий; писа-н-ий, підписа-н-ий; залюбл-єн-ий, бач-єн-ий; вилл-т-ий, коло-т-ий</i> ; активні дієприкметники (<i>пишучий</i>) українській мові не притаманні, що функцію виконують описові конструкції – що (або який) пише; 4) безособова на -но/-то: <i>написа-но, зробле-но, прожж-то, випи-то</i> ; 5) дієприслівник: <i>пиш-учи, любл-ючи, підписа-вши, полуби-вши</i>
Дійсний спосіб	Дійсний спосіб (індикатив) означає реальну дію; є найуживанішим. Дієслова в дійсному способі змінюються за часами	показує, що дія розглядається як реальний факт у теперішньому, минулому або майбутньому часі	Значення минулого і теперішнього часу є реальним, а майбутнього – гіпотетичним, тому він може мати відтінок значення недійсного способу
Додаток	частина мови (іменник, займенник, дієслово тощо), яка йде за головним дієсловом речення (присудком) і відповідає на питання <i>що? кого?</i>	Додаток може бути прямим, непрямым, прийменниковим. Додаток може бути виражений іменником, займенником, інфінітивом, герундієм, цілим підрядним додатковим реченням. (I can see <i>a bus</i> . They ask me to help)	Додаток виражається тими ж частинами мови, що й підмет

1	2	3	4
yes/no питання	Загальне питання, яке потребує відповіді <i>yes (так)/no (ні)</i>	Строгий порядок слів	Довільний порядок слів
Займенник (Pronouns)	слово, що вживається замість іменника (A boy reads books – <i>He</i> reads books). Займенники за своїм лексичним значенням і морфологічними ознаками поділяються на кілька розрядів	8: особисті (Personal), присвійні (Possessive), зворотні (Reflexive), питальні (Interrogative), вказівні (Demonstrative), відносні (Relative), невизначені (Indefinite), взаємні (Reciprocal)	9: особові, зворотний, питальні, відносні, присвійні, вказівні, означальні, неозначені, заперечні
Заперечне речення	речення з часткою not (ne) (He <i>doesn't</i> speak English. We <i>don't</i> like classical music)	Строгий порядок слів	Довільний порядок слів
Зворотні займенники	Від присвійних займенників my, our, your тощо додаванням закінчень	The Reflexive Pronouns – <i>myself, yourself, himself, herself, itself, oneself, ourselves, yourselves, themselves</i>	<i>себе</i>
Іменник	ім'я (<i>Robert</i>), особа або річ (<i>a teacher</i> – вчитель, <i>a table</i> – стіл), дія (<i>a conversation</i> – розмова)	Не мають граматичного роду	Мають граматичний рід
Іменникове означення	іменник, який є означенням іншого <i>a stone bridge</i> (тобто кам'яний міст)	Існує	Немає
Інфінітив	початкова форма дієслова	як правило, з часткою to – <i>to write</i> (He likes to write letters)	без частки: <i>писа-ти, говори-ти, літа-ти, гримі-ти, мерзну-ти, диєвуа-ти</i>
Кількісні прислівники	незмінна самостійна частина мови, що виражає ознаку дії, стан предмета або ознаку якості	слово, що вказує на кількість чогось: <i>many, much, some, any</i>	відповідає на запитання як? де? звідки? наскільки? якою мірою?
Коротка відповідь	відповідь, що містить підмет+дієслово-присудок (Who came? – <i>Mike did</i>)	Строгий порядок слів	Довільний порядок слів
Множина	більше від одного (<i>girls, men, children, rooms</i>). Див. однина	Додавання закінчення s, крім винятків (там чергування літер у слові або закінчення en)	Додавання різних флексій залежно від роду іменника
Модальні дієслова	це функціонально-семантична категорія, яка виражає відношення змісту висловлювання до дійсності та мовця до змісту висловлювання. Модальність є істотною ознакою речення	<i>can, could, may, might, will, would, shall, should, must, ought to, need, needn't, used to</i>	Граматично модальність виражається поєднанням дієслова (чи іншого предиката) з модальними частками, прислівниками, дієсловами, а також словосполученнями та реченнями. Портвняйте також різну модальність у прикладі: <i>піді-но, піді, ти б пішов, нехай би ти пішов, бодай би ти пішов</i>

1	2	3	4
Найвищий ступінь	форма прикметника або прислівника, що виражає найвищу міру	<i>the largest, the most important</i>	Проста форма найвищого ступеня утворюється від форми вищого ступеня за допомогою префікса <i>най-</i> . Для підсилення можуть вживатися префікси як- і що-. Складена форма ступенів порівняння прислівників утворюється додаванням до звичайного прислівника: для вищого ступеня слів <i>більш, менш</i> ; для найвищого ступеня слів <i>найбільш, найменш</i>
Наказовий спосіб	дієслово у формі наказу, вказівки тощо	Строгий порядок слів (Open the book! Don't smoke!)	Довільний порядок слів
Необчислювані іменники	іменники, що не вживаються у множині: <i>air</i> – повітря, <i>snow</i> – сніг, <i>milk</i> – молоко	Існують як винятки. Мають або форму однини, або форму множини	Існують як винятки. Мають або форму однини, або форму множини
Неправильні форми	ті, які змінюються на за звичайними правилами. Див. правильні форми	(дієслова – <i>be was/where been, write wrote written, ступенів порівняння прикметників/прислівників – good better best</i>)	Немає
Однина	одна річ або особа (<i>a girl, a man, a child, a room</i>)	Існує та вживається з артиклем	Існує та вживається без артикля
Окличне речення	виражає здивування, гнів тощо (What a nice day! – Який чудовий день!)	Строгий порядок слів	Довільний порядок слів
Особа	граматична особа займенника	(1-ша особа – <i>I, me</i> ; 2-га особа – <i>you</i> ; 3-тя особа – <i>he, she, it, one, they</i>)	(1-ша особа – <i>я, мене</i> ; 2-га особа – <i>ти</i> ; 3-тя особа – <i>він, вона, воно, вони</i>)
Пасивний стан	дія спрямована на підмет	<i>be</i> +дієприкметник – <i>A cup was broken</i> (чашку розбито)	Немає
Питання	альтернативне, загальне <i>yes/no</i> питання до підмета, роз'єднувальне, спеціальне <i>wh</i> -питання	Строгий порядок слів	Довільний порядок слів
Питання до підмета	питання, що запитує про підмет (утворюється без допоміжного дієслова)	Строгий порядок слів (Who came late? – Jack did.)	Довільний порядок слів
Підмет	іменник, займенник або інша частина мови, що передує головному дієслову (присудку)	<i>A car has two doors. We are happy</i>	Машина має двоє дверей. Ми щасливі

1	2	3	4
Почуттєві дієслова	такі дієслова, як <i>feel</i> (почувати), <i>hear</i> (чути), <i>look</i> (дивитися), <i>smell</i> (відчувати запах, нюхати), <i>sound</i> (звучати)	Існують синтаксичні особливості використання	Немає синтаксичних особливостей використання
Правильні форми	ті, які змінюються за звичайними правилами. Див. неправильні форми	(дієслова – <i>play</i> – <i>played</i> – <i>played</i> , ступенів порівняння прикметників/прислівників – <i>small</i> – <i>smaller</i> – <i>the smallest</i>)	Всі дієслова правильної форми
Прийменники	вживаються перед іменником для позначення місця, часу, напрямку	такі слова, як <i>at</i> , <i>in</i> , <i>on</i> , <i>to</i> , <i>under</i> , <i>near</i> (<i>in the street</i> , <i>on Wednesday</i> , <i>at home</i>)	Існують, але не мають буквального перекладу з англійської
Прикметник	описує особу, річ, подію тощо Див. ступеневі порівняння прикметників	не змінюване за числом, родом та відмінком слово (a <i>tall</i> boy, a <i>happy</i> end, a <i>long</i> holiday)	змінюване за числом, родом та відмінком слово
Присвійний відмінок	утворений за допомогою флексій	апострофа й літери <i>s</i> (' <i>s</i>): 'с додається до іменників (загальних чи власних), що указати на належність John's father (батько Джона)	Або зміні слова (мама – мамин, тато – татів)
Присвійні займенники	Означає належність до когось або чогось	<i>mine</i> , <i>yours</i> , <i>his</i> , <i>hers</i> , <i>ours</i> , <i>theirs</i>	Мій, твій, його, її, наш, їхній
Прислівник	не змінюване за числом, відмінком слово, що вказує, як, коли, куди, де тощо відбувається дія	He speaks <i>slowly</i>	Він говорить швидко
Прислівники ступеня	такі слова, як <i>enough</i> (дості), <i>fairly</i> (досить, цілком), <i>hardly</i> (ледве, насили), <i>quite</i> (цілком, зовсім, абсолютно, повністю), <i>rather</i> (швидше, краще, переважно)	Існують (The film was <i>quite</i> good)	Існують
Простий час	(минулий, теперішній, майбутній) виражає негритивалу дію	Існує (The <i>dance</i> well. He <i>lived</i> in Kyiv)	Не існує
Речення	наказове, окличне, питальне, розповідне, заперечне, стверджувальне	Строгий порядок слів	Довільний порядок слів
Роз'єднувальне питання	коротка питальна частина, що йде за розповідною	Існує. (He speaks English, <i>doesn't he?</i> He <i>doesn't</i> speak English, <i>does he?</i>)	Немає
Розповідне речення	в якому щось стверджується або заперечується (He speaks English. He <i>doesn't</i> speak English)	Строгий порядок слів	Довільний порядок слів
Складний	(іменник, займенник тощо) з двох або більше частин	<i>schoolboy</i> (школяр), <i>somebody</i> (хтось)	Хто-небудь, вухогорлоніс
Спеціальне wh-питання	питання, яке починається з питальних слів <i>who(m)</i> , <i>what</i> , <i>when</i> , <i>which</i> , <i>why</i> , <i>where</i> , <i>whose</i>	Строгий порядок слів	Довільний порядок слів

1	2	3	4
Сполучник	такі слова, як <i>and</i> (і, а), <i>but</i> (але), <i>when</i> (коли), <i>because</i> (тому що, бо), що з'єднують речення	He likes hard rock <i>but</i> I like classical music	Йому подобається тяжкий рок, але мені подобається класична музика
Стверджувальне речення	не заперечне і не питальне (He speaks English. We like classical music)	Строгий порядок слів	Довільний порядок слів
Тривалий час	форма дієслова, що утворюється з допоміжного дієслова <i>be</i> і смислового дієслова із закінченням <i>ing be + V-ing</i>	вказує, що дія відбувається, відбулась або відбуватиметься у розвитку. (She is <i>writing</i> (вона пише, тобто зараз). She <i>was writing</i> (вона писала). She <i>will be writing</i> (вона писатиме)	Немає
Умовний спосіб	показує, що мовець розглядає свою дію як реальний факт, як щось допустиме, бажане	Існує	Не існує
Форма -ing	дієслово, прикметник або іменник, що закінчується на <i>-ing</i>	He is <i>reading</i> a boring book; I like <i>reading</i>	Немає
Фразове дієслово	дієслово, що вживається з прийменниками	<i>Look at</i> the picture. <i>Come in</i>	Немає
Час	форма дієслова, що вказує на час	12 форм дієслова залежно від часу (теперішній – <i>present</i> , минулий – <i>past</i> , майбутній – <i>future</i>)	3 форми дієслова залежно від часу
Часові позначення	прислівники, що вказують, коли відбувається дія	<i>last year, today, in 1994, on Sunday</i>	Аналогічно

У тематичних словниках біля кожного слова вказана його властивість, наприклад,

буферизувати/ABGH	кілобайтовий/V	клатурний/V	кодосумісний/V	консоль/ij	криптозахисний/V
відформатувати/AB	кілобайт/efg	Кобол/е	комбостисок/ab	конфігуратор/efg	крос-асемблер/efg
декодувати/ABGH	кілобітовий/V	кодек/efg	комутований/V	копілефт/е	крос-компілятор/efg
кешувати/ABGH	кілобіт/efg	кодер/efg	конкатенація/ab	копірайт/е	кука/ab
кирилічний/V	кілобод/efg	кодогенератор/efg	консольний/V	криптографічний/V	курсорний/V

де V – прикметник, A – дієслово, а групи a b c d o описують деякі іменники (рис. 2, б).

id	ordering	state	flag	type	lang	mask	find	repl	id	ordering	state	word	lang
26	26	1	a	SFX	uk	ін	ін	оном	1	1	1	після	uk
27	27	1	a	SFX	uk	ін	ін	оні	2	2	1	між	uk
28	28	1	a	SFX	uk	іг	іг	огу	3	3	1	are	en
29	29	1	a	SFX	uk	іг	іг	огові	4	4	1	and	en
30	30	1	a	SFX	uk	іг	іг	огом	5	5	7	між	uk
31	31	1	a	SFX	uk	іг	іг	озі	6	6	1	been	en
32	32	1	a	SFX	uk	[^л]ід	ід	оду	7	7	1	has	en
33	33	1	a	SFX	uk	[^л]ід	ід	одові	8	8	1	their	en
34	34	1	a	SFX	uk	[^л]ід	ід	одом	9	9	1	any	en
35	35	1	a	SFX	uk	[^л]ід	ід	оді	10	10	1	the	en
36	36	1	a	SFX	uk	[^пг]лід	ід	ьоду	11	11	1	with	en
37	37	1	a	SFX	uk	[^пг]лід	ід	ьодові	12	12	1	таких	uk
38	38	1	a	SFX	uk	[^пг]лід	ід	ьодом	13	13	1	їхніми	uk
39	39	1	a	SFX	uk	[^пг]лід	ід	ьоді	14	14	1	как	ru
40	40	1	a	SFX	uk	[пг]лід	ід	оду	15	15	1	такої	uk
41	41	1	a	SFX	uk	[пг]лід	ід	одові	16	16	1	на	uk
42	42	1	a	SFX	uk	[пг]лід	ід	одом	17	17	1	на	ru
43	43	1	a	SFX	uk	[пг]лід	ід	оді	18	18	1	ними	uk

Рис. 3. Таблиця правил приведення до основи слова (а) та заблокованих слів (б)

SFX a ін ону	ін # загін загону (Д.Р.)	SFX a ін оном	ін # загін загonom (О.)
SFX a ін онові	ін # загін загонові (Д.)	SFX a ін оні	ін # загін загоні (М.)
третій рядок описує # Іменники із закінченням на -іг з чергуванням -і -о			
SFX a ігогу	іг # батіг батогу (Д.Р.)	SFX a ігогом	іг # батіг батогом (О.)
SFX a ігогові	іг # батіг батові (Д.М.)	SFX a ігозі	іг # батіг батозі (М.)
дев'ятий рядок описує # Іменники із закінченням на -ід з чергуванням -і -о			
SFX a ід оду [^л]ід	# провід проводу (Д.Р.)	SFX a ід одом [^л]ід	# провід проводом (О.)
SFX a ід одові [^л]ід	# провід проводові (Д.)	SFX a ід оді [^л]ід	# провід проводі (М.)

На рис. 3, б наведено таблицю слів, які заблокував модератор, тобто слів, які не можуть бути ключовими словами, але їх врахування під час аналізу текстів істотно впливає на кінцевий результат.

Висновки і перспективи подальших наукових розвідок

Складність генерування термінальних ланцюжків у англійській мові полягає в тому, що дерево є довшим та ширшим з-за наявності артиклів та залежності послідовності іменникових груп, що поєднані відповідними прийменниками, однієї від одної. Генерування термінальних ланцюжків українською мовою ускладнюється лише наявністю додаткових відмінків та особливістю змін флексій залежно від роду та відмінку вживаного терму в контексті.

1. Бех П. Англійська мова / П. Бех, Л. Биркун. – К.: Либідь, 1993. – 232 с. 2. *Methods based on ontologies for information resources processing* / [V. Lytvyn, V. Vysotska, L. Chyrun, D. Dosyn] // LAP Lambert Academic Publishing, Germany, 2016. – 324 с. 3. Берко А. Системи електронної контент-комерції / А. Берко, В. Висоцька, В. Пасічник. – Л.: Вид-во Нац. ун-ту “Львівська політехніка”, 2009. – 612 с. 4. Математична лінгвістика / [В. Пасічник, Ю. Щербина, В. Висоцька, Т. Шестакевич]. – Л.: Новий світ-2000, 2012. – 359 с. 5. Висоцька В. Методи і засоби опрацювання інформаційних ресурсів в системах електронної контент-комерції: автореф. дис. ... канд. техн. наук: 05.13.06 – інформаційні технології / В. Висоцька. – Львів, 2014. – 27 с. 6. Berko A. Features of information resources processing in electronic content commerce / A. Berko, V. Vysotska, L. Chyrun // ACS journal. – 10(2). – Poland, 2014. – P. 5–19. 7. Vysotska V. Web Content Processing Method for Electronic Business Systems / V. Vysotska, L. Chyrun // International Journal of Computers & Technology. – 12(2). – 2013. – P. 3211–3220. 8. Висоцька В. Моделювання етапів життєвого циклу комерційного web-контенту / В. Висоцька, Л. Чирун, Л. Чирун // Вісник Нац. ун-ту “Львівська політехніка”. – 2011. – № 715. – С. 69–87. 9. Висоцька В. Особливості проектування та впровадження систем електронної комерції / В. Висоцька // Вісник Нац. ун-ту “Львівська політехніка”. – 2008. – № 631. – С. 55–84. 10. Vysotska V. Analysis and evaluation of risks in electronic commerce / V. Vysotska,

I. Rishnyak, L. Chyrun // CADSM'07. – Lviv, 2007. – P. 332–333. 11. Vysotska V. Linguistic Analysis of Textual Commercial Content for Information Resources Processing / V. Vysotska // TCSET'2016. – 2016. Lviv-Slavske, Ukraine. – P. 709–713. 12. Бісікало О. Виявлення ключових слів на основі методу контент-моніторингу україномовних текстів / О. Бісікало, В. Висоцька // *Радіоелектроніка. Інформатика. Управління*. – 1(36). – Запоріжжя: ЗНТУ. – 2016/1. – С. 74–83. 13. Chyrun L. Informational resources processing intellectual systems with textual commercial content linguistic analysis usage constructional means and tools development / L. Chyrun, V. Vysotska, I. Kozak // *Econtechmod*. – Lublin ; Rzeszow, 2016. – 5(2). – P. 85–94. 14. Кондратєв Є. Контент-аналіз текстових масивів даних / Є. Кондратєв, В. Висоцька // *4 ІКС-2015*. – 2015, Львів-Славське. – С. 170–171. 15. Vysotska V. Features of the content-analysis method for text categorization of commercial content in processing online newspaper articles / V. Vysotska, L. Chyrun // *ACS journal*. – 11(1). – Poland, 2015. – P. 5–19. 16. Vysotska V. Linguistic analysis and modelling semantics of textual content for digest formation / V. Vysotska, L. Chyrun // *MEST Journal*. – 3(1). – P. 127–148. 17. Висоцька В. Особливості моделювання синтаксису речення слов'янських та германських мов за допомогою породжувальних контекстно-вільних граматик / В. Висоцька // *Вісник Нац. ун-ту “Львівська політехніка”*. – № 814. – Львів, 2015. – С. 246–276. 18. Висоцька В. Особливості рубрикації текстового комерційного контенту / В. Висоцька // *Вісник Нац. ун-ту “Львівська політехніка”*. – 2015. – № 826. – С. 359–367. 19. Застосування контент-аналізу для опрацювання текстових масивів даних / [Я. Кісь, В. Висоцька, Л. Чирун, В. Фольтович] // *Вісник Нац. ун-ту “Львівська політехніка”*. – 2015. – № 814. – С. 282–292. 20. Чирун Л. Особливості методів контент-аналізу текстових масивів даних web-ресурсів в межах регіону / Л. Чирун, В. Кучковський, В. Висоцька // *Вісник Нац. ун-ту “Львівська політехніка”*. – 2015. – № 829. – С. 296–320. 21. Шестакевич Т. Моделювання семантики речення природною мовою за допомогою породжувальних граматик / Т. Шестакевич, В. Висоцька, Л. Чирун, Л. Чирун // *Вісник Нац. ун-ту “Львівська політехніка”*. – 2015. – № 814. – С. 335–352. 22. Бісікало О. Експериментальне дослідження пошуку значущих ключових слів україномовного контенту / О. Бісікало, В. Висоцька // *Вісник Нац. ун-ту “Львівська політехніка”*. – № 829. – 2015. – С. 255–272. 23. Берко А. Лінгвістичний аналіз текстового комерційного контенту / А. Берко, В. Висоцька, Л. Чирун // *Вісник Нац. ун-ту “Львівська політехніка”*. – 2015. – № 814. – С. 203–227. 24. Vysotska V. Generative regular grammars application to modeling the semantics of sentences in natural language / V. Vysotska // *Вісник Нац. ун-ту “Львівська політехніка”*. – 2014. – № 808. – С. 43–56. 25. Висоцька В. Особливості генерування семантики речення природною мовою за допомогою породжувальних необмежених та контекстно-залежних граматик / В. Висоцька // *Вісник Нац. ун-ту “Львівська політехніка”*. – 2014. – № 783. – С. 271–292. 26. Висоцька В. Застосування породжувальних граматик для моделювання синтаксису речення / В. Висоцька, Т. Шестакевич, Ю. Щербина // *Вісник Нац. ун-ту “Львівська політехніка”*. – 2012. – № 743. – С. 175–190. 27. Берко А. Застосування методу контент-аналізу для формування інформаційних ресурсів в системах електронної контент-комерції / А. Берко, В. Висоцька, М. Сороковський // *Вісник Нац. ун-ту “Львівська політехніка”*. – 2012. – № 743. – С. 3–15. 28. Висоцька В. Утворення речень англійською та німецькою за допомогою породжувальних граматик / В. Висоцька, Т. Шестакевич, Ю. Щербина // *Вісник Нац. ун-ту “Львівська політехніка”*. – 2012. – № 744. – С. 142–152. 29. Висоцька В. Інтелектуальна система розподілу дайджестів між працівниками електронних засобів масової інформації / В. Висоцька, О. Окрушко // *Вісник Нац. ун-ту “Львівська політехніка”*. – 2012. – № 744. – С. 41–53. 30. Щербина Ю. Утворення українських дієприкметників за допомогою породжувальних граматик / Ю. Щербина, Ю. Нікольський, В. Висоцька, Т. Шестакевич // *Вісник Нац. ун-ту “Львівська політехніка”*. – 2011. – № 715. – С. 354–369. 31. Проект створення відкритих словників для перевірки орфографії української мови. – Режим доступу : https://code.google.com/archive/p/spell-uk/wikis/aspell_uk.wiki. 32. Hunspell/Aspell for Ukrainian. – Режим доступу : <https://sourceforge.net/projects/ispell-uk/>.