

УДК 519.62

Д.В. Преснякова

Харківський національний університет радіоелектроніки, Харків

МОДЕЛЬ ПОВЕРХНЕВОГО СЕНСУ ПРИРОДНОЇ МОВИ ІЗ ВИКОРИСТАННЯМ СЕМАНТИЧНИХ ФУНКЦІЙ

У статті запропонована модель інтегрованої семантики, яка базується на формулах числення предикатів, елементами яких є семантичні функції. Так предикатні константи мають значення, що задаються семантичними функціями. Аргументами предикатів, крім змінних і констант, можуть бути семантичні функції. Предикатні константи можуть мати атрибути, за рахунок яких можливо у логічній наслідок включати такий фактор, як час. Для виведення логічних наслідків модифіковано метод резолюцій, для чого в операцію уніфікації введено доповнення: уніфікуючими вважаються еквівалентні літерали та ті, що знаходяться у розміщенні семантичних функцій.

Ключові слова: семантичні функції, природна мова, формалізація семантики.

Постановка проблеми

Проблема інтелектуального аналізу і текстів природною мовою (ПМ) з'явилася одночасно зі створенням комп'ютерів. Дослідники в області ПМ поділяють на загальну і прикладну NLP (Natural Language Processing – NLP). Завданням загальної NLP є розробка моделей обробки мови людиною, а моделі при цьому повинні бути комп'ютерно ефективними. В основу таких моделей полягає концепція загального розуміння текстів і їх сенсу, як це спостерігається в роботах Чарнака, Кулліана, Шенка, Філмора [1 – 4].

Формалізація сенсу бере початок в роботах Н. Хомського. Для відображення сенсу використовується структура безпосередньо складових (БС). Перетворення природно-мовної конструкції в структуру безпосередньо складових відбувається за допомогою породжуючої граматики. В процесі перетворення ПМ-конструкції в БС не використовується ніяка семантична інформація. Надалі з'ясувалося, що породжуюча граMATика не в змозі адекватно «мовному світу» видавати сенс аналізованих конструкцій. Однакові структури БС можуть призводити до різних сенсів. У той же час в породжуючій граматиці відображено, що різні структури можуть мати один і той же сенс. Якщо доповнити породжуючу граматику моделями і використовувати їх у трансформаціях, можна досягти більшої адекватності моделі реаліям «мовного світу».

У даній роботі запропонована модель інтегрованої семантики. Отримує подальший розвиток модель сенсу у вигляді семантичних функцій [5], суперпозиції функцій, що відображають зміст словосполучень мов з фразовою структурою. Суперпозиція функцій дозволяє композувати сенси великих ПМК з менших, допускаючи при цьому формальні перетворення конструкцій, що відображають сенс. Ця

модель повинна дозволити формальний висновок з природно-мовної конструкції, використовуючи необхідну інформацію про підсенси.

Модель сенсу у вигляді семантичних функцій повинна узгоджуватися з вже апробованими моделями, пропонуючи нові можливості і допускати формальний доказ еквівалентності сенсу мовних конструкцій.

Подання сенсу у вигляді певної формули вимагає обчислення цієї формули із використанням програмної системи (аналізатора ПМК), структура якої визначається заданою моделлю сенсу.

Модель поверхневої семантики у вигляді семантичних функцій

Модель поверхневої семантики у вигляді семантичних функцій, які можуть бути використані як інтерпретатор моделі лексичної семантики для слів, зазначених в λ -словнику та простих словосполучень природної мови: іменник + іменник у родовому відмінку, іменник + прикметник та словосполучень, що містять дієслово.

Показано, що одержання виражень формальної семантики без урахування поверхневої семантики може призводити до помилок при одержанні логічних наслідків. Інтеграція формальної та поверхневої семантики вимагає формальну мову (метамову) для опису тлумачень. Метамова для опису лексичної семантики має відповідати таким вимогам:

- мати описові засоби, еквівалентні засобам природної мови;
- дозволяти обчислення словникової статті слів і словосполучень природної мови;
- давати можливість установити еквівалентність для пари текстуально різних семантичних функцій, які вважаються еквівалентними, якщо вони мають однакові значення. Значенням семантичної

функції вважається номер словарної статті слова в λ -словнику.

Як метамова в роботі обрана функціональна мова. Функціональними константами обрані позначення V і F імя. Функція $V_i(\alpha)$ позначає лексичне значення слова α , що задається i -ю словниковою статтею в тлумачному словнику. Сама словникова стаття виражена суперпозицією функцій V і F . Функціями F можна описати сполучуваність слів у словосполученні, включаючи речення, як особливий тип словосполучення. Назвемо такі функції семантичними функціями, і описуватимемо з їхньою допомогою сполучення слів природної мови. Наприклад, семантику речення «жителі більшого міста дихають смогом» можна виразити семантичною функцією

$$\begin{aligned} &F_предложение(V_i(\text{дышать}), F_генетив \\ &(V_j(\text{житель}), F_прилагательное \\ &(V_k(\text{город}), \\ &Y_m(\text{большой}))), V_r(\text{смог})), \end{aligned} \quad (1)$$

де $F_генетив$ – функція, що відповідає словосполученню «іменник - іменник у родовому відмінку», $F_прилагательное$ – функція для словосполучення «прикметник - іменник».

Відзначимо, що кожному словосполученню відповідає свій тип семантичної функції, від якого залежить розподіл ролей для аргументів. Так, в (1), якщо перший аргумент – дієслово, що виражає дію певного типу, другим аргументом є той, хто робить дію. Слова природної мови виражають поняття, які можуть утворювати ієрархічну залежність, на верху якої перебувають «ОБ'ЄКТ, ДІЯ, АБСТРАКЦІЯ, ВІДНОШЕННЯ». Із цього погляду, словосполучення можуть так само мати трактування в більш загальних поняттях, ніж поняття словосполучення. Пояснимо цю тезу на прикладі. У роботі обрано модель семантики словосполучень природної мови, що базується на численні предикатів, у яке вбудовані семантичні функції. Так для формули (1), що відображує речення, формула числення предикатів матиме вигляд:

$$\forall x (V_i(\text{жить}) - (x, F_прилагательное(V_j(\text{город}), V_k(\text{большой}))) \supset V_m(\text{дышать}) - (x, V_r(\text{смог}))),$$

де $V_m(\text{дышать})$, $V_i(\text{жить})$ – є предикатні константи, x – змінна.

Інші аргументи предикатів представлені семантичними функціями.

Таке подання предикатних формул дозволяє в процесі логічного висновку враховувати сполучення слів у словосполученні. Сформульовано визначення семантичної функції. Позначимо через $L(\text{ПМ})$ множину конструкцій ПМ, тобто множину словосполучень, а через $L(\text{СФ})$ – множину конструкцій мови,

словника статті. Зазначимо, що $L(\text{ПМ})$ неформальна мова, а $L(\text{СФ})$ – формальна. Нехай Σ термінальний алфавіт для подання ланцюжків в $L(\text{ПМ})$, Σ_T – частина Σ , що використовується тільки для подання слів $L(\text{ПМ})$. Через $\alpha \in \Sigma^* \setminus \Sigma_T$ позначимо мовну конструкцію α , що складається із символів алфавіту Σ (включаючи порожнє слово ϵ). Тоді

$$\Sigma + \Sigma_T = \Sigma^* \setminus \{\epsilon\}.$$

N -алфавіт для представлення конструкцій мови $L(\text{СФ})$, і $N = \{V, F_и, (,), \} \cup \Sigma_T$.

Нехай $f(\alpha_1, \alpha_1, \dots, \alpha_n)$ – функція, яка відповідає словосполученню, що складається зі слів $\alpha_1, \alpha_1, \dots, \alpha_n \in \Sigma + \Sigma_T$. Значенням цієї функції є зміст, що може бути обчислений за допомогою відповідних процедур, тобто $f(\alpha_1, \alpha_1, \dots, \alpha_n) = \beta$, де $\beta \in N^+$. Мовна конструкція β представляє семантичну функцію

$$F_и(X_1, \dots, X_m), \quad (2)$$

де $X_j = V_i(\alpha_k)$ або $X_j = F_и$ ($1 \leq j \leq m$, $1 \leq i \leq n$, k – номер словникової статті в семантичному словнику); $и$ – назва функції (словосполучення); $F_и$ – семантична функція, вкладена в $F_и$; α_k – слово зі словосполучення.

Отже, можна записати

$$f(\alpha_1, \alpha_1, \dots, \alpha_n) = F_и(X_1, \dots, X_m), \quad (3)$$

де права частина рівності визначена в (2).

Наведемо приклади семантичних функцій в українській мові. Нехай розглядається словосполучення «ціна товару в упаковці». Йому відповідає семантична функція

$$F_родового\ відмінка(V_j(\text{ціна}), F_прийменник(V_r(v), V_e(\text{товару}), V_s(\text{упаковки}))).$$

Для словосполучення «ціна товару в гривнях» семантична функція має вигляд:

$$F_родового\ відмінка(F_прийменник(V_i(v), V_j(\text{ціна}), V_k(\text{гривнях}), V_e(\text{товару}))).$$

В обох прикладах індекси i, j, k, e, r, s позначають номер словникової статті в семантичному словнику.

Обрані приклади демонструють факт, що в слов'янських мовах слова й словосполучення можуть бути в $L(\text{ПМ})$ не підряд, а розірваними. Семантична ж функція представляє всі слова і словосполучення у фіксованому порядку.

Функція $f(\alpha_1, \alpha_1, \dots, \alpha_n)$ з (3) має областю свого визначення множину слів, а областю значень – множину рядків (термов), що є семантичною функцією. У свою чергу семантична функція може інтерпретуватися як функція, аргументами якої є інші функції, і яка має свою область значень.

Задаючи правила обчислення семантичних функцій, можна одержувати їхні значення. Із цього погляду рівність (3) можна переписати у вигляді

$$f(\alpha_1, \alpha_1, \dots, \alpha_n) = P^{m+1}(F_{\text{и}}, X_1, \dots, X_m), \quad (4)$$

де P^{m+1} – $(m+1)$ місцева функція з відомими правилами обчислення (обчислюються значення X_1, \dots, X_m , а потім – значення функції $F_{\text{и}}$).

Областю значень $F_{\text{и}}$ є множина семантичних функцій. Виникає питання в чому різниця між $F_{\text{и}}$ й множиною її значень, оскільки значенням семантичної функції є інша семантична функція. В ПМ значення словосполучення може відбивати конкретику цього словосполучення, а може бути узагальненим. Конкретика відбивається функцією $F_{\text{и}}$, а узагальнення – значенням $F'_{\text{и}}$.

За допомогою семантичних функцій можна виразити синтаксичну структуру словосполучення з елементами семантики. Якщо залучення λ -словника не дозволяє побудувати семантичну функцію, то конструкція вважається помилковою.

Мова для подання семантичних функцій є формальною і складається з термів – рядків, побудованих на алфавіті N . Наведемо визначення терму.

Слово $\alpha \in \Sigma^+ \text{рт}$ є термом.

Слово $V_i(\alpha)$ є термом.

Слово $F_{\text{и}}(t_1, \dots, t_n)$, де t_1, \dots, t_n – терми, є термом.

Одномісна семантична функція $V_i(\alpha)$ і багатомісна функція $F_{\text{и}}$ мають множиною визначення й множиною значень множину термів.

Множина семантичних функцій підрозділяється на типи, і кожна семантична функція ставиться до певного типу. Тип семантичної функції визначає ролі її аргументів. Для кожної конкретної мови (російська, українська, англійська) визначаються свої типи семантичних функцій, які пов'язані із граматичною структурою мови.

Визначено семантичні функції для часто зустрічальних словосполучень.

Наприклад, словосполученню $\alpha_1 \alpha_2$, де α_1 – іменник і α_2 – іменник у родовому відмінку, називаному в лінгвістиці «генетив», відповідає семантична функція

$$F_{\text{генетив}}(V(\alpha_1), V(\alpha_2)).$$

У такому словосполученні може бути більше двох слів, тому що α_1 і α_2 самі можуть бути словосполученнями. У цьому випадку семантична функція має вигляд:

$$F_{\text{генетив}}(F_{\text{имя1}}(V(\alpha_1), X_1, \dots, X_k), \\ F_{\text{имя2}}(V(\alpha_2), Y_1, \dots, Y_e),$$

де $X_1, \dots, X_k, Y_1, \dots, Y_e$ – семантичні функції.

Обробка формул потребує порівняння формул з метою встановлення еквівалентності двох семантичних функцій.

Для цього розроблено метод встановлення еквівалентності семантичних функцій, який дозволяє обчислювати еквівалентність при визначенні контексту в λ -формулах для дієслів.

Введемо позначення: СФ1 і СФ2 – семантичні функції; Д1 і Д2 – дерева для функцій СФ1 і СФ2 відповідно; ТекД1 і ТекД2 – пари оброблюваних вершин (поточних) з відповідних дерев; СемСл – семантичний словник.

Метод обчислення еквівалентності семантичних функцій

Вхід: СФ1, СФ2.

Вихід: «так», «ні».

1. Перетворення СФ1 у Д1, СФ2 у Д2.
2. Обхід Д1 і Д2 у глибину з фіксацією пари вершин у ТекД1 і ТекД2.
3. Якщо ТекД1 = ТекД2 (як терми), то виконати пункт 2.
4. Якщо ТекД1 = 'V(α)' і ТекД2 = F_1, то перетворення Д1: ТекД1 стає коренем піддерева, що позначений F_2, де F_2 – семантична функція, що є значенням V(α) у СемСл. Показки на вершини ТекД1 і ТекД2 не змінюються. Виконати пункт 3.
5. Якщо ТекД1 = F_и і ТекД2 = 'V(α)', аналогічно пункту 4.
6. Якщо ТекД1 = 'V(α)' і ТекД2 = 'V(β)', то перетворити Д1 і Д2. ТекД1 стає коренем для СФ F_1, ТекД2 – коренем піддерева для СФ F_2, де F_1 і F_2 суть значення V(α) і V(β) відповідно в СемСл. Показки в ТекД1 і ТекД2 не міняються. Виконати пункт 3.
7. Якщо ТекД1 = 'F_1' і ТекД2 = 'F_2', то обрати дерево для перетворення й змінити його відповідно до правил обчислення значення для функції типу F.

Загальна складність алгоритму має оцінку $O(m \cdot n)$. Семантична структура речення відображується семантичною функцією, яка названа пропозицією. У пропозиції однозначно представлена семантична структура словосполучення. Для побудови пропозицій запропоновано використовувати λ -словник і семантичні примітиви. Семантичні примітиви – це елементи елементарної семантики, які утворюють онтологію. На вершині цієї онтології знаходяться примітиви ОБ'ЄКТ, АБСТРАКЦІЯ, ДІЯ, ВІДНОШЕННЯ.

Такий підхід дозволяє формалізувати і використовувати інформацію із тлумачних словників. Запропонований метод встановлення еквівалентності семантичних функцій дозволяє обчислювати еквівалентність при визначенні контексту в λ -формулах для дієслів.

Представлена модель інтегрованої семантики, яка базується на формулах числення предикатів, елементами яких є семантичні функції. Так предикатні константи мають значення, що задаються семантичними функціями типу V. Аргументами предикатів, крім змінних і констант, можуть бути семантичні функції. Предикатні константи можуть

мати атрибути, за рахунок яких можливо у логічний наслідок включати такий фактор, як час. Для виведення логічних наслідків модифіковано метод резолюцій, для чого в операцію уніфікації введено доповнення: уніфікуючими вважаються літерали еквівалентні та ті, що знаходяться у розміщені семантичних функцій.

ВИСНОВКИ

У даній статті знаходить своє відображення підхід до формалізації сенсу словосполучень природної мови у вигляді семантичних функцій.

Семантичні функції пропонують формальний апарат для машинної обробки сенсів, дозволяють встановлювати еквівалентність словосполучень ПМ. Точність обчислень регламентується якістю семантичного словника.

Запропоновано обчислювальна схема для конструювання семантичних функцій, що відображають зміст конкретних словосполучень.

Подальший напрямок досліджень: відокремлення всієї множини семантичних функцій у різних природних мовах; визначення повної множини операцій встановлення еквівалентності семантичних функцій; формальні методи представлення сенсу в семантичному словнику; застосування семантичних функцій в машинному перекладі.

Список літератури

1. Chomsky N. *Three Models for Description of Language* / N. Chomsky // *IRE Trans. Informat. Theory*, 1956. – v. IT-2. – P. 113-124. Перевод: Хомський Н. Три моделі описання мови / Н. Хомський // *Кибернетический сборник*. – М.: ИЛ, 1961. – Вып. 2. – С. 237-266.
2. Жолковский А.Н. Модель <Смысл-Текст> / А.Н. Жолковский // *Энциклопедия кибернетики*. – К.: Украинская Сов. Энциклопедия, 1974. – Том 2. – С. 46-48.
3. Мельчук И.А. Опыт теории лингвистических моделей "Смысл-Текст" / И.А. Мельчук. – М.: Наука, 1974.

4. Melchuk I.A. *The Russian Language in the Meaning – Text Perspective* / I.A. Melchuk. – *Wiener Slawistischer Almanach, Sonderband, 39, Moskau-Wien*, 1995.

5. Баранов А.Н. *Англо-русский словарь по лингвистике и информатике. Т.1* / А.Н. Баранов, Д.О. Добровольский, М.Н. Михайлов, П.Б. Паришин, О.И. Романова. – М.: Помовский и партнеры, 1996.

6. Boole George. *Investigation of the Laws of Thought* / George Boole. – London, Eng., Walton, 1854.

7. Katz Jerrold J. *The philosophy of language* / Jerrold J. Katz. – New York -London, 1966. Перевод главы: Катц Дж. *Семантическая теория* / Дж. Катц // *Новое в зарубежной лингвистике*. – Прогресс, 1981. – Вып. 10.

8. Katz J. *The Structure of Semantic Theory* / J. Katz, J. Fodor // *Language*. – 1963. – Vol. 39.

9. Апресян Ю.Д. *Лексическая семантика. Синонимические средства языка* / Ю.Д. Апресян. – М.: Наука, 1974. – 367 с.

10. Апресян Ю.Д. *Лексическая семантика. Синонимические средства языка: автореф. монографии ...* // *Филол. наук: 10.02.19* / Апресян Ю.Д. – Минск: Белорусский гос. ун-т, 1983. – 37 с.

11. Минский М. *Фреймы и представление знаний* / М. Минский. – М.: Энергия, 1979. – 150 с.

12. Fillmore Ch. T. *The case for case / Ch. T. Fillmore // Universal in linguistic theory*. – New York, 1968. (Перевод: *Новое в зарубежной лингвистике*. – М.: Прогресс, 1981. – Вып. 10. – С. 369-495.)

13. Мельчук И.А. *Толково-комбинаторный словарь русского языка. Опыт семантико-синтаксического описания русской лексики* / И.А. Мельчук, А.К. Жолковский. – Wien: Wiener Slawistischer Almanach., 1984. – 992 с.

14. Halliday M. A. K. *An Introduction to functional Grammar* / M. A. K. Halliday // Edward Arnold (Publishers), 1985.

15. Ожегов С.И. *Словарь русского языка: Ок. 57 тыс. слов* / С.И. Ожегов; под ред. чл.-корр. АН СССР Н.Ю. Шведовой. – 17-е изд., стереотип. – М.: Рус.яз, 1985. – 797 с.

Надійшла до редколегії 21.07.2015

Рецензент: д-р техн. наук, проф. С.Ю. Шабанов-Кушнаренко, Харківський національний університет радіоелектроніки, Харків.

МОДЕЛЬ ПОВЕРХНОСТНОГО СМЫСЛА ЕСТЕСТВЕННОГО ЯЗЫКА С ИСПОЛЬЗОВАНИЕМ СЕМАНТИЧЕСКИХ ФУНКЦИЙ

Д.В. Преснякова

В статье предложена модель интегрированной семантики, основанная на формулах исчисления предикатов, элементами которых являются семантические функции. Так предикатные константы имеют значения, задаваемые семантическими функциями. Аргументами предикатов, кроме переменных и констант, могут быть семантические функции. Предикатные константы могут иметь атрибуты, за счет которых возможно в логическое следствие включать такой фактор, как время. Для вывода логических следствий модифицирован метод резолюций, для чего в операцию унификации введено дополнение: унифицируемых считаются литералы эквивалентные и находящиеся в размещении семантических функций.

Ключевые слова: семантические функции, естественный язык, формализация семантики.

NATURAL LANGUAGE SUPERFICIAL SENSE MODEL WITH THE USE OF SEMANTIC FUNCTIONS

D.V. Presnyakova

The paper proposes a model of integrated semantics, based on the predicate calculus formulas, whose elements are semantic functions. So predicate constants have the values semantic functions. The arguments of the predicate, except variables and constants may be semantic functions. Predicate constants can have attributes, by which it is possible to include a logical consequence of such factors as the time. Modified method of resolutions, which in the operation of unification entered addition.

Keywords: semantic functions, natural language formalization of semantics.