

УДК 519.876.5:004.81

### Алгебраїчна модель лінгвістичного процесора

Бісікало О.В., ВДАУ, м. Вінниця, Україна, [bisikalo@vsau.org](mailto:bisikalo@vsau.org)

#### Постановка задачі

Актуальність досліджень в області розробки лінгвістичного процесора зростає пропорційно з розширенням мережі Інтернет і глобалізацією економіки. Запропонований в [1] підхід до аналізу природно-мовних конструкцій на основі подій дозволяє моделювати появу первинних асоціативних зв'язків між образами, проте необхідно побудувати формальний апарат для накопичення та обробки в процесорі значимих параметрів окремих речень аналогічно накопиченню асоціативних зв'язків між образами при сприйнятті людиною події.

#### Формалізація

Інтерпретувавши простір асоціативних пар у вигляді булеану [2], представимо модель образного мислення за допомогою двохосновної алгебраїчної системи:

$$\text{Algebra} = \langle B; \Omega \rangle, \quad (1)$$

де  $B = \{Word, Number\}$  – (2)

основи, а  $\Omega = \{OP, RE, IF\}$  – (3)

сигнатура системи. В межах першої з основ *Word* варто розрізняти наступні види символічних послідовностей (слів):

$$\text{Word} = \{Image - Word, Link - Type, Pr onoun, Name, Author \},$$

де слова, що позначають образи *Image-Word* будемо класифікувати в залежності від їх ролі в синтагмі, а саме *Object-Quality* – якість об'єкту; *Object* – об'єкт; *Notion* – поняття; *Method* – метод; *Method-Quality* – якість методу:

$$\text{Image} - \text{Word} = \{Object - Quality, Object, Notion, Method, Method - Quality \};$$

- словами *Link-Type* будемо позначати ролі образів у реченні, що відповідає синтагмі за допомогою множини типів асоціативних зв'язків:

$$\text{Link} - \text{Type} = \{r1, r2, r3, r4, r5, r6, r7\},$$

де  $r1$  – визначення,  $r2$  – присудок,  $r3$  – підмет,  $r4$  – обставина місця,  $r5$  – обставина часу,  $r6$  – обставина,  $r7$  – додаток, – подальша вербальна деталізація основних типів асоціативних зв'язків може бути здійснена у вигляді підмножин відповідних питальних займенників *Inter-Pronoun*;

- словами *Name* та *Author* будемо позначати назви та авторів текстів, що вносяться в блок пам'яті.

В якості другої основи *Number* будемо розглядати числа таких видів:

$$\text{Number} = \{Bi, Logic, Integer, Time\},$$

де  $Bi$  –  $n$ -вимірний бінарний код вузлів булеану;

*Logic* – логічні значення 0 (False) та 1 (True):

$$\text{Logic} = \{0,1\};$$

*Integer* – цілі невід'ємні числа:

$$\text{Integer} = \{x / x \in Z^+\};$$

*Time* – дійсні невід'ємні числа:

$$\text{Time} = \{x / x \in R^+\}.$$

Для  $n$ -вимірного бінарного коду будемо позначати як  $Bi_i$  – код з  $i$  одиницями та, відповідно,  $n-i$  нулями, вважаючи, що вузол з кодом  $Bi_i$  належить до  $i$ -го шару графа булеану [2]. Формально:

$$Bi = \{Bi_j, j = \overline{1, n}\}. \quad (4)$$

В структуру лінгвістичного процесора закладемо наступну формальну ієрархію **основних понять** моделі образного мислення:

- **Синтагма** – множина бінарних кодів *Syntagma* відповідає таким вузлам булеану, які, в якості вершин дерева графу «піраміди сенсу», виражають образне представлення події і є формальним аналогом простого речення

$$Syntagma = \{x / x \in Bi_i, i = \overline{3, n}\}; \quad (5)$$

- **Текст** – множина бінарних кодів *Text* відповідає вузлам булеану, що є об'єднанням певної послідовності синтагм в таку зв'язну підмножину, де кожна з синтагм має з усіма іншими хоча б один спільний образ

$$Text = \left\{ x \in Bi_i, i = \overline{3, n} / \bigcup_j Bi - Sy_j \subseteq x \right\}, \text{ де } Bi - Sy \in Syntagma; \quad (6)$$

- **Довготермінова пам'ять** – вершина булеану з бінарним кодом *Long-memory* (глобальна одиниця) об'єднує синтагми всіх текстів в загальну модель часової послідовності (стрічки) подій

$$Long - memory := x \in Bi_n / \bigcup_j Bi - Te_j \subseteq x, \text{ де } Bi - Te \in Text; \quad (7)$$

- **Асоціативна пам'ять** – множина бінарних кодів *Assoc-memory* всіх вузлів другого шару булеану

$$Assoc - memory = \{x / x \in Bi_2\}; \quad (8)$$

- **Надоперативна пам'ять** – множина бінарних кодів *Super-memory* п вузлів першого шару булеану, кожний з яких може бути активізований в будь-який момент часу внаслідок сприйняття відповідного образу органами відчуттів

$$Super - memory = \{x / x \in Bi_1\}; \quad (9)$$

- **Фокус уваги** – вказівник на один образ з множини образів з бінарним кодом *Focus* в кожний з дискретних тактів часу

$$Focus := x / x \in Super - Memory; \quad (10)$$

- **Оперативна пам'ять** – множина бінарних кодів *Oper-memory* позначає такі вузли булеану від п'ятого шару до дев'ятого, кожний з яких є активним на даний момент часу та відрізняється тим, що до його складу обов'язково входить образ з фокусом уваги

$$Oper - memory = \{x / x \in Bi_i, i = \overline{5, 9}\}. \quad (11)$$

### Висновки

За допомогою двохосновної алгебраїчної системи (1)÷(3) в лінгвістичний процесор закладено формальну ієрархію основних понять (4)÷(11) моделі образного мислення. Таким чином показано, що вихідні природно-мовні конструкції процесора можуть відображати не тільки символічне значення, але й змістовне навантаження відомих психофізіологічних феноменів.

### Література

1. Бісікало О.В. Структура блоку пам'яті на основі моделі образного мислення людини // Искусственный интеллект – 2007. – № 3. – с. 461-468.
2. Бісікало О.В. Дослідження простору асоціативних пар в контексті бази знань електронного підручника // Вимірювальна та обчислювальна техніка в технологічних процесах – 2006. – № 2 (28). – с. 109-113.