

СИНТЕЗ ЛОГІКО-ЛІНГВІСТИЧНИХ МОДЕЛЕЙ РЕЧЕНЬ ПРИРОДНОЇ МОВИ ЯК ЗАСІБ ПОБУДОВИ ЗМІСТОВНОЇ МОДЕЛІ ТЕКСТУ

А.І. Вавіленкова, *Національний авіаційний університет*, a_vavilenkova@mail.ru

Текст вважається універсальним засобом представлення, накопичення та передачі знань, тому технології роботи з природно-мовними текстами завжди вважалися найважливішими для штучного інтелекту. Текст – це комплекс взаємопов'язаних одне з одним речень, що володіє певною автономністю по відношенню до аналогічних комплексів, змістовною цінністю, яка забезпечується єдністю комунікативного напрямку [1]. Цілісність тексту перетворює його в систему, в якій елементи залежать і передбачають один одного (речення, висловлювання, абзаци, розділи).

Для побудови змістовної моделі тексту необхідно здійснити ряд кроків.

1. Визначити тип зв'язного тексту, в межах якого формується глобальний зв'язок. Кожен тип тексту має свої мовні та когнітивні особливості, на основі чого для кожного з них можна побудувати текстову базу та ситуаційну модель [2]. Текстова база включає в себе семантичне представлення тексту та фіксує його зміст у вигляді взаємопов'язаних пропозицій. В залежності від ключових слів, що входять до текстової бази, можна виділити тексти розмовного, наукового, офіційно-ділового, публіцистичного та художнього стилів.

2. Виявити можливі типи зв'язків між складними синтаксичними частинами, тобто побудувати схематичну структуру тексту, що аналізується (у тексті виділяють три основні частини: зачин, розгортання думки і висновок). В залежності від типу зв'язного тексту таких складових може бути більше, тому для конкретної текстової бази буде свій набір складних синтаксичних частин F .

Аналіз кожної складної синтаксичної частини $f_j \in F$ полягає у:

- визначенні структури міжфразової єдності (відіграє велику конструктивну роль, визначає тематичну перспективу всього міжфразового простору);
- виявленні співвідношень складної синтаксичної частини $f_j, j = \overline{1, m}$ та множини абзаців, які входять до неї $A = \{a_1, \dots, a_k, \dots, a_q\}, k = \overline{1, q}, q$ – кількість абзаців;
- виділенні засобів міжфразового зв'язку (власне лексичні – синоніми, антоніми, лексичні повтори; лексико-граматичні – анафора, катафора; граматичні – порядок слів, сполучники, форми часу, модальні слова, обставини часу, причини, місця і т.д.);
- визначенні провідного зв'язку в складній синтаксичній частині.

3. Визначити тип кожного з абзаців $a_k \in A$. Так,

- аналітико-синтетичний абзац містить аналітичну частину в першій позиції, а узагальнюючу – у другій;
- синтетико-аналітичний абзац починається з узагальнюючої, стрижневої фрази, зміст якої розкривається в наступних реченнях;
- рамочний абзац має суміщену структуру: зачин розпочинає тему, далі – пояснююча частина, і завершується абзац узагальнюючою фразою).

4. Визначити тип зв'язку між реченнями в кожному абзаці, що є одним із найважливіших етапів проведення лінгвістичного аналізу тексту. Зв'язок між реченнями у тексті буває ланцюговим (контактним) або паралельним (дистанційним).

Ланцюговий зв'язок – думка проходить від одного речення до іншого та здійснюється зазвичай при повторі слова з попереднього речення та розгортанні її у наступному реченні. При цьому виді зв'язку наступні речення використовують задане раніше позначення предмету для поширення відомостей про нього, містять у собі наслідок, причину, спираються на вказаний раніше час та місце. Досить часто залежність

між реченнями в такому виді зв'язку настільки велика, що без попереднього речення зміст наступних не зрозумілий. Засобами ланцюгового зв'язку може бути: лексичний повтор, синоніми, вказівні слова, особові займенники, займенникові прислівники, сполучники, словесний пропуск і т.д.

Паралельний зв'язок – полягає у вживанні речень, в яких однаковий порядок слів, однотипні граматичні форми вираження членів речення. Таким чином речення при такому виді зв'язку рівноправні між собою, при цьому у них здійснюється перерахування чогось, або вони зіставляються, або порівнюються. Основним засобом реалізації паралельних зв'язків виступає синтаксичний паралелізм, тобто однакова або схожа побудова речень, що найчастіше проявляється в однаковій послідовності слів та єдності часових форм дієслів-присудків (прєдикатів).

5. Визначити засоби побудови та відновлення логічних зв'язків у кожному абзаці $a_k \in A$. Зв'язний текст представляє собою деяку закінчену послідовність речень, пов'язаних між собою за змістом в рамках задуму автора. Задача дослідження – встановити види такого зв'язку та визначити правила його передачі.

Основні способи зв'язку речень у тексті [3]: заміна іменника на займенник; спільнокореневі слова; лексично виправданий повтор; лексично виправданий повтор та вказівний займенник; заміна обставини (іменника, прислівника) прислівниками місця «там», «тут», «туди» і т.д.; антитеза – використання лексичних та контекстних антонімів, протиставних сполучників; висловлювання поточного речення замінюється на слово у наступному реченні; контекстні синоніми; наявність у наступному реченні вказівника на частину цілого, про яке говориться у поточному реченні; наступне речення – неповне (парцеляція).

6. Представити кожен текст $t \in T$ із множини електронних документів у вигляді впорядкованої четвірки:

$$t = \langle C, F, B, A \rangle,$$

де $C = \{c_1, \dots, c_i, \dots, c_n\}$ – множина існуючих типів текстів, $i = \overline{1, n}$, n – кількість типів;

$F = \{f_1, \dots, f_j, \dots, f_m\}$ – множина складних синтаксичних частин тексту, $j = \overline{1, m}$, m – кількість складних синтаксичних частин;

B – текстова база, що складається з набору ключових слів тексту та взаємопов'язаних пропозицій;

$A = \{a_1, \dots, a_k, \dots, a_q\}$ – множина абзаців тексту, $k = \overline{1, q}$, q – кількість абзаців.

Якщо об'єднати основні положення структурного аналізу, лінгвістики тексту та математичного моделювання, то можна формалізувати процес створення змістовної моделі тексту. Це можливо шляхом побудови та синтезу логіко-лінгвістичних моделей речень природної мови.

Знаючи тип зв'язку речень у кожному з абзаців, можна передбачити можливі типи зв'язків у самих реченнях, а також засоби когезії, що при цьому використовуються.

Розглянемо три абзаци науково-технічного тексту та побудуємо для кожного з них логіко-лінгвістичну модель.

Абзац 1. «Для оцінки динамічних властивостей системи прийнято досліджувати їх реакцію на типові вхідні впливи».

Логіко-лінгвістична модель заданого абзацу:

$$\text{Прийнято \& досліджувати} \left(\begin{array}{l} , \text{ реакцію} \{ix\} [\text{впливи} \{ \text{типові, вхідні} \}], \\ \text{оцінки} [\text{властивостей} \{ \text{динамічних} \}] [\text{системи}] \end{array} \right) \quad (1)$$

$$P_1 \& P_2 \langle x_2 \{c_{21}\} [x_3 \{c_{31}, c_{32}\}], x_4 [x_5 \{c_{51}\}] [x_6] \rangle, \quad (2)$$

де $P_1 \& P_2$ – прєдикат; x_2, x_3, x_4, x_5, x_6 – прєдикатні змінні – аргументи; $c_{21}, c_{31}, c_{32}, c_{51}$ – прєдикатні константи, що описують відповідні аргументи.

У абзаці 1 займенник «їх» вказує на «динамічні властивості системи», замінюючи таким чином словосполучення. Тому в логіко-лінгвістичну модель (1)-(2) можна внести наступні зміни: $c_{21} \equiv x_5\{c_{51}\}[x_6]$, після чого:

$$\text{Прийнято \& досліджувати} \left(\begin{array}{l} , \text{реакцію}[\text{властивостей}\{\text{динамічних}\}[\text{системи}]], \\ [\text{впливи}\{\text{типові, вхідні}\}], \\ \text{оцінки}[\text{властивостей}\{\text{динамічних}\}[\text{системи}]] \end{array} \right) \quad (3)$$

$$P_1 \& P_2 \left(x_2[x_5\{c_{51}\}[x_6]], [x_3\{c_{31}, c_{32}\}], x_4[x_5\{c_{51}\}[x_6]] \right) \quad (4)$$

Абзац 2. «Один із підходів, що дасть змогу уніфікувати і тим самим полегшити вирішення даної задачі та заснований на лінійності об'єкта, базується на принципі суперпозиції».

Логіко-лінгвістична модель заданого абзацу:

$$\begin{array}{l} \text{Базується} \left(\text{дин \& підходи, принципи}[\text{суперпозиції}] \right) \rightarrow \\ \text{дасть \& змогу \& уніфікувати \& полегшити} \left(\text{вирішення}[\text{задачі}\{\text{даної}\}] \right) \& \\ \text{засований} \left(\text{лінійності}[\text{об'єкта}] \right) \end{array} \quad (5)$$

$$P_1 \left(x_{11} \& x_{12}, x_3[x_4] \right) \rightarrow P'_1 \& P'_2 \& P'_3 \& P'_4 \left(x'_2[x'_3\{c_{31}\}] \right) \& P'_5 \left(x'_4[x'_5] \right) \quad (6)$$

Абзац 3. «Довільний сигнал можна розкласти на суму типових впливів та дослідити реакцію системи на кожну складову, потім, користуючись принципом суперпозиції, отримати результуючу зміну вхідної величини».

Логіко-лінгвістична модель заданого абзацу:

$$\begin{array}{l} \text{Можна \& розкласти} \left(\text{сигнал}\{\text{довільний}\}, \text{суму}[\text{впливів}\{\text{типових}\}] \right) \& \\ \text{дослідити} \left(\text{реакцію}[\text{системи}\{\text{складову}\{\text{кожну}\}}] \right) \rightarrow \\ \text{отримати} \left(\text{зміну}\{\text{результуючу}\}[\text{величини}\{\text{вихідної}\}] \right) \& \\ \text{користуватися} \left(\text{принципом}[\text{суперпозиції}] \right) \end{array} \quad (7)$$

$$P_1 \& P_2 \left(x_2\{c_{21}\}, x_3[x_4\{c_{41}\}] \right) \& P_3 \left(x_5[x_6[x_7\{c_{71}\}]] \right) \rightarrow P'_1 \left(x'_2\{c'_{21}\} \right) \& P'_2 \left(x'_4[x'_5] \right) \quad (8)$$

В другому та третьому абзаці відсутні лексичні повтори, тому ніяких заміन не відбувалося. Між трьома запропонованими абзацами простежується ланцюговий зв'язок, через це доцільно здійснити пошук елементів, що сприяють утворенню логічного поєднання абзаців у текст.

Так, у першому та третьому абзаці знайдено слова, що відіграють ключову роль у реченнях і виступають предикатними змінними – аргументами: «впливів», «типових», «системи», «принцип», «суперпозиції».

Здійснивши заміну елементів, що повторюються, а також представивши кожен абзац, як одновимірний масив, упорядкованих за індексами елементів (математично) та за семантико-синтаксичною будовою (лінгвістично), матимемо змістовну модель тексту:

$$P_{1[1]} \& P_{2[1]} \left(x_{2[1]}[x_{5[1]}\{c_{5[1]}\}][x_{6[1]}], [x_{3[1]}\{c_{3[1]}, c_{3[2]}\}], x_{4[1]}[x_{5[1]}\{c_{5[1]}\}][x_{6[1]}] \right) \quad (9)$$

$$\begin{array}{l} P_{1[2]} \left(x_{1[2]} \& x_{12[2]}, x_{3[2]}[x_{4[2]}] \right) \rightarrow P'_{1[2]} \& P'_{2[2]} \& P'_{3[2]} \& P'_{4[2]} \left(x'_{2[2]}[x'_{3[2]}\{c_{3[2]}\}] \right) \& \\ P'_{5[2]} \left(x'_{4[2]}[x'_{5[2]}] \right) \end{array} \quad (10)$$

$$\begin{array}{l} P_{1[3]} \& P_{2[3]} \left(x_{2[3]}\{c_{2[3]}\}, x_{3[3]}[x_{3[1]}\{c_{3[1]}\}] \right) \& P_{3[3]} \left(x_{5[3]}[x_{6[1]}[x_{7[3]}\{c_{7[3]}\}]] \right) \rightarrow \\ P'_{1[3]} \left(x'_{2[3]}\{c'_{2[3]}\} \right) \& P'_{2[3]} \left(x'_{3[3]}\{c'_{3[3]}\} \right) \& P'_{2[3]} \left(x'_{3[1]}[x_{4[1]}] \right) \end{array} \quad (11)$$

Логіко-лінгвістичні моделі (9)-(11) утворюють змістовну модель заданого тексту.

Список використаної літератури

1. Кронгауз М.А. Семантика – М.: Издательский центр «Академия», 2005. – 352 с.
2. Кобозева И.М. Лингвистическая семантика. – М.: Эдитореал УРСС, 2000. – 352.
3. Гальперин И.Р. Текст как объект лингвистического исследования. Изд. 5-е, стереотипное. – М.: КомКнига, 2007. – 144 с.