

СТРУКТУРНО-ФУНКЦІОНАЛЬНИЙ РІВЕНЬ ОРГАНІЗАЦІЇ ЛІНГВІСТИЧНОГО ПРОЦЕСОРА

Ю.І. КИСЛЕНКО, А.В. ХІМІЧ

Анотація. Мовленнєва діяльність людини актуалізується індивідуальною мовною системою, яка (за Л. Щербою) визначається взаємодією лінгвістичного процесора, відповідального за структурний рівень організації мовленнєвої діяльності, та природно-мовної бази знань, де зберігається весь поточний когнітивний потенціал як символічного, так і образного рівня. Із позицій урахування інтеграційного підходу до аналізу мовленнєвої діяльності презентовано структурно-функціональний рівень організації лінгвістичного процесора для опрацювання повідомлень моно/поліпредикатних рівнів. Такий підхід, що поєднує особливості формування та взаємодії лінгвістичного процесора з базою знань, є плідною основою моделювання індивідуальної мовної системи для вирішення важливих проблем інформаційних технологій

Ключові слова: лінгвістичний процесор, індивідуальна мовна система, базова семантико-синтаксична структура, квант знань, ситуація, інформаційні природно-мовні технології, інформаційна база лінгвістичного процесора.

ІНТЕГРАЦІЙНИЙ ПІДХІД ДО АНАЛІЗУ ТЕКСТОВОЇ ІНФОРМАЦІЇ

Моделювання мовленнєвої діяльності постає натеper однією з найбільш актуальних і затребуваних у різних сферах наукових та прикладних досліджень проблем сьогодення, починаючи від перспектив розвитку Інтернет (семантичний Web і формування узагальненої моделі знань) і закінчуючи аналізом складних проблем когнітивно-комунікативних процесів з виходом на якісний перекилад. Усі ці питання пов'язані з аналізом, розумінням та моделюванням мовленнєвої діяльності людини, найбільш адекватним проявом та відтворенням якої зрештою є її здатність до аналізу та синтезу текстової інформації. Мовленнєва діяльність особистості актуалізується *індивідуальною мовною системою* (ІМС). У науковий обіг поняття ІМС уведено нашим співвітчизником Л. Щербою [1] ще в тридцятих роках минулого століття як сукупності лінгвістичного процесора (ЛП), відповідального за структурний рівень мовної організації та бази знань (БЗ), де постійно нагромаджується когнітивний потенціал людини. У загальному випадку структуру ІМС, що охоплює не лише інформацію мовного, а й образного (сенсорного) рівня, зображено на рис. 1.

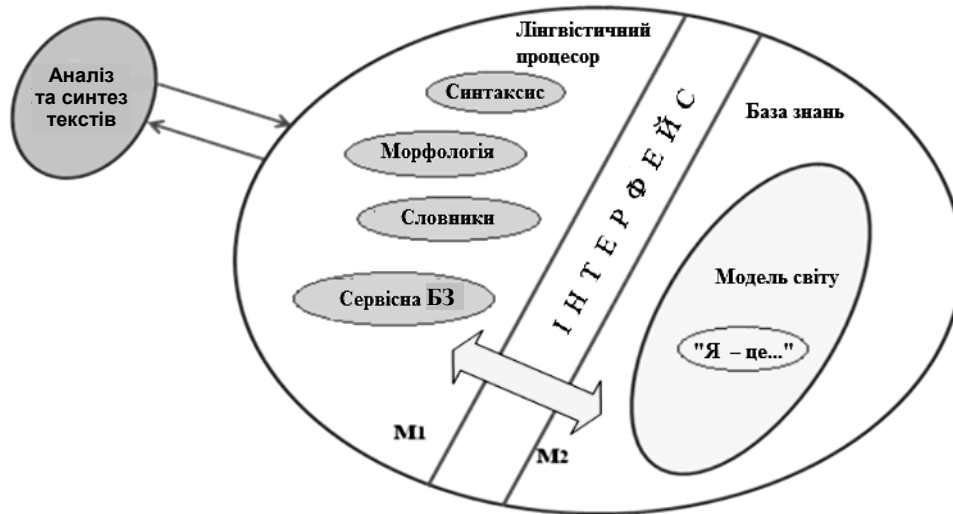


Рис. 1. Індивідуальна мовна система: лінгвістичний процесор — відповідальний за структурний рівень організації повідомлення; база знань — когнітивний потенціал суб'єкта

Лінгвістичний процесор охоплює всі знання про структурний рівень мовної організації, що забезпечують процедури синтезу/аналізу повідомлення. **База знань** нагромаджує сукупні знання як символічного (мовного), так і сенсорного рівня. **Інтерфейс** — схема взаємодії ЛП–БЗ. Важливо враховувати, що складові ЛП і БЗ диференціюють знання про структурне та змістовне навантаження БЗ, хоча обидві складові актуалізовані в єдиній нейросубстанції.

У загальному випадку ІМС спирається як на знання про структурно-функціональний рівень мовної організації, що акумульовані в сучасних підручниках та академічних виданнях (рівень ЛП), так і на весь нагромаджений натеper когнітивний потенціал окремого індивіда, отриманий через сприйняття інформації як образного (сенсорного), так і символічного рівня та збережений в пам'яті людини.

Структурно-функціональний рівень організації ЛП визначається структурним рівнем мовної організації. Ретельно ця проблема висвітлена автором у праці «Back to Basics of Speech Activity» [2], де з позицій інтегрування різнопланових напрямів сучасних досліджень мовленнєвої діяльності з боку біологів, нейрофізіологів, психологів, філософів, кібернетиків і, звісно ж, лінгвістів сформований узагальнений інтеграційний підхід до аналізу мовленнєвої діяльності.

У цій праці подано також аксіоматику мовленнєвої діяльності, до висновків окремих положень якої будемо постійно звертатися. Одним з найважливіших положень постає думка Б.Ю. Городецького, що «багато недоречностей в мовознавстві пов'язані з фактом, що до цього часу мова все ще вважається схемою відтворення думки, а не способом організації та подання знань» [3]. Це важливий дороговказ на шляху вирішення заявленої проблематики.

Фрагментарно головні висновки інтеграційного підходу зводяться до такого: чітко визначено поняття **ситуації** як окремого кванта знань образного рівня, що сприймається зоровим трактом. **Ситуація** — це фрагмент

зорової складової довкілля, що потрапляє на центральну ямку сітківки та опрацьовується за повною програмою. А що це за повна програма, відповідь знаходимо у Семіра Зекі [4], який узагальнив результати багаторічних досліджень зорового тракту нейрофізіологами. **На третьому рівні сітківки вирішується проблема «фон-фігура» — визначаються об'єкти/суб'єкти, на четвертому/п'ятому — ідентифікується динаміка рухомих об'єктів, а далі (в зоровій ділянці головного мозку) визначаються всі ознаки знайдених складових та міра цих ознак.** Ці функції власне і визначають повну програму опрацювання конкретної ситуації, які надалі на образному рівні передаються в глибинні структури мозку. Якщо окрема ситуація довкілля визначається **триєдністю категорій часу, простору та дії** (філософська концепція), то і її відбиток у нейромережі також визначається пов'язаною множиною відповідних елементів нейросубстанції.

Коли ж у людини далі виникає комунікативна інтенція стосовно трансляції цих знань з образного на мовний рівень, то для адекватного відтворення окремої ситуації лишається тільки єдина можливість — «наклеїти» на визначені складові окремої ситуації (Obj, Subj, Mov), їх ознаки Attr(Obj), Attr(Subj), Attr(Mov) та ще й міри всіх ознак Attr (Attr) мовні ярлички — слова та пов'язати їх між собою відповідним чином для відтворення триєдності їх множини у часі, просторі та дії. Саме функціональне навантаження зорового аналізатора і визначає диференціацію частин мови, що сформувалася на довгому еволюційному шляху розвитку людини. Коли ж згадати ще й тезу лінгвістів, що будь-яка ситуація зовнішнього світу може бути презентована лише чотирма частинами мови, то чітко бачимо, що **мовна презентація ситуації є похідною від структурно-функціонального рівня нейроорганізації зорового тракту.** Цей висновок презентується табл. 1.

Таблиця 1. Співвідношення між системами опрацювання образної та мовної інформації

№ з/п	Функціональне навантаження зорового аналізатора	Відповідні мовні категорії (частини мови)
1	Obj / Subj	Іменник
2	Attr(Obj/Subj)	Прикметник
3	Attr(Attr)	Прислівник
4	Mov	Дієслово
5	Attr(Mov)	Прислівник
6	Attr(Attr)	Прислівник

Вербалізація (трансляція на мовний рівень) окремої ситуації актуалізується у вигляді окремої **базової семантико-синтаксичної структури** (БССС), яка є основою формування будь-якого тексту. **Базова семантико-синтаксична структура — це двоскладова монопредикатна схема опису довільної ситуації реального або віртуального світу, всі складові якої актуалізовані на атрибутивному рівні.** Структурний рівень організації БССС зображено на рис. 2.

Важливо, що з посиланням на О.М. Гвоздева [5] уже матимемо основні етапи становлення мовленнєвого ладу дитини, що тривають близько трьох

(2,5 – 3) років. У цьому віці дитина вже опановує схему транслявання окремої ситуації з образного рівня на символічний — мовний. Це головний етап становлення мовленнєвої організації людини: з одного боку, дитина вже опанувала загальну схему вербалізації довільної ситуації, а з другого — вона вже може користуватися і проміжними (неповними) структурами, починаючи від одно-, дво-, трискладових конструкцій і закінчуючи повідомленнями з обмеженням (сім плюс-мінус дві) складових, що визначається структурно-функціональною організацією сітківки. Цей період визначає можливість опанування структур монопредикатного рівня.

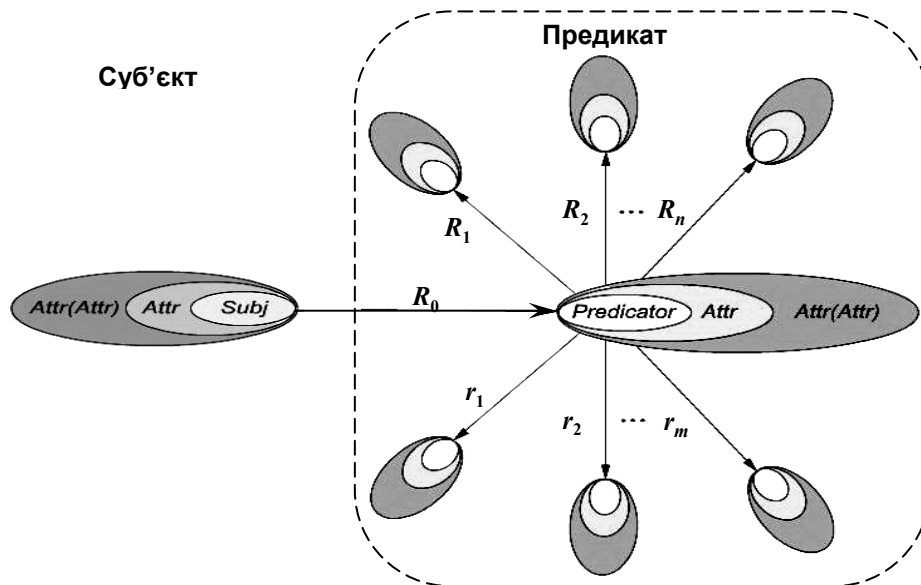


Рис. 2. Базова семантико-синтаксична структура. **Subj** — суб'єкт БССС; **Pred** — предикат як узагальнена характеристика дії; **Predicator** — ядро предиката; $r_1, r_2, r_3, \dots, r_m$ — ситуаційні відношення; R_1, R_2, \dots, R_n — предикативні відношення, що формують модель керування дієслова

Наступний рівень опанування мовної організації характеризується здатністю формування та використання повідомлень поліпредикатного рівня, де в одному повідомленні поєднується множина структур монопредикатного рівня. Це відкриває шлях до схем сурядного та підрядного поєднань багатьох структур БССС, виходячи згодом на рівень рекурсивної організації повідомлення. Отже, у структурному сенсі організації текстової інформації (за ідеологією ВІСА-14 [2]) виходимо на дослідження структурного рівня організації повідомлення за дихотомією моно/поліпредикатної схеми організації повідомлення.

Висновок інтеграційного підходу досить радикальний і настільки ж продуктивний: всі ми володіємо мовою, не знаючи часто-густо як вона організована лише за однієї умови — будь-який текст, як і мовна організація загалом визначаються лише однією стандартною структурою опису довільної ситуації, яку вона опановує у віці двох – трьох років.

У загальному випадку довільний текст виражається множиною однотипних структур, організованих на моно/поліпредикатному рівні. Коли дитина/людина опанувала структурний рівень мовної організації, виходячи на рекурсивний рівень організації повідомлення, це означає, що вона, сприй-

маючи довільне текстове/мовне повідомлення, абсолютно вільно (хоча може і підсвідомо) виконує його декомпозицію за окремими квантами знань, кожен з яких окреслюється стандартною структурою БССС. Завдання ж лінгвістичного процесора — перевести на формалізований рівень процедуру декомпозиції довільного повідомлення за стандартними однотипними структурами БССС з ідентифікацією як усіх складових кожної структури, так і відношень між ними.

Тож, коли намагаємось перекласти на комп'ютер функції лінгвістичного процесора як складової ІМС, маємо послідовно реалізувати дві функції: *по-перше, виконати декомпозицію повідомлення за окремими квантами знань у вигляді структур БССС, по-друге, проаналізувати структурний рівень організації кожного кванта знань з визначенням його складових Obj, Subj, Mov з урахуванням їх атрибутивного оточення, а також ідентифікувати відношення ситуаційного та предикативного рівнів для кожного предиката.* Останнє зауваження визначається принциповим етапом моделювання процесу розуміння повідомлення. Якщо знову звернутися до структури БССС, презентованої рис. 2, то чітко вбачаємо цю відмінність — *людина, сприймаючи таке повідомлення, стикається виключно лише з лексичною її презентацією, тоді як система відношень R_0 , ситуаційні відношення $r_1, r_2, r_3, \dots, r_m$ та предикативні відношення R_1, R_2, \dots, R_n , що формують модель керування дієслова, ідентифікуються лише за рахунок мовної компетенції, сформованої людиною ще у процесі опанування мови, яка зберігається в пам'яті БЗ і не потребує словесної сатисфакції.*

Результати аналізу кожного повідомлення потрібно запам'ятовувати, імплементувавши їх до когнітивного потенціалу БЗ. Завдання, на перший погляд, виглядає невідомим або досить важким, проте матеріали інтеграційного підходу відкривають певні орієнтири для його вирішення. Проблема розбудови структурно-функціонального навантаження ЛП поділяється на етапи опрацювання структур моно/поліпредикатних рівнів.

ФУНКЦІОНАЛЬНЕ НАВАНТАЖЕННЯ ЛІНГВІСТИЧНОГО ПРОЦЕСОРА

Функціональне навантаження ЛП визначається головними етапами становлення мовленнєвої діяльності людини, презентованої матеріалами роботи ВІСА-14 [2]. При цьому *когнітивний потенціал ЛП (як здатність опрацьовувати тексти довільної складності) повинен поступово нарощуватися, починаючи від структур монопредикатного рівня (проходячи послідовно етапи від односкладного повідомлення до повної структури БССС) та виходячи згодом на рівень опрацювання рекурсивних схем організації повідомлень довільної складності.* Тож варто чітко диференціювати, власне, і етапи проектування ЛП, орієнтованих спочатку на опрацювання текстів монопредикатного рівня, а далі і поліпредикатного. Важливо завжди пам'ятати, що ЛП індивідуальної мовної системи кожного суб'єкта завжди спирається на когнітивний потенціал, презентований його БЗ, і нова інформація завжди опрацьовується з урахуванням поточного отриманого досвіду. *Як тільки нове повідомлення опрацьовано лінгвістичним процесором і передано у БЗ, воно відразу ж презентується вже як контекст до певного фрагмента нейромережі та поповнює когнітивний потенціал суб'єкта.*

Це принципова відмінність живої нейросубстанції від комп'ютерних моделей, де поки що індивідуальна мовна система задекларована на концептуальному рівні без реального наповнення складових ЛП та БЗ. Зауважимо, що на кафедрі технічної кібернетики НТУУ «КПІ ім. Ігоря Сікорського» в межах розвитку інформаційних природно-мовних технологій принципово опрацьований структурно-функціональний рівень організації природно-мовної бази знань [6]; тож актуальною постає вже проблема розроблення лінгвістичного процесора як важливої іншої складової індивідуальної мовної системи. На часі тепер завдання — наповнити обидві складові потужним когнітивним потенціалом, здатним вирішувати нагальні проблеми сучасних інформаційних природно-мовних технологій. Проект орієнтований на опрацювання текстів флективних мов, складніших у структурному сенсі відносно аналітичних.

Структурно-функціональне навантаження ЛП монопредикатного рівня

Лінгвістичний процесор у загальному випадку визначає мовну компетенцію щодо структурно-функціонального рівня мовної організації, яка (згідно з працею О.М. Гвоздева [5]) формується поступово, починаючи від опанування структури БССС монопредикатного рівня (вік 2,5–3 роки) і закінчуючи у віці близько 10–12 років опануванням повідомлень поліпредикатного рівня, виходячи вже на рівень сприйняття рекурсивної організації повідомлення. Звісно, ці межі дуже залежні від умов виховання та навчання дитини, проте певним чином вони окреслюють інтервали становлення мовної компетенції у повному обсязі.

Як монопредикатні, так і поліпредикатні рівні спираються на однакову систему відношень — ситуаційних та предикативних, але, у першому випадку вони поєднують між собою лише складові структури БССС, що ідентифікуються окремими частинами мови, тоді як у другому — трансформуючись певним чином, поєднують між собою вже окремі структури БССС. Таке співвідношення моно/поліпредикатних структур і зумовлює особливості формування організації ЛП: на першому етапі проектування враховуються особливості презентації повідомлень монопредикатного рівня, тоді як на другому — ідентифікуються вже залежності формування структур поліпредикатного рівня.

Структура лінгвістичного процесора монопредикатного рівня (ЛП-моно) визначається однозначно структурною організацією БССС (рис. 2). Мабуть, розглядаючи це питання, потрібно певним чином враховувати особливості нейроорганізації зорового тракту, що опрацьовує довкілля поетапно — з дискретою в одну ситуацію, яка визначається окремим **квантом знань образного рівня**, з наступною її вербалізацією (коли виникає в цьому потреба) у вигляді структури БССС, яку вважатимемо **квантом знань символічного (мовного) рівня**.

Нагадаймо послідовність сприйняття довкілля зоровим трактом. Довкілля квантується очима окремими ситуаціями обсягом сім плюс/мінус дві складові Obj, Subj, що зумовлюється актуальним зоровим полем; серед них визначаються динамічні характеристики Mov (усіх або одного об'єкта); далі ця інформація надходить у зорову кору головного мозку, де визначаються всі ознаки складових ситуації Attr(Obj/Subj), Attr(Mov) та ще й міра

цих ознак у вигляді складових $Attr(Attr)$. Для кожної ситуації визначаються також просторово-часові характеристики, які порівнюються з попередніми їх оцінками. За такою схемою опрацьовується кожна ситуація, бо цей процес — отримання адекватної інформації про поточний стан довкілля — є надважливим для виживання людини у швидкоплинному динамічному середовищі.

Коли ж у людини виникає потреба транслювати отриману інформацію на мовний рівень, вона її послідовно транслює у вигляді множини структур БССС, пов'язаних відповідним чином. Мовне повідомлення аналізується поетапно з дискретою в одну ситуацію. Проаналізуємо цю процедуру: завданням для ЛП-моно постає процедура аналізу повідомлення, тобто однозначне визначення всіх складових ($Subj, Obj_1, \dots, Obj_n, Mov$), включаючи їх атрибутивне оточення $Attr(Obj/Subj), Attr(Mov)$, та міру цих ознак $Attr(Attr)$, а також ситуаційних відношень $r_1, r_2, r_3, \dots, r_m$ (час, простір, причина, умова, наслідок тощо) та предикативних відношень (R_1, R_2, \dots, R_n), що формують модель керування дієслова.

Множина структур БССС, презентованих у межах окремого абзацу, є зв'язною сукупністю цих структур, триєдиних у часі, просторі та динаміці. Сучасні надбання класичної лінгвістики багато в чому дозволяють виконувати процедуру аналізу повідомлень монопредикатного рівня. Проте знову доводиться озвучувати тезу, що *індивідуальна мовна система завжди функціонує за рахунок плідної взаємодії ЛП, відповідального за аналіз структурного рівня мовної організації, і БЗ, де зберігається та нагромаджується весь когнітивний потенціал, на якому відбувається змістовна (семантична) інтерпретація кожного кванта знань у вигляді БССС.*

Можемо тепер окреслити поетапно окремі кроки на шляху сприйняття ЛП повідомлення у вигляді множини стандартних структур БССС. Важливо враховувати, що аналіз довільного повідомлення будь-якого рівня (моно/полі) потрібно починати обов'язково з ідентифікації комунікативних засобів, які принципово не входять до структури БССС. Перелік та класифікацію таких засобів можна знайти в монографії [7], які визначають окремі аспекти ставлення автора до конкретної інформації, презентованої поточним квантом знань.

Отже, аналіз окремого повідомлення монопредикатного рівня, окресленого окремою структурою БССС, поділяється на окремі етапи:

1. **Ідентифікація комунікативних засобів повідомлення** (звісно, коли вони там є), що визначають певне ставлення автора як до окремих складових, так і загалом до повідомлення, презентованого структурою БССС.

2. **Ідентифікація складових БССС** з визначенням частин мови, їх характеристик (множина, одинина, відмінки, час тощо). На сучасному етапі розвитку інформаційних технологій існують і використовуються досить потужні словники, які багато в чому допомагають виконувати етапи визначення частин мови, їх характеристик та особливості використання. Це дозволяє успішно ідентифікувати усі складові з урахуванням їх атрибутивного оточення.

3. **Ідентифікація ситуаційних відношень кожної БССС** вбачається досить складним етапом. Дитина опановує їх у послідовності: дія, час, простір, причина, умова, наслідок тощо (О.М. Гвоздев [5]), постійно

аналізуючи співвідношення отриманої інформації з когнітивним її надбанням. Натепер нагромаджено багато досліджень з цього напрямку, що можуть використовуватися для реалізації ЛП. Можемо подати перелік праць з аналізу часових та просторових відношень (Д.О. Поспедов [8], Ю.І. Кисленко [7], Ю.І. Кисленко, О.С. Черевко [9], Ю.І. Кисленко, А.В. Хіміч [10]).

4. **Ідентифікація предикативних відношень R_1, R_2, \dots, R_n** (моделей керування окремого дієслова) досі вбачається досить проблематичним, оскільки натепер ще не формалізовані ці дані для комп'ютерних технологій. Проте проглядається можливість їх формування та використання в процесі нагромадження когнітивного потенціалу БЗ.

Структурно-функціональне навантаження ЛП поліпредикатного рівня

Згідно з матеріалами ВІСА-14 [2] після засвоєння дитиною (людиною) монопредикатного рівня організації повідомлення та якщо досить досконало опрацьована система відношень окремої структури БССС (як ситуаційного, так і предикативного рівня), то згодом мовець переходить уже до опрацювання та використання повідомлень поліпредикатного рівня. Це означає, що на «вакантне місце» кожної складової структури БССС ставиться вже не слово, а окрема ситуація, тобто виникає можливість опанування поліпредикатні схеми організації повідомлення, що, у свою чергу, уможливує опанування та використання (синтезу/аналізу) рекурсивних схем організації повідомлення. При цьому дещо змінюється схема інкорпорації однієї структури в іншу, і саме це — розуміння особливостей формування поліпредикатного рівня організації повідомлення — дозволяє впевнено користуватися рекурсивною схемою організації комунікативного процесу. Аналізуючи такі схеми формування та опрацювання подібних повідомлень, важливо диференціювати окремі схеми взаємодії множини структур БССС у межах цілісного поліпредикатного повідомлення. Особливості формування таких структур подамо у стислому вигляді, хоча частково їх презентовано у праці [7].

Схеми ускладнення предикатора. Існує ряд об'єктивно-суб'єктивних важелів відтворення ситуацій, коли неможливо адекватно ідентифікувати окрему ситуацію, не враховуючи цілий спектр характеристик або відношень самого суб'єкта до певної динамічної ситуації. Прикладом такої конструкції може бути ситуація: *«Хлопець бажає почати навчатись опанувати професію пілота»*. У загальному випадку таку ситуацію можна подати схемою, коли предикатор презентує не лише динамічні особливості складових ситуацій, але й охоплює значно ширший спектр динамічних характеристик як власне суб'єкта, так і дії, включаючи кілька предикатів, кожен з яких охоплює певні динамічні характеристики дії: *модальність виконання дії, стадійність процесу, можливість навчання, опанування фаху* тощо. Зауважимо, що такий ускладнений предикат підпорядкований лише одному суб'єкту та ще й визначається певною послідовністю актуалізації. Річ у тім, що в реальному середовищі та у складних схемах взаємодії людини з динамічним довкіллям дуже часто майже неможливо адекватно окреслити динамічну ситуацію лише одним конкретним предикатором, не враховуючи взаємодію з іншими. Такий ускладнений предикат презентується ієрархічною множиною окремих підпорядкованих предикатів, коли кожен з наступних

уточнює попередній. Така схема формування поліпредикатного повідомлення часто подається в реальних текстах і має також опрацьовуватися ЛП.

Схеми трансформування предикаторів. Ця схема має широкі можливості трансформування предикатора (дієслова) у різновиди атрибутивного рівня у формах дієприкметника, дієприслівника або номінативної функції субстантиву. Тоді маємо схему організації поліпредикатного повідомлення, де один предикатор у формі дієслова Mov трансформується за зовнішніми флективними ознаками в дієприкметник, виконуючи функцію атрибутивної складової (Attr (Obj/Subj)) будь-якого елемента кореневої структури, чи в дієприслівник, окреслюючи ознаки конкретного предикатора (Attr (Mov)), або в субстантив, актуалізований вже повним предикатом. Особливість таких схем трансформування в тому, що у кожному випадку **трансформується лише предикатор Mov у ту чи іншу схему, тоді як власне весь предикат (у сукупності множини ситуаційних та предикативних складових) лишається майже у незмінному вигляді.** Звісно, лінгвістичний процесор повинен ідентифікувати такі різновиди взаємодії двох структур, ураховуючи певні особливості їх актуалізації. Сфери використання схем трансформування структур БССС за вказаними напрямками можуть різнитися для різних мов. Особливо це потрібно враховувати у перекладі з однієї мови на іншу.

Поліпредикатна інтерпретація повідомлень з однорідними членами речень. До структур поліпредикатного рівня варто віднести і повідомлення з однорідними членами. Таке повідомлення реципієнтом сприймається як множина однотипних структур, що можуть відрізнятися одним елементом або їх множиною. Проте, коли завданням лінгвістичного процесора є аналіз структурно-функціонального рівня організації довільного повідомлення (**що особливо важливо як для адекватного сприйняття та «розуміння» повідомлення людиною, так і не менш важливим визначається для адекватного опрацювання повідомлень у сучасних системах інформаційних технологій**), то принципово важливим є питання врахування та опрацювання таких структур. Усі можливі схеми формування повідомлень з однорідними членами можуть бути формально презентовані двома структурами БССС — **«хмаринками» структур БССС, що довільним чином «наповзають» одна на одну.**

Формування структур поліпредикатних повідомлень за функціональним навантаженням елементів БССС. Чітке розуміння можливих схем організації повідомлення за схемами моно/поліпредикатних рівнів дає змогу значним чином поліпшити процедуру автоматичного опрацювання текстової інформації.

Якщо знову звернутися до витоків мовленнєвої діяльності та згадати, що структура БССС є похідною від структурно-функціонального рівня нейроорганізації зорового тракту з визначенням складових Obj/Subj на множині ознак Attr (Obj/Subj) та їх міри Attr(Attr), а також з ідентифікацією їх динамічних характеристик Mov в околі ознак Attr (Mov) та Attr(Attr) (див. рис. 2), то можемо презентувати ще одну можливість формування повідомлень поліпредикатного рівня.

Якщо звернутися знову до структури складових елементів БССС, то виходимо ще на один рівень формування структур поліпредикатного рівня.

Приклад формування таких повідомлень знаходимо у «Лісовій пісні» Л. Українки: «Той, що в скелі сидить», «Той, що греблю рве на волі». У загальному випадку — це поліпредикатні структури, де головне речення презентоване лише вказівним займенником «той», а підрядне визначає власне окрему ситуацію, презентовану предикатом «Той сидить в скелі». Це приклад використання поліпредикатної схеми, де підрядна конструкція ідентифікує суб'єкт (об'єкт) головного речення, презентований окремою структурою БССС.

Аналіз такої схеми організації повідомлення поліпредикатного рівня підштовхнув автора до цікавого узагальнення, пов'язаного з функціональним навантаженням займенників. Займенник — це однозначна частина мови, статус якої дуже розпорошений, невизначений та нечітко окреслений; однак функціонально він може ідентифікувати (замінювати) будь-яку частину мови. Подамо приблизну класифікацію займенників за їх функціональним навантаженням (табл. 2).

Таке розмаїття та багатофункціональне навантаження займенників відкриває ще пласт можливих схем формування структур поліпредикатного рівня, який покриває функціональне навантаження всіх складових БССС. Варто зауважити, що людина однозначно та безпомилково фіксує їх семантичне навантаження, тоді як для моделювання відповідних залежностей — це досить складна проблема. На монопредикатному рівні ці засоби виконують роль окремих складових БССС, тоді як на поліпредикатному рівні вони постають у парному варіанті, поєднуючи вже окремі структури БССС, одна з яких ідентифікується займенниками, а друга презентується вже цілою структурою.

Таблиця 2. Функціональне навантаження займенників відносно складових БССС (монопредикатний рівень)

Тип займенника	Семантичне навантаження	Функції БССС
Особові	я, ти, він, вона, воно, ми, ви, вони	Subj, Obj
Безособові	себе (не має ні роду, ні числа)	Subj
Присвійні	мій, твій, його, її, наш, ваш, їх, їхній	Attr(Obj/Subj)
Вказівні	цей, сей, той, та, те	Attr(Obj/Subj)
Означальні	весь, всякий, кожний, інший	Attr(Obj/Subj)
Питальні	хто, що, який, котрий	Attr(Obj/Subj), Obj
Заперечні	ніхто, ніщо	Subj, Obj
Неозначені	дехто, дещо, хто-небудь, що-небудь	Subj, Obj
Вказівні	такий, який, котрий	Attr(Attr)

Для формування структур поліпредикатного рівня найчастіше використовуються певні схеми, ідентифікаторами яких постають співвідношення елементів: «той, хто . . .», «... того, кого ...», «... тому, кому . . .» тощо. Аналогічна схема може використовуватися у випадку пояснення атрибутивного оточення (Attr (Obj/Subj)) певного елемента головного речення з використанням відповідних елементів «...такий, який ...», «...такому, що...», або ще і зв'язок може формуватися на рівні міри атрибутивного елемента Attr(Attr) — аналог системи «...так, як...». Як бачимо, маємо безмежну (але добре структуровану за канонами БССС) схему формування повідомлень поліпредикатного рівня. Варіанти таких схем взаємодії повідомлень поліпредикатного рівня можуть бути подані досить широко у вигляді табл. 3.

Таблиця 3. Схеми формування повідомлень поліпредикатного рівня

Ситуаційні відношення	Схеми формування структур поліпредикатного рівня	Приклади реалізації
Часові	Коли...БССС1..., то ... БССС2...	Коли ..., то ...
Просторові	Там...БССС1..., де ... БССС2...	Там ..., де....
Причини	Через...БССС1... буде ... БССС2	Через ... буде щось.
Умови	За умови БССС1... буде ... БССС2	За умови – буде щось
Наслідок	У результаті БССС1 буде ... БССС2	У результаті буде
Порівняння	БССС1 ..., оскільки БССС2...	Зроблено так, як ...

ПРИКЛАДНІ АСПЕКТИ ВИКОРИСТАННЯ ЛІНГВІСТИЧНОГО ПРОЦЕСОРА

Вище проаналізовано досить ретельно особливості організації та використання мовних засобів формування повідомлень як монопредикатного рівня, що використовуються в засобах вербалізації окремої ситуації, так і поліпредикатного, які визначають особливості вербалізації вже окремого цілісного фрагмента дійсності, презентованого множиною окремих ситуацій, триєдиних у часі, просторі та дії. Інтеграційний підхід до аналізу мовленнєвої діяльності зумовлює чіткий системний погляд на структурний рівень мовної організації, що спирається на поняття окремого кванта знань, презентованого структурою БССС, і реалізується на моно/поліпредикатних рівнях.

Перш ніж окреслити перспективи моделювання ІМС, важливо визначити те підґрунтя і підстави, які забезпечують такий оптимізм. Знакова публікація автора у цьому напрямі — це праця «Back to basics of speech activity» [2], де з позицій інтеграційного підходу (із залученням досліджень нейрофізіологів, філософів, психологів, кібернетиків та ін.) чітко були визначені поняття ситуації як окремого кванта знань образного рівня та схема її вербалізації у вигляді структури БССС. Там чітко й аргументовано показано, що людина впевнено володіє мовою лише тому, що їй на всі випадки життя достатньо лише однієї стандартної структури. Професор Києво-Могилянської академії, доктор філологічних наук О.М. Демська після ретельного ознайомлення з цією працею чітко сформулировала свій вердикт: *«Ця робота — вершина структуралізму»*, бо все розмаїття структурного рівня мовної організації спирається лише на одну-єдину структуру, що постає похідною від структурно-функціонального рівня нейроорганізації зорового тракту та використовується на моно/поліпредикатному рівні.

Ця робота стала поштовхом до моделювання індивідуальної мовної системи як пов'язаної сукупності ЛП та БЗ (див. рис. 1). Важливо, що такий комплексний підхід до моделювання мовленнєвої діяльності відкриває шлях до вирішення вузлових проблем інформаційних природно-мовних технологій, визначених інтернет-спільнотою як «Семантичний Web» та «Формування моделі світу», без вирішення яких не варто сподіватись на принципове трансформування якості пошуку.

Натепер роботи з моделювання індивідуальної мовленнєвої системи виконують аспіранти на кафедрі технічної кібернетики НТУУ «КПІ» ім. Ігоря Сікорського за двома напрямками: «Проектування природно-мовної бази знань» (Д.С. Сергеев) та «Формування лінгвістичного процесора» (А.В. Хіміч). Уже достатньо повно опрацьовано конструктивний підхід до структурно-функціонального рівня організації природно-мовної бази знань, презентований працею «*Cognitive architecture of speech activity and modeling thereof*» [6]; на часі також опрацювання структурно-функціонального рівня організації ЛП.

Функціональне навантаження лінгвістичного процесора для синтезу тексту монопредикатного рівня. Лінгвістичний процесор як невід’ємна складова ІМС може і повинен працювати в двох режимах — синтезу та аналізу мовного повідомлення. Послідовно проаналізуємо ці режими.

У контексті моделювання ЛП, орієнтованого на виконання функцій синтезу мовного матеріалу, цілком слушним буде посилання на попередню працю автора, виконану у співпраці з Г.А. Амплєєвою [10] ще в 2000 р. Цей програмний продукт може сприйматися тепер уже в дещо іншій площині, а саме: у руслі моделювання роботи ЛП для синтезу тексту через послідовність структур БССС монопредикатного рівня.

Вхідні дані. Об’єкт дослідження — процес автоматичного формування опису сцени, синтезованої довільним чином на екрані монітора. У пам’яті комп’ютера синтезовано дві невеликі БЗ, що моделювали фрагмент візуального світу: перша — стосувалася візуалізації на сцені графічної інтерпретації просторових відношень типу *напрямку* та *відстані* (з ідентифікацією відношень «попереду», «позаду», «ліворуч», «праворуч», «далеко», «недалеко» тощо та їх можливою комбінацією), а друга — презентувала алгоритм розпізнавання об’єктів (багатокутників, чи фігур з елементами сталої кривизни). На моніторі випадковим чином синтезувалася сцена за рендомною реалізацією: об’єктів, їх розмірів, орієнтації, розташування, кольору тощо; знову випадковим чином на цій сцені розміщувався автономний робот, ідентифікований своїми координатами, габаритами та напрямом зору, а також своєю схемою семантичної інтерпретації просторових відношень.

Поставлено конкретне завдання: обстежити задану сцену, знайти всі фігури, розпізнати їх, виконавши процедуру ідентифікації з визначенням усіх параметрів кожного об’єкта (лінійні та кутові характеристики, площа, центр мас, кривизна відповідних фрагментів, колір тощо), провести їх класифікацію та визначити просторове положення кожної фігури відносно робота. **Після обстеження сцени отримані дані потрібно вербалізувати, тобто подати звіт у символічному (текстовому) вигляді.**

Звіт формується у двох варіантах: або в абсолютній системі відліку просторових відношень (відносно положення робота та його напрямку зору), або у відносній системі (відносно окремо вибраного об’єкта — у цьому випадку відносно об’єкта з мінімальною площею). Тут, власне, цікавитимуть лише певні аспекти цієї процедури у зв’язку з деякими *особливостями процедури вербалізації*, що реалізується ЛП. Текст формується у вигляді послідовності речень — структур БССС.

Програмний продукт моделює основні функції зорового тракту: визначає всі об'єкти Obj, обстежує їх контур, визначає всі ознаки Attr(Obj) кожного з них (лінійні та кутові розміри, площу, центр мас), фіксує також ступінь знайдених ознак Attr (Attr) (великий, малий, жовтий, рожевий тощо); окрім того, фіксує своє просторове положення, постійно відслідковує напрям зору. Процедура синтезу тексту визначається структурою БССС з ідентифікацією ситуаційних відношень для кожної ситуації.

Звісно, це дуже спрощена модель аналізу сцени з наступною процедурою презентації отриманих результатів у вигляді послідовності структур БССС. Проте це найважливіший інформаційний аспект мовленнєвої діяльності людини, пов'язаний з вербалізацією отриманої зоровим трактом інформації. Доречно тут послатися також на відгук Г. Осіпова (голови асоціації «Штучний інтелект» РАН) на конференції «Інтелектуальні системи прийняття рішень» (2005 р.) щодо оцінки згаданого програмного продукту: ***на практиці добре відомі праці напряму текст–рисунок, а роботи напряму рисунок–текст ще не траплялися***. Це, мабуть, одна з перших спроб моделювання складної інтелектуальної функції людини у напрямі моделювання співвідношення «дійсність–текст».

Перспективи використання лінгвістичного процесора в інформаційних природно-мовних технологіях. Дійсно, проблема проектування ЛП у повному обсязі (з функціями аналізу/синтезу текстової інформації) видається досить складною, але тут вже досліджені головні особливості структурної організації повідомлення на як монопредикатному, так і поліпредикатному рівні, що притаманні всім мовам, бо окрема структура БССС постає схемою вербалізації окремої ситуації, яка, у свою чергу, є похідною від структурно-функціонального рівня нейроорганізації зорового тракту людини (Семір Зекі [4]).

Можемо поки що презентувати потужність запропонованої ідеології формування ЛП на засадах інтеграційного підходу на прикладі моделювання здатності такого ЛП стосовно можливості ідентифікації трансформування дієслова за схемами дієприкметника, дієприслівника та субстантиву. У праці Ю. Кисленка та А.В. Терентьєва [11] презентовані результати емуляції пошуку в Інтернеті конкретного запиту за умови можливості трансформування предикатора запиту від дієслівної форми до дієприкметника, дієприслівника та субстантиву. Результати експерименту підтверджують, що лише врахування можливості ідентифікувати в запиті похідні форми дієслова (що видається не надто складним) дозволяють значним чином підвищити релевантність пошуку.

Проаналізуємо отримані результати у зв'язку з моделюванням ЛП на інтеграційних засадах, головною функцією якого стає озвучена теза: ***«довільний текст презентується множиною однотипних структур БССС, організованих на моно/поліпредикатному рівні»***. Результати експерименту та методика проведення пошуку презентується на рис. 3 за результатами праці [11]. Презентуватимемо суть такого експерименту за матеріалами цієї праці.

Експеримент. Суть віртуального експерименту модифікації інформаційного пошуку полягає у такому: вибирається довільний запит і опрацьовується у п'яти провідних пошукових системах; двадцять перших

документів видачі (презумпція упорядкування за релевантністю для кожного пошукача) вибираються як тестові масиви для реалізації «пошуку за знаннями», коли під «квантом знань» визначено окрему структуру БССС. Пошук на тестових масивах виконується вручну експертом, котрий для визначення релевантності та трансформування структур користується своїми знаннями. При цьому автори використовували для віртуального пошуку лише три з усього спектра трансформацій: вхідний запит з дієслівною формою предикатора трансформувалася послідовно в субстантивну, ад'єктивну та адвербіальну форми. У тестових масивах для кожної пошукової системи вручну проводився пошук одночасно за трьома трансформованими структурами і за певним критерієм визначалася релевантність документів видачі.

Об'єктом дослідження, таким чином, постає інтернет-система, точніше — технологія пошуку інформації. Мета віртуального експерименту — порівняти ефективність існуючих схем пошуку (пошук за ключовими словами) відносно запропонованої схеми структурованого пошуку (пошук за знаннями). Тестування проводилося для п'яти провідних систем: **Google** — найпопулярнішої пошукової системи ([http // www.google.com /](http://www.google.com/)); **Yahoo** — однієї з провідних американських пошукових систем ([http // yahoo.com /](http://yahoo.com/)); **Bing** — популярної системи корпорації Microsoft ([http // www.bing.com /](http://www.bing.com/)); **Yandex** — найвідомішої російської пошукової системи ([http // www.yandex.ru /](http://www.yandex.ru/)); **Мета** — української пошукової системи ([http // meta.ua /](http://meta.ua/)).

Як запит обрано фразу: «**Письменник, який отримав Пулітцерівську премію в 2009 році**». Для пошуку за ключовими словами ця інформаційна потреба подається у вигляді послідовності: $Q_1 = (\text{письменник} \wedge \text{отримав} \wedge \text{Пулiтцерiвськy} \wedge \text{премию} \wedge \text{«2009 рік»})$, де \wedge означає логічне «І». У такому вигляді запит опрацьовується на всіх пошукових системах; при цьому для кожної з них фіксується видача перших двадцяти документів, за якими надалі вручну виконується пошук «за знаннями».

Тут маємо структуру природно-мовного повідомлення, реалізовану на множині категорій «письменник», «отримав», «премію», «Пулiтцерiвськy», «2009 рік», пов'язаних певними відношеннями R_0, R_1, r_i . Саме сукупність категорій та функцій, що їх пов'язують, однозначно визначають зміст (семантику) повідомлення. Зазначена структура вже як повнотекстова на словому рівні вибирається як запит для віртуального пошуку на отриманих вибірках з двадцяти документів для кожної пошукової системи.

Окрім того, виходячи з авторського бачення структурної організації текстової інформації, семантично адекватними запитом будуть і повідомлення, фрагменти яких подано його трансформаціями. Таким чином, «пошук за знаннями» одночасно включає в себе пошук за всіма різновидами запиту:

Q_2 — Письменник, який отримав Пулітцерівську премію в 2009 році.

Q_{21} — Письменник, отримуючий Пулітцерівську премію в 2009 році.

Q_{22} — Отримання письменником Пулітцерівської премії в 2009 році.

Q_{23} — Письменник, отримуючи Пулітцерівську премію в 2009 році.

Важливо звернути увагу на крапки в кінці кожного різновиду запиту — це означає, що релевантними будуть документи, повідомлення яких містять щонайменше зазначені різновиди структур $Q_2, Q_{21}, Q_{22}, Q_{23}$.

Для порівняння ефективності пошуку за різними стратегіями скористаємося оцінкою релевантності отриманих результатів. Для пошуку

за ключовими словами – запит – Q_1 – відповідність документа видачі інформаційній потребі користувача (його релевантність) оцінювалася таким чином. Якщо в документі видачі в одному повідомленні збіглися п'ять ключових слів, то виставлялася оцінка «п'ять», якщо «чотири», то оцінка «чотири» і т. д. до оцінки «одиниця». Отримані дані з фіксацією середніх значень у вигляді графіків для кожної пошукової системи показано на рис. 3.

У іншому випадку Q_2 — «пошук за знаннями» — порівнюються вже фрагменти структур (як сукупність категорій, пов'язаних певними відношеннями). Релевантна інформація потребує ідентифікації не тільки категорій, а і функцій, що їх пов'язують. Таким чином, якщо в документі (окремих його повідомленнях) є структура, що збігається цілком із структурою запити з п'яти елементів з урахуванням їх функціональних зв'язків, то виставляється оцінка «п'ять», якщо збігаються чотири елементи спільно з їх функціональними зв'язками, то виставляється оцінка «чотири» і т. д. Результати «пошуку за знаннями» для всіх тестованих пошуковиків (оцінка релевантності) подано на рис. 3, б.

Під час порівняння результатів відразу стає очевидним значне підвищення релевантності видачі «пошуку за знаннями» для всіх пошукових систем: Google (+0,9), Yahoo (+1,55), Bing (+1,15), Yandex (+1,0), Meta (+1,5). Ці результати отримано на малих вибірках і, ймовірно, заздалегідь були прогнозовані. **Однак слід відзначити істотні обставини:**

- 1) трансформації, використовувані для пошуку «за знаннями», цілком доступні для реалізації в пошукових системах за формальними критеріями;
- 2) вирішується важливе принципове питання — отримується гарантія того, що після проведення пошуку в подібному режимі не випускаються цікаві релевантні документи;
- 3) можна більш точно в запиті формулювати інформаційну потребу, оскільки порівняння ведеться «за знаннями».

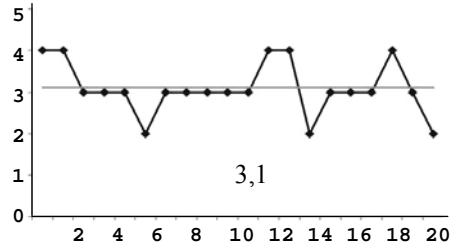
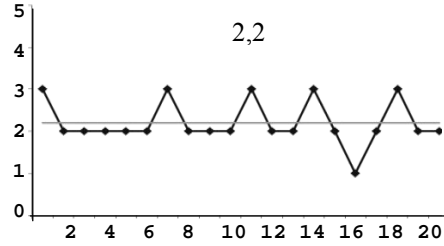
Таким чином, у результаті віртуального експерименту отримано порівняння машинного пошуку з пошуком інформації людиною. Пріоритет, звичайно ж, лишається за людиною, і такий результат можна було б передбачити заздалегідь, урахувавши, що машина:

- все ще не забезпечена знаннями про сферу пошуку;
- не має достатнього досвіду спілкування із середовищем;
- не володіє інтуїцією, передбаченням і багатьма іншими здібностями;
- не має тієї мовної компетенції, яка «спресована» в незчисленних лінгвістичних фоліантах і мало враховується в сучасних технологіях.

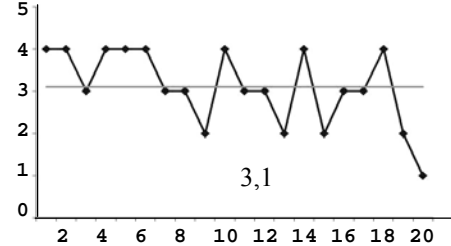
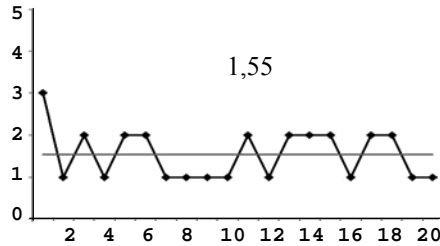
Можемо зробити певні висновки: досі маємо все ще приблизне уявлення про сам об'єкт дослідження — індивідуальну мовну систему людини, яка реалізовує мовленнєву діяльність у режимах синтезу / аналізу мовного повідомлення, і в сучасні інформаційні технології вкладається лише дуже незначна і несистемна децима знань про цю одну з найбільш складних інтелектуальних функцій людини. В експерименті враховано лише два складники мовної організації: окремі схеми трансформування повідомлення, доповнені ще й особливостями структурної організації мови, що розглядають довільний текст як множину однотипних синтаксичних структур. І перше, і друге потребують певних формалізованих процедур, що стосуються трансформацій як запити, так і текстів документів. Однак натепер за сучас-

ного розвитку інформаційних технологій — це рутинні стандартні процедури, які без особливих зусиль можуть бути перекладені на комп'ютер.

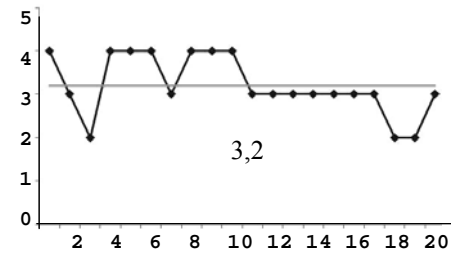
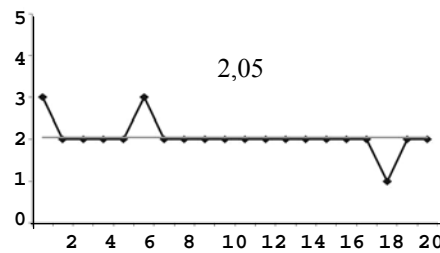
Google



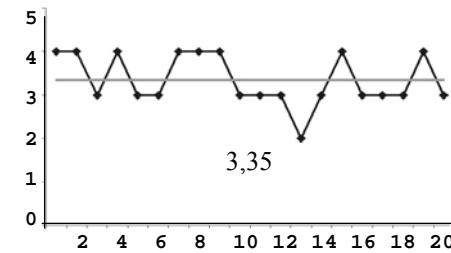
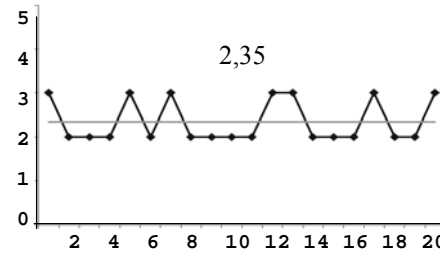
Yahoo



Bing



Yandex



Мета

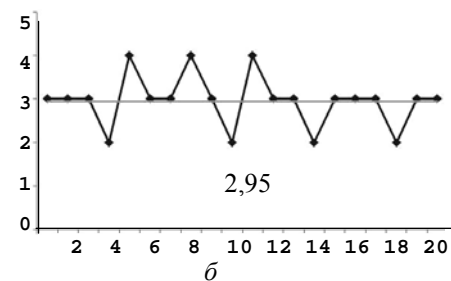
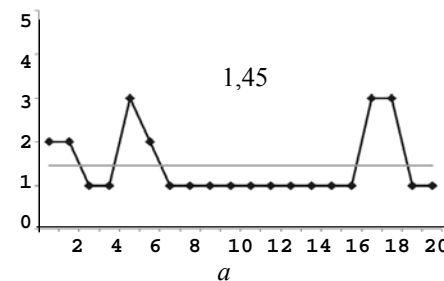


Рис. 3. Порівняльна характеристика оцінювання релевантності: *a* — оцінка релевантності за допомогою пошуку за ключовими словами; *б* — оцінка релевантності за допомогою пошуку за знаннями

ВИСНОВОК

Презентована робота присвячена аналізу індивідуальної мовної системи (як сукупності ЛП та БЗ), відповідальної у загальному випадку за актуалізацію мовленнєвої діяльності людини. Зокрема розглядається структурно-функціональний рівень організації ЛП, відповідального за структурний рівень формування повідомлення у процесі синтезу/аналізу мовного повідомлення.

Основою дослідження постає інтеграційний підхід до аналізу структурного рівня мовної організації, що охоплює результати сучасних досліджень мовленнєвої діяльності у багатьох помежованих напрямках. Такий підхід істотно відрізняється від класичної лінгвістики з огляду на формальне визначення базової семантико-синтаксичної структури як основи формування довільного повідомлення. Спочатку розглядаються особливості формування повідомлень на монопредикатному рівні, а потім і на поліпредикатному.

Заключна частина присвячена як прикладним питанням використання підходу до синтезу мовного повідомлення (система автоматичного формування опису сцени), так і можливостям його застосування для пошукових систем.

ЛІТЕРАТУРА

1. *Щерба Л.В.* О тройком аспекте языковых явлений и эксперименте в языкознании: избр. тр. / Л.В. Щерба // Языковая система и речевая деятельность. — М.: Наука, 1974. — С. 24–39.
2. *Kyslenko Yu.I.* Back to basics of speech activity / Yu.I. Kyslenko // *Biologically Inspired Cognitive Architectures* (2014) 8. — P. 47–69.
3. *Городецкий Б.Ю.* Компьютерная лингвистика: моделирование языкового общения (вступ. статья) / Б.Ю. Городецкий // Сб. НЗЛ. — Вып. 24. — М.: «Прогресс», 1989. — С. 5–31.
4. *Зеки Семир.* Зрительный образ в сознании и мозге / Семир Зеки // В мире науки. — 1992. — № 11–12. — С. 33–41.
5. *Гвоздев А.Н.* Формирование у ребенка грамматического строя русского языка / А.Н. Гвоздев. — М.: Изд-во Академии пед. наук, 1949. — 192 с.
6. *Kislenko Yuri.* Cognitive architecture of speech activity and modeling thereof / Yuri Kislenko, Danylo Sergeiev // *Biologically Inspired Cognitive Architectures* (2015) 12. — P. 134–143.
7. *Кисленко Ю.И.* Архитектура мови (Лінгвістичне забезпечення інтелектуальних інтегрованих систем) / Ю.И. Кисленко. — К.: ІЗМН, 1998. — 344 с.
8. *Кисленко Ю.И.* Структурно-функциональный уровень организации лингвистического процессора / Ю.И. Кисленко, О.С. Черевко // Искусственный интеллект. — Донецк. — 2012. — С. 143–150.
9. *Кисленко Ю.И.* Інформаційна база лінгвістичного процесора / Ю.И. Кисленко, А.В. Хіміч // Адаптивні системи автоматичного управління. — 2016. — № 1 (28). — С. 42–48.
10. *Кисленко Ю.И.* О моделировании соотношения «Действительность – Текст» / Ю.И. Кисленко, Г.А. Амплеева // Искусственный интеллект. — Донецк. — 2000. — № 3. — С. 475–487.
11. *Кисленко Ю.И.* Проблемы и перспективы развития поисковых систем / Ю.И. Кисленко, А.В. Терентьев // Искусственный интеллект. — 2011. — № 3.

Надійшла 19.11.2017