

УДК 510.635:004.891(045)

СТРУКТУРИЗАЦІЯ ТЕКСТОВОЇ ІНФОРМАЦІЇ ЗА ДОПОМОГОЮ ЛОГІКО-ЛІНГВІСТИЧНОГО МОДЕЛЮВАННЯ

А.І. Вавіленкова

Національний авіаційний університет

e-mail: vavilenkovaa@gmail.com

Одним із способів представлення текстової інформації у формальному вигляді є побудова логіко-лінгвістичних моделей речень природної мови та їх об'єднання для опису текстових фрагментів [1]. Представлення тексту у вигляді формальних моделей можна поділити на декілька етапів.

1. Виявлення словосполучень. На відміну від стандартної методики поділу речень природної мови на групи слів, що формуються згідно з їх місцем розташування у реченні, логіко-лінгвістичне моделювання базується на ідеї знаходження груп слів, що пов'язані між собою за змістом, тобто словосполучень.

Керуючись гіпотезою "правило на правило", згідно з якою між граматичною структурою речення та логічною формою існує повна відповідність, можливістю застосування теореми про повноту, а також враховуючи те, що кожне речення природної мови характеризується граматичною організацією: будовою та граматичним вираженням (оформленням) членів речення, маємо таку градацію:

- підмет – суб'єкт x ;
- присудок – відношення p ;
- додаток – об'єкт y або предмет z відношення;
- означення – характеристика суб'єкта g , об'єкта q або предмета відношення r ;
- обставина – характеристика відношення h .

Нехай група слів, пов'язаних між собою логічними зв'язками, буде позначатися $sp_j, j = \overline{1, m}$, де m – кількість словосполучень у реченні. За правилами української мови словосполучення можуть утворювати між собою наступні члени речення (за їх наявності):

- "означення – підмет" – $sp_j = g \cup x$;
- "присудок – додаток" – $sp_j = p \cup y$;
- "означення – додаток" – $sp_j = q \cup y$;
- "додаток – додаток" – $sp_j = y \cup z$;
- "означення – додаток" – $sp_j = r \cup z$;
- "обставина – присудок" – $sp_j = h \cup p$.

Нехай потрібно виявити словосполучення у простому реченні: *"Цікава лекція заохочує студентів до подальшого навчання"*. Спочатку відбувається ідентифікація ролей для кожного із слів речення:

- підмет – суб'єкт x – *лекція*;
- присудок – відношення p – *заохочує*;
- додаток – об'єкт y – *студентів*;
- додаток – предмет z відношення – *навчання*;
- означення – характеристика суб'єкта g – *цікава*;

- означення – характеристика об'єкта $q - 0$;
- означення – характеристика предмета відношення $r - \text{подальшого}$;
- обставина – характеристика відношення $h - 0$.

Таким чином, у заданому реченні наявні такі словосполучення:

- "означення – підмет" – $sp_j = g \cup x - \text{цікава лекція}$;
- "присудок – додаток" – $sp_j = p \cup y - \text{заохочує студентів}$;
- "додаток – додаток" – $sp_j = y \cup z - \text{студентів навчання}$;
- "означення – додаток" – $sp_j = r \cup z - \text{подальшого навчання}$.

2. Побудова логіко-лінгвістичних моделей речень природної мови. На основі отриманих на першому етапі структуризації словосполучень, які входять до речення як його складові частини і є будівельним матеріалом, створюється формальна логіко-лінгвістична модель [2], що об'єднує в собі всі члени речення природної мови.

Нехай потрібно побудувати логіко-лінгвістичну модель для простого речення: "Цікава лекція заохочує студентів до подальшого навчання". Формальна модель будується на основі виділених вище словосполучень, що підставляються в шаблон:

$$L(S) = L^S = p(x, g, y, q, z, r, h) = p(x, g, y, 0, z, r, 0),$$

$$L(S) = L^S = \text{заохочує (лекція, цікава, студентів, 0, навчання, подальшого, 0)}.$$

Залежно від типу речення та наявних у ньому сурядних та підрядних зв'язків загальна форма логіко-лінгвістичної моделі набуває окремих форм.

Наприклад, для простого речення природної мови, ускладненого однорідними присудками "Цікава лекція заохочує та спонукає студентів до подальшого навчання" логіко-лінгвістична модель матиме вигляд:

$$L(S) = L^S = p_1(x, g, y, q, z, r, h) \& p_2(x, g, y, q, z, r, h) = \\ = p_1(x, g, y, 0, z, r, 0) \& p_2(x, g, y, 0, z, r, 0),$$

$$L(S) = L^S = \text{заохочує (лекція, цікава, студентів, 0, навчання, подальшого, 0)} \& \\ \text{спонукає (лекція, цікава, студентів, 0, навчання, подальшого, 0)}.$$

Для простого речення природної мови, ускладненого однорідними означеннями до підмета "Цікава та інформативна лекція заохочує студентів до подальшого навчання" логіко-лінгвістична модель буде мати вигляд:

$$L(S) = L^S = p(x, g_1, y, q, z, r, h) \& p(x, g_2, y, q, z, r, h) = \\ = p(x, g_1, y, 0, z, r, 0) \& p(x, g_2, y, 0, z, r, 0),$$

$$L(S) = L^S = \text{заохочує (лекція, цікава, студентів, 0, навчання, подальшого, 0)} \& \\ \text{заохочує (лекція, інформативна, студентів, 0, навчання, подальшого, 0)}.$$

Для простого речення природної мови, ускладненого однорідними означеннями та додатками "Цікава лекція заохочує лінєвих і працюючих студентів та практикантів до подальшого навчання" логіко-лінгвістична модель виглядає наступним чином:

$$L(S) = L^S = p(x, g, y_1, q_1, z, r, h) \& p(x, g, y_2, q_1, z, r, h) \& \\ p(x, g, y_1, q_2, z, r, h) \& p(x, g, y_2, q_2, z, r, h) = \\ = p(x, g, y_1, q_1, z, r, 0) \& p(x, g, y_2, q_1, z, r, 0) \& \\ p(x, g, y_1, q_2, z, r, 0) \& p(x, g, y_2, q_2, z, r, 0),$$

$L(S) = L^S =$ *заохочує (лекція, цікава, студентів, лінивих, навчання, подальшого, 0) & заохочує (лекція, цікава, практикантів, лінивих, навчання, подальшого, 0) & заохочує (лекція, цікава, студентів, працюючих, навчання, подальшого, 0) & заохочує (лекція, цікава, практикантів, працюючих, навчання, подальшого, 0).*

Логіко-лінгвістична модель простого речення природної мови, ускладненого однорідними додатками *"Цікава лекція заохочує студентів до подальшого навчання та роботи"* має такий вигляд:

$$L(S) = L^S = p(x, g, y, q, z_1, r, h) \& p(x, g, y, q, z_2, r, h) = \\ = p(x, g, y, 0, z_1, r, 0) \& p(x, g, y, 0, z_2, 0, h),$$

$L(S) = L^S =$ *заохочує (лекція, цікава, студентів, 0, навчання, подальшого, 0) & заохочує (лекція, цікава, студентів, 0, роботи, 0, 0).*

Тобто формування словосполучень на першому етапі структуризації тексту дає можливість створити неподільні за змістом одиниці, атомарні предикати, як складові логіко-лінгвістичних моделей речень природної мови.

3. Створення масивів логіко-лінгвістичних моделей. Довільний текстовий фрагмент складається з пов'язаних між собою логічними, змістовними зв'язками речень природної мови. Тому формально текстовий фрагмент можна представити за допомогою одновимірною масиву логіко-лінгвістичних моделей $t = \{L^{S_1}, L^{S_2}, \dots, L^{S_\delta}, \dots, L^{S_{N(t)}}\}$, де $\delta = \overline{1, N(t)}$, $N(t)$ – загальна кількість речень у текстовому фрагменті.

Нехай необхідно формально представити такий фрагмент тексту: *"Цікава лекція заохочує лінивих і працюючих студентів та практикантів до подальшого навчання. Від її змісту, динаміки викладання та демонстраційних матеріалів залежить те, чи отримає знання студент. Тому викладачі повинні ретельно готуватися до лекцій, постійно удосконалювати свої знання та стежити за сучасними тенденціями у науці."*

Масив логіко-лінгвістичних моделей для нього буде мати вигляд:

$$t = \{L^{S_1}, L^{S_2}, \dots, L^{S_\delta}, \dots, L^{S_{N(t)}}\} = \{L^{S_1}, L^{S_2}, L^{S_3}\} = \\ \{L^{S_1} = p_1(x_1, g_1, y_{11}, q_{11}, z_1, r_1, 0) \& p_1(x_1, g_1, y_{12}, q_{11}, z_1, r_1, 0) \& \\ p_1(x_1, g_1, y_{11}, q_{12}, z_1, r_1, 0) \& p_1(x_1, g_1, y_{12}, q_{12}, z_1, r_1, 0), \\ L^{S_2} = [p_2(x_2, 0, y_{21}, q_{21}, 0, 0, 0) \& p_2(x_2, 0, y_{22}, 0, z_{21}, 0, 0) \& p_2(x_2, 0, y_{23}, q_{22}, 0, 0, 0)] \rightarrow \\ p'_2(x'_2, 0, y'_2, 0, 0, 0, 0) = [p_2(x_2, 0, y_{21}, x_1, 0, 0, 0) \& p_2(x_2, 0, y_{22}, 0, z_{21}, 0, 0) \& \\ p_2(x_2, 0, y_{23}, q_{22}, 0, 0, 0)] \rightarrow p'_2(y_{11}, 0, y'_2, 0, 0, 0, 0), \\ L^{S_3} = p_{31} \& p_{32}(x_3, 0, y_{31}, 0, 0, 0, h_{31}) \& p_{31} \& p_{33}(x_3, 0, y_{32}, q_{31}, 0, 0, h_{32}) \& \\ p_{31} \& p_{34}(x_3, 0, y_{33}, q_{32}, z_3, 0, 0) = p_{31} \& p_{32}(x_3, 0, x_1, 0, 0, 0, h_{31}) \& \\ p_{31} \& p_{33}(x_3, 0, y'_2, q_{31}, 0, 0, h_{32}) \& p_{31} \& p_{34}(x_3, 0, y_{33}, q_{32}, z_3, 0, 0)\}. \\ t = \{L^{S_1}, L^{S_2}, \dots, L^{S_\delta}, \dots, L^{S_{N(t)}}\} = \{L^{S_1}, L^{S_2}, L^{S_3}\} = \\ \{L^{S_1} = \text{заохочує (лекція, цікава, студентів, лінивих, навчання, подальшого, 0) \& заохочує} \\ \text{(лекція, цікава, практикантів, лінивих, навчання, подальшого, 0) \&} \\ \text{заохочує (лекція, цікава, студентів, працюючих, навчання, подальшого, 0) \&} \\ \text{заохочує (лекція, цікава, практикантів, працюючих, навчання, подальшого, 0).} \\ L^{S_2} = [\text{залежить (те, 0, змісту, лекції, 0, 0, 0)} \&$$

залежить (те, 0, динаміки, 0, викладання, 0, 0) &
 залежить (те, 0, матеріалів, демонстраційних, 0, 0, 0) →
 отримає (студент, 0, знання, 0, 0, 0, 0),

L^{S_3} = повинні & готуватися (викладачі, 0, лекцій, 0, 0, 0, ретельно) &
 повинні & удосконалювати (викладачі, 0, знання, свої, 0, 0, постійно) &
 повинні & стежити (викладачі, 0, тенденціями, сучасними, науці, 0, 0) }.

Геометричною інтерпретацією логічних зв'язків між реченнями природної мови у текстовому фрагменті можуть служити мережі переходів з кінцевим числом станів, у яких вершини – це речення, а дуги – відношення, за якими речення пов'язані між собою за змістом [3]. Так, для розглянутого текстового фрагменту мережа переходів з кінцевим числом станів буде мати вигляд (рис. 1):

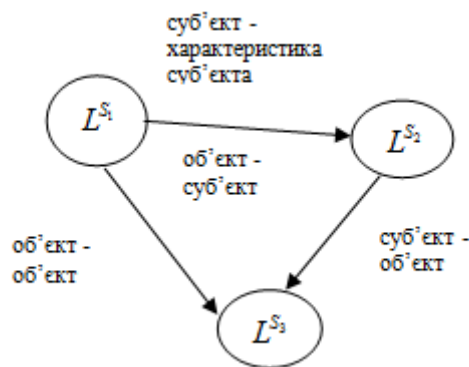


Рис. 1. Геометрична інтерпретація логічних зв'язків між реченнями природної мови у текстовому фрагменті

При побудові масиву логіко-лінгвістичних моделей застосовуються основні принципи їх синтезу [4], що дозволяються здійснити заміну складових логіко-лінгвістичних моделей речень природної мови, які пов'язані між собою за змістом.

Отже, структуризація текстової інформації шляхом логіко-лінгвістичного моделювання передбачає виявлення і пошук логічних зв'язків, починаючи з найнижчого рівня – побудови словосполучень та закінчуючи пошуком логічних зв'язків у фрагментах та текстах.

Список використаної літератури

1. Маслова В.А. Когнитивная лингвистика: учеб. пособ. / Маслова В.А. – Минск: ТетраСистемс, 2008. – 272 с.
2. Vavilenkova A.I. The formal model of knowledge retrieval and processing / A.I. Vavilenkova // Black Sea Scientific journal of academic research. – (Series «Technical and applied sciences»). – 2014. – Vol.16, N 09. – P. 115 – 119.
3. Алгоритмы: построение и анализ / [Кормен Т., Чарльз И., Лейзерсон Р. и др.]. – [3-е изд.]. – М.: Издательский дом «Вильямс», 2013. – 1328 с.
4. Вавиленкова А.И. Основные принципы синтеза логико-лингвистических моделей / А.И. Вавиленкова // Кибернетика и системный анализ. – 2015. – Т. 51, № 5. – С. 176 – 185.