

КАТЕГОРІЯ ПРОСТОРУ В ІНФОРМАЦІЙНИХ ПРИРОДНО-МОВНИХ ТЕХНОЛОГІЯХ

Анотація: В роботі аналізуються мовні засоби відтворення просторових відношень на площині (2D) та просторі (3D) для потреб автоматичного опрацювання мовного матеріалу. Акцентується увага на структурній організації засобів подачі інформації щодо фіксації просторових відношень на площині та просторі. Ретельно аналізуються мовні засоби абсолютного та відносного характеру для відтворення простору, простежуються структурні особливості їх організації та взаємодії на шляху відтворення мовними засобами складних просторово відношень.

Ключові слова: базова семантико-синтаксична структура, пошукові системи, експертні системи, автоматичні перекладачі, природно-мовні інтерфейси, бази знань.

Вступ

Інформаційні технології, які намагаються моделювати окремі прояви мовленнєвої поведінки людини на сьогодні складають потужний кластер, до якого можуть бути прираховані пошукові системи, експертні системи, автоматичні перекладачі, природно-мовні інтерфейси, бази знань тощо.

Загальною, найважливішою (і, мабуть, найбільш проблематичною) складовою всіх технологій постає задача аналізу текстової/мовленнєвої інформації. Враховуючи сучасний стан класичної лінгвістики, яка постає основою розбудови аналізаторів всіх технологій, найбільшчим часом ми не можемо сподіватися на якісні зміни у цих розробках. Аргументовано ця позиція висвітлена в роботі [1], де ретельно проводиться зіставлення класичної лінгвістики із запропонованою моделлю системної організації мови, представленою працею [2]. Особливість підходу – в інтегруванні сучасних досягнень всіх помежованих напрямків дослідження мовленнєвої діяльності людини. Це дозволило вийти на системний рівень структурної організації мови, де головним елементом постає єдина базова семантико-синтаксична структура (БССС), яка чітко визначена на формальному рівні. Далі весь структурний рівень мовної організації представляється на монопредикатному або поліпредикатному рівнях, тож, ключовим напрямком розбудови аналізаторів постає проблема ідентифікації цієї структури. Важливими складовими БССС постають обставини часу та простору. Опрацювання цих категорій постає важливою складовою розбудови аналізаторів. В попередній публікації [3] автори проаналізували лише одновимірний варіант відтворення цих категорій. Дана ж публікація продовжує аналіз мовних засобів для ідентифікації сірконстант простору у вимірах 2D і 3D.

Мета даної статті полягає в аналізі організації мовних засобів для відтворення П-відношень на площині (2D) та у просторі (3D) для потреб їх

використання в системах аналізу текстової інформації. Класична лінгвістика в таких питаннях йде шляхом накопичення статистичних даних на шляху аналізу множини текстів. Щоб якось систематизувати мовні засоби відтворення П-відношень скористайтесь даними, накопиченими фахівцями точних наук. Такий підхід гарантує системний характер та повноту отриманих результатів.

Формування П-відношень у просторі 2D, 3D

Попередній аналіз структурної організації П/Ч-відношень поданих в матеріалах [3], практично, стосувався лише мовних засобів для фіксації певних відрізків, а саме – одновимірних характеристик часу та простору (1D), тоді як, власне, простір не має обмежень, а отже існує також у вимірах 2D та 3D. Послідовно розглянемо особливості відтворення цих властивостей у мові. Тож, переходимо спочатку до аналізу мовних засобів для відтворення просторових відношень на площині, тобто, у двовимірному просторі – 2D.

Одиниці виміру площини добре відомі – завжди можемо посилатися на стандарти СІ. Для потреб навігації, людством опрацьована система одиниць, сукупність яких покриває всю земну поверхню та будь-які окремі її фрагменти. Сучасні технічні засоби дозволяють з високим ступенем точності визначати як координати будь-якого об'єкта (довготу та широту), так і положення довільного фрагмента поверхні; система навігації GPS, наприклад, працює з технічними засобами точного визначення координат рухомих об'єктів. Як і для одновимірного варіанту, П-відношення на поверхні можуть відтворюватися або певною кількістю поіменованих конкретних одиниць виміру “квартали 44, 21, 22”, наприклад, або посиленням на порядок їх використання “у 44-му кварталі” тощо.

Але при відтворенні П-відношень на площині виникають й інші питання, що пов'язані, зокрема, з орієнтацією. Досить суттєвою вбачається проблема фіксації напряму, яка далі розпадається на три рівні, що умовно позначаються маркерами “ДЕ?”, “КУДИ?”, “ЗВІДКИ?”. Перший рівень відповідалний за орієнтацію людини на площині, інші два – будуть похідними від попереднього (дивись табл.1.).

П-відношення типу “ДЕ?” відтворюють статичні (незмінні) відношення, тоді як типи “КУДИ?” та “ЗВІДКИ?” обслуговують сферу відтворення П-відношень динамічних ситуацій. В роботі автора [2] раніше ретельно було проаналізовано використання першого типу П-відношень. Особливість підходу визначалась тим, що змістовне (семантичне) навантаження відношень типу “ДЕ?”, подавалось через образотворчі засоби їх інтерпретації [4]. Завдяки такому підходу вдалося досить ефективно моделювати процедури синтезу/аналізу П-відношення мовного повідомлення, з метою їх використання в системах управління роботами. В перспективі простежується тенденція поширення використання П-відношень за схемами “КУДИ?” та “ЗВІДКИ?”, що дозволить моделювати значно ширший спектр динамічних відношень.

Мовні засоби орієнтації у просторі 2D

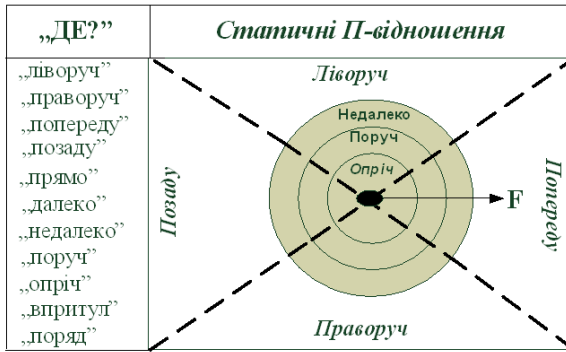
Статичні відношення		Одиничні конструкції	Комбіновані схеми
	<i>Одиниці виміру напрямку</i>		ліворуч, праворуч попереду, позаду, збоку
<i>Одиниці виміру відстані</i>		далеко, поруч, недалеко, близько	далеко попереду ліворуч, недалеко праворуч, поруч з А ...
Динамічні відношення	<i>“куди”</i>	вдалині, вліво, вправо, вперед, назад, вбік	вліво вперед, вправо вперед, вперед вліво, вперед вправо, назад вліво, назад вправо, вправо назад, вліво назад, прямо вбік ...
	<i>“звідки”</i>	здалеку, зліва, справа, збоку, спереду ...	здалеку справа, зблизька справа, здалеку зліва, зблизька зліва ...

У першому випадку орієнтація визначається відносно поточного положення самого рухомого об'єкта чи суб'єкта, тоді як у другому - місцеположення визначатиметься відносно конкретного об'єкта/суб'єкта довкілля.

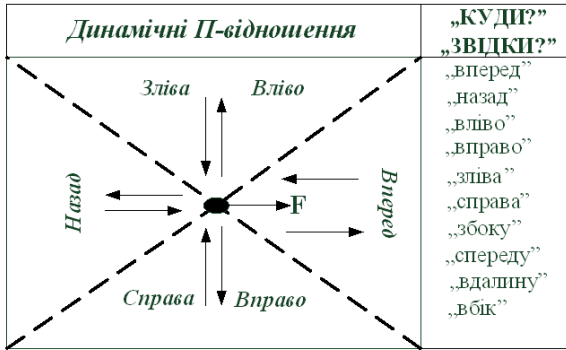
На графічному рівні можемо зафіксувати співвідношення окремих одиниць, зафіксованих таблицею 1. Лише слід враховувати при цьому, що початок координат визначається положенням суб'єкта або робота, а початковий напрям визначається або напрямком зору F або орієнтацією магнітних полюсів Землі, або просторовими характеристиками двох об'єктів, сузір'їв тощо.

Слід зауважити, що окрім лексичного складу нас цікавитиме також семантичне навантаження окремих одиниць виміру П-відношень. У роботі автора [2] певним категоріям П-відношень були зіставлені їх семантичні еквіваленти на графічному рівні, що дозволило певним чином моделювати режими синтезу та аналізу (розуміння) певного класу П-відношень. Це було використано для систем орієнтації та керування інтелектуальним роботом.

На рисунку 1.а відтворюються просторові відношення опису довкілля, що використовуються на побутовому рівні мови. Початок координат тут співпадає з положенням суб'єкта, а відлік напрямку ведеться від напрямку зору F (напрямку сприйняття довкілля). Множина одиниць виміру простору у загальному випадку повинна задовольняти двом умо-



а



б

Рис. 1 – Графічна інтерпретація П-відношень у виміру 2D:
 а. – П-відношення статичних ситуацій;
 б. – П-відношення динамічних ситуацій

вам: по-перше, покривати весь навколишній простір, а по-друге, бажано, щоб перекриття окремих одиниць П-відношень було мінімальним, інакше одиниці виміру втрачають свою диференціюючу функцію.

Рисунок 1.б демонструє змістовний рівень використання П-відношень для динамічних ситуацій, коли напрям або відстань, або і напрям, і відстань змінюються в часі. З таблиці 1 видно, що у відповідності із функціональним навантаженням П-відношень змінюється певним чином і їх лексична реалізація. Для систем сприйняття текстової інформації (аналізаторів/синтезаторів мови) така диференціація вважається суттєвою.

Рисунки 1.а та 1.б фіксують П-відношення для абсолютної системи координат. Можливий варіант використання відносної системи відліку П-відношень, коли відносини типу “ДЕ?” або “КУДИ?”, “ЗВІДКИ” формуються не відносно суб’єкта сприйняття доквілля Subj, а відносно до-

вільних складових середовища.

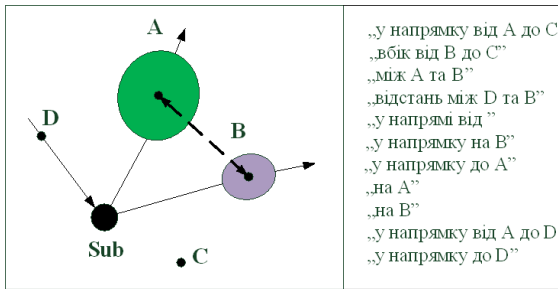


Рис. 2 – Відносна система напрямку на площині (2D)

На рис.2 представлені варіанти формування П-відношень за відносною системою відліку. Приклади реалізації мовних засобів засвідчують можливість формування П-відношень як лінійного (вимір 1D), так і просторового (вимір 2D) характеру.

П-відношення відносної системи ефективно використовуються в системах навігації рухомих засобів (повітряних, сухопутних, водних). Якщо знехтувати незначною кривизною земної поверхні, то відтворення П-відношень у вимірі 3D може здійснюватися за законами 2D. Тоді в якості початкових координат вибираються магнітні полюса Землі, відносно яких формується сітка відліку (меридіани і паралелі) лінійних та просторових характеристик систем навігації (північ, південь, захід, схід, північний захід тощо).

Стандартом відліку напрямку приймається орієнтація “північ-південь” – напрям F , всі інші – будуть похідними від нього. Наявність даної схеми також надає можливість використання для ідентифікації П-відношень таких понять (одиниць виміру простору) як “широта” та “довгота”, де за допомогою градусної сітки фіксується місцерозташування конкретного об’єкта. В даному випадку надається можливість видавати дуже точну просторову інформацію, наприклад, “географічні координати Києва -51 північної широти та 31 східної довготи”.

Якщо попередній варіант використання П-відношень на площині доповнити одиницями виміру третьої координати “зверху”, “знизу”, то значно розширимо сферу використання та моделювання П-відношень і перейдемо до виміру 3D. По аналогії з простором 2D окрім формування статичних відношень типу “ДЕ?”, тут виникає можливість використання П-відношень типу “КУДИ?” (“вгору”, “вниз”) та “ЗВІДКИ?” (“згори”, “знизу”), що обслуговують динамічні ситуації.

За допомогою даного способу подачі П-відношень, стане можливим не лише уточнити положення суб’єкта/об’єкта на площині, а і в просторі взагалі, оскільки весь світ – це об’ємна просторова модель, а орієнтація на площині являється досить спрощеною та умовною системою подачі знань,

що не завжди може задовольняти вимогам і потребам чіткого відтворення категорії простору в символній інформації.

Насамкінець лишається зауважити, що незважаючи на принципову нечіткість мовних засобів для відтворення П/Ч-відношень, існують також засоби для заперечення однозначної інтерпретації не лише якісних, а й кількісних одиниць виміру часу та простору. Реалізується така семантична трансформація шляхом використання словоформ, із значенням невпевненості, можливості, приблизності тощо.

Засоби нечіткої логіки використовуються в ситуаціях, коли неможливо однозначно вказати на якісну чи кількісну одиницю виміру часу та простору. В таких випадках використовується комбінація елементів чіткої та нечіткої логіки для відтворення просторово-часових відношень. Такий механізм формування П/Ч-відношень, зазвичай, часто використовується на побутовому рівні, оскільки окреслює лише приблизно обсяг П/Ч-відношень. Для прикладу, можемо подати наступні форми фіксації П/Ч-відношень: *“мабуть, за два кілометри звідси”, “можливо, через півгодини”, “може, на п’ятій хвилині”, “приблизно за два кілометри на південь”, “біля п’яти метрів ліворуч”,*

Висновки

Дана робота у сукупності з попередньою публікацією [3] завершує аналіз мовних засобів відтворення часу та простору. Особливість даного підходу – у послідовному використанні співвідношення “Дійсність-Текст” як основи системного підходу до вирішення проблеми структурної організації мовних засобів відтворення часу та простору. Результати досліджень закладаються в структуру аналізатора текстової інформації як складової лінгвістичного процесора для інформаційних природно-мовних технологій. Важливим моментом вважається також використання образотворчих (графічних) засобів семантики для відтворення просторових категорій в задачах аналізу/синтезу мовного повідомлення.

Література

1. Кисленко Ю.І. Мовленнева діяльність та її моделювання. - Адаптивні системи автоматичного управління, 16. – Дніпропетровськ: “Системні технології”, 2010. – с.49-57.
2. Кисленко Ю.І. Архітектура мови (Лінгвістичне забезпечення інтелектуальних інтегрованих систем). – учбовий посібник. – К.: Віпол, 1998. – 344 с.
3. Кисленко Ю.І., Черевко О.С. Категорії часу та простору в інформаційних природно-мовних технологіях. – 2010. – 14 с, подана до друку.
4. Апресян Ю.Д. Лексическая семантика. – М.: Наука, 1974.

Отримано 08.12.2011 р.