

ТЕОРІЯ КОМПІЛЯТОРІВ

Лекція 11

ОПИС СЕМАНТИКИ КВ-МОВ МЕТОДОМ МАТЕМАТИЧНОЇ ІНДУКЦІЇ

2020

**Повний текст лекції буде розміщений
на сайті baklaniv.at.ua**

На даний момент не існує універсального загальноприйнятого підходу для формального опису семантики. Розроблено безліч різних моделей і методів:

- граматичні моделі, засновані на додаванні розширень до граматики, що визначає цю мову (атрибутних граматики, імперативний або операційний метод, віденський метод, W-граматики);

- аплікативні моделі, в яких безпосередньо конструюється визначення функції, яку обчислює кожна програма, написана на цій мові (аксіоматичний метод, метод денотаційної семантики [6], метод функціональної семантики);
- моделі специфікацій, в яких описуються відносини між різними функціями мови і, якщо програма реалізує ці відносини, то вона коректна по відношенню до специфікації (метод специфікацій, алгебраїчний метод).

У всьому перерахованих методах опис семантики здійснюється шляхом виділення деякої типології смислів - категорій значень (семантичних категорій, примітивів), через які і за допомогою яких на деякому семантичному мовою задається більш складний зміст.

Для опису семантики контекстно-вільних мов згадуються три найбільш поширених методи: *продукційної* (дедуктивної, аксіоматичної), *денотаційної* (математичної) та *операційної* семантики, які є фактично єдиними потенційно придатними. Остання обставина пояснюється гнучкістю цих методів, можливістю статичного розширення набору семантичних категорій (математичних об'єктів, команд віртуальної машини, і ін.), з використанням яких здійснюється опис семантики.

Продукційний метод базується на обчисленні предикатів, де результат обчислення описується через взаємозв'язок змінних до і після застосування конструкцій мови. Продукційна семантика фактично побудована на основі математичної логіки, орієнтована на людину і призначена швидше для доведення правильності програм, ніж для формального визначення семантики.

Як приклади продукційних методів опису семантики можна привести аксіоматичний метод Хоара та метод індуктивних тверджень Флойда.

Денотаційний метод ґрунтується на функціональних обчисленнях, в яких вбудовані операції мови відображаються в однозначні математичні об'єкти, які потім застосовуються для опису семантики мовних конструкцій.

Зокрема, даний підхід ілюструється теорією обчислень Скотта, заснованої на семантичних доменах, де спочатку перераховуються стандартні і визначаються (кінцеві) домени, потім, на основі визначення конструкторів, задається семантика конструкцій мови у вигляді формули над доменами і конструкторами.

Зауважимо, що денотаційне опис мови використовується для породження компіляторів, проте жодного працездатного компілятора на основі цього методу до цих пір не побудовано.

В операційному методі операції мови описуються через команди деякої абстрактної машини. Сенс виконуваних оператором дій виражається в зміні стану цієї машини. Прикладами абстрактних машин є SECD-машина і категоріальна машина.

Як показала практика, опису, засновані на низкоуровневом моделюванні, практично не приносять користі, оскільки занадто громіздкі і незручні для сприйняття людиною.

Найбільшого поширення на практиці отримало операційний опис семантики контекстно-вільних мов в системі генерації синтаксичних аналізаторів YACC. На вхід програми подається контекстно-вільна граматика, з кожної продукцією якої зв'язується дію, описане на мові програмування Cі, і яке виконує щоразу, коли створюється вузол дерева граматичного розбору, відповідний цієї продукції.

Жоден з перерахованих методів визначення семантики не опинився ефективним ні для користувача, ні для розробника мови. Операційний метод досить зручний для реалізації і може бути корисний розробнику, але для користувача цей метод не має великого значення, так як породжувані їм опису містять занадто багато непотрібних йому подробиць.

Розробник навряд чи зможе керуватися функціональним і денотаційним методами, а для користувача вони, як правило, виявляються занадто складними, щоб їх можна було використовувати безпосередньо.

Користувачеві легше зрозуміти аксіоматичний метод, але при спробі скласти повне визначення мови він породжує надзвичайно складне опис, а для розробника цей метод і зовсім непридатний.

Незважаючи на деякі успіхи в описі семантики формальних мов завдання формального визначення семантики природної мови так і не вирішена через його «семантичної замкнутості». Семантично замкнутим називається мова, яка містить в собі як вираження, що відносяться до деяких позамовних об'єктів, так і вирази, що відносяться до характеристики самої мови.

Останнє призводить до виникнення в ньому суперечностей і парадоксів. Наприклад, намагаючись відповісти на питання, істинно або хибно висловлювання «Дана пропозиція помилково», ми приходимо до протиріччя.

Однак поняття істинності притаманне не об'єктивній дійсності, а воно висловлює не більше ніж одне з можливих властивостей висловлювань.

Отже, сенс «суперечливих» висловлювань може бути різний, в залежності від контексту їх застосування. Наприклад, раніше використовувалося пропозицію, яке в одній зі своїх інтерпретацій породжує антиномію. Але цей вислів не зруйнувало семантичну структуру справжнього тексту, так як було застосовано в особливому контексті і для вираження особливого сенсу.

Таким чином, проблеми опису семантики мов, що з'являються внаслідок виникнення протиріч, викликані тим, що мова виявляється настільки широким, що дозволяє визначати внутрішньо суперечливі ситуації і поняття, наприклад таке, як «безліч всіх множин».

Будь-яку формальну теорію, шляхом введення нових понять або надмірного розширення обсягу існуючих понять можна зробити суперечливою. Тому розширення понятійного базису не тільки у формальній, а й змістовної теорії повинно супроводжуватися тестом на несуперечливість.

Щоб уникнути виникнення протиріч А. Тарським запропоновано розділити мову на дві частини: об'єктивний (предметний) мова, службовець для вираження висловлювань, і метамова, призначений для опису семантики цих висловлювань. На предметному мовою говорять про ту чи іншу предметну область, а на метамові обговорюють властивості предметного мови.

Завдяки цьому поділу в мові не можуть з'являтися пропозиції, що говорять про самих себе.

На відміну від предметних мов, метамова є, як правило, деяким спеціальним чином створеним мовою, що не містить будь-якого роду двозначностей, що перешкоджають використанню його як знаряддя для створення точних висловлювань про предметному мовою. Однак метамова, як і будь-яка мова, також потребує описі.

Для визначення метамови потрібно метамова більш високого порядку, що призводить до деякої ієрархічної структури метаязикової засобів.

З цієї причини семантика формальних систем в даний час задається шляхом визначення деякого універсального метамови, який замикає ієрархію, будується на основі кінцевого безлічі первинних семантичних категорій і містить формальний апарат, за допомогою якого і через первинні категорії якого виражається більш складний зміст.

Так як метамова по своїй природі має бути більш виразною, ніж предметна мова, то при її використанні виникли труднощі, пов'язані з необхідністю динамічного розширення категоріального апарату в залежності від того, які вислови предметного мови зустрілися при аналізі.

Суть даного підходу полягає в тому, що для опису семантики одного з класів формальних мов - контекстно-вільних, застосовується *метод математичної індукції*, що полягає в семантичному замиканні формальної мови і в використанні семантичних категорій, які визначаються в міру необхідності, в процесі опису формальної граматики мови і засобами самої мови, де під семантичною категорією розуміється нетермінальних поняття граматики визначається мови.

Базою індукції в цьому випадку виступають первинні семантичні категорії, які оголошуються перед використанням, для чого служить аксіома - єдине необхідне і достатнє правило, включається в граматику всіх визначених мов. Якщо виникає задача опису семантики базових категорій, то таке завдання може бути вирішена відомими методами, наприклад шляхом безпосередньої реалізації цих категорій командами цільової обчислювальної платформи або визначення їх семантики на деякій метамові.

СИНТАКСИС І СЕМАНТИКА

Визначення 1. Формальною мовою L над алфавітом (кінцевою множиною знаків) T називається довільна підмножина множини U рядків кінцевої довжини в алфавіті T , $L \subseteq U(T)$.

Приклад 1. Задамо мову натуральних чисел, представлених в двійковій системі числення:

$$T = \{0, 1\}, U(T) = \{0, 00, 01, 10, 11, 000, 001, 010, K\}, L = \{0, 1, 10, 11, 100, 101, 110, 111, K\}, L \subset U(T).$$



Визначення 2. Множина S називається формальною семантикою формальної мови L , якщо задано відображення φ рядків мови L в елементи множини S ; тоді для рядка $\gamma \in L$ елемент $\varphi(\gamma) \in S$ буде її змістом.

Приклад 2. Визначимо семантику мови натуральних чисел L з прикладу 1 відображенням φ його рядків в елементи множини натуральних чисел $S = \{0, 1, 2, K\}$,
 $\varphi(0) = 0$, $\varphi(1) = 1$, $\varphi(10) = 2$, $\varphi(11) = 3$, ... ♦

В основі опису семантики формальних мов лежить принцип композиційного Г. Фреге, що полягає у визначенні семантики цілого через семантику частин, тобто між семантичними правилами інтерпретації та синтаксичні правила утворення виразів мови встановлюється відповідність.

Синтаксис формальних мов традиційно задається формальними граматиками, а контекстно-вільною мовою називається мова, для якої існує контекстно-вільна граматики, яка породжує цю мову.

Визначення 3. Контекстно-вільною граматикою G називається формальна система

- T, N, P, I , що задається чотирма елементами: T - кінцевим безліччю термінальних знаків; N - кінцевою множиною нетермінальних (допоміжних) знаків таких, що $T \cap N = \emptyset$; P - кінцевою множиною правил виведення (продукцій) вигляду $A \rightarrow \alpha$, де $A \in N$, α - рядок кінцевої довжини в об'єднаному алфавіті граматики, $\alpha \in U(T \cup N)$; I - початковим нетермінальним знаком (аксіомою) граматики, $I \in N$.

Приклад 3. Контекстно-вільна граматика мови L натуральних чисел в двійковій системі числення,

$$G = \langle \{0,1\}, \{I, J\}, P, I \rangle,$$

де

$$P = \{I \rightarrow 0, I \rightarrow J, J \rightarrow 1, J \rightarrow J0, J \rightarrow J1\}. \blacklozenge$$

Визначення 4. Висновком в контекстно-вільній граматиці $G = \langle T, N, P, I \rangle$ називається кінцева послідовність рядків, яка завершується рядком, що не містить нетермінальних знаків, і отримується з аксіоми граматики I шляхом послідовної заміни входжень нетермінальних знаків з лівих частин продукцій P на рядки з їх правих частин,

$$I \rightarrow \alpha_1 \rightarrow \alpha_2 \rightarrow \dots \rightarrow \alpha_{n-1} \rightarrow \gamma, \gamma \in U(T).$$

Приклад 4. Висновок довжини 4 в граматиці G з прикладу 3,

$$I \rightarrow J \rightarrow J1 \rightarrow J01 \rightarrow 101. \blacklozenge$$

Для опису семантики однієї формальної мови (предметної мови) будемо ставити відображення його рядків в рядки іншої формальної мови (семантичної мови). Принцип композиції в цьому випадку дозволяє спростити визначення семантики, тому що стає можливим використання кінцевих (конструктивних) засобів для зіставлення рядків предметної і семантичної мов.

З розглянутих прикладів видно особливості відомих методів опису семантики формальних мов:

- семантика формальної мови - це його інтерпретація за допомогою іншої мови;
- опис семантики здійснюється конструктивними (формальними) засобами;
- принцип композиції реалізується шляхом використання апарату формальних граматики;

- фіксується кінцева множина простих семантичних категорій, через які описується більш складний зміст;
- визначається тільки одна семантика (задається тільки одна інтерпретація).

СЕМАНТИЧНЕ ЗАМИКАННЯ

Метамовний шлях опису семантики формальних мов вважається неконструктивним, бо призводить до нескінченної рекурсії: для опису семантики метамови потрібно інший метамова, який, в свою чергу, також потребує описі.

Для вирішення цієї проблеми виконаємо замикання метамови на його предметну мову, суть якого полягає в тому, що, по-перше, множина первинних семантичних категорій залишається відкритою і поповнюється на етапі вирішення конкретної прикладної задачі, по-друге, використовується індуктивна граматична форма визначення більш складного сенсу через первинні (базові) семантичні категорії, а також через семантичні категорії, визначені раніше.

Для цього предметну мову будемо розглядати як двійку, що складається з предметного синтаксису, або правил побудови рядків мови, і предметної семантики, або правил інтерпретації цих рядків за допомогою іншої мови (рис. 1).



Рис. 1. Семантическое замыкание

Для опису предметної мови скористаємося метамовою. Відповідно до визначеного вище розподілом предметної мови, в метамові виділимо два види виразних засобів: виразні засоби для опису предметного синтаксису (синтаксичний мова) і виразні засоби для опису предметної семантики (семантичний мова).

Зв'язок метамови і предметної мови здійснюється через деяку систему аксіом, наприклад, що регламентують правила низькорівневого уявлення, структурування та кодування текстів.

Ототожнюючи складові частини метамови з відповідними частинами предметного мови, а семантику предметної мови будемо визначати на самій предметній мові, яка в цьому випадку грає роль своєї метамови, а то загальне, що притаманне визначення довільних предметних мов винесемо в протомову.

У підсумку отримуємо квазіприродню мову, яка складається з аксіоми, універсальної протомови і деякої множини проблемних (спеціалізованих) мов (рис. 2).

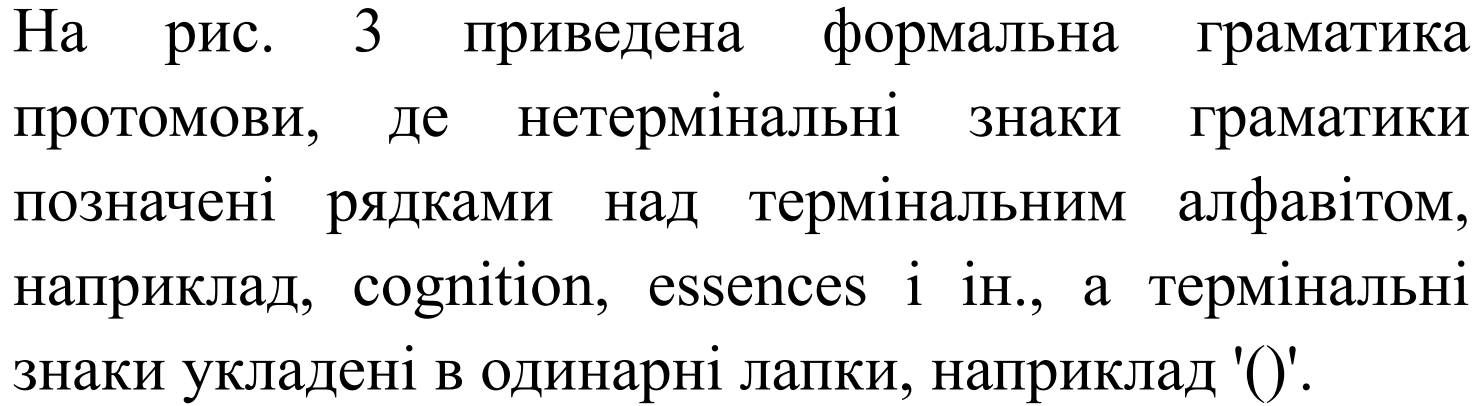


Рис. 2. квазиестественный згкое замыкание

У свою чергу одна або кілька проблемних мов, що визначаються в рамках однієї квазіприродної мови, використовується не тільки для визначення своєї семантики, але служать і для опису деякої ситуації в предметній області і її опосередкованої інтерпретації.

ПРОТОМОВА

Протомова, або первинна мова, призначена для опису синтаксису проблемних мов і реалізує аксіому, яка необхідна для оголошення базових семантичних категорій при індуктивному опису семантики проблемних мов. Таким чином, протомова є формальною мовою з фіксованими синтаксисом і семантикою.

На рис. 3 приведена формальна граматика протомови, де нетермінальні знаки граматики позначені рядками над термінальним алфавітом, наприклад, *cognition*, *essences* і ін., а термінальні знаки укладені в одинарні лапки, наприклад '()'.


cognition → essences [cognition]
essences → '()' notion '()' [intension]
intension → sentence [intension]
sentence → syntax '{ }'
syntax → item [axiom] [syntax]
item → notion | lexeme
axiom → '[' [terminals]]'
lexeme → pattern | term
pattern → ''' [terminals] '''
term → ''' [terminals] '''
terminals → terminal [terminals]

Рис. 3. Граматика протомови

Грамматика задана з точністю до прогалін і позначених курсивом нетермінальних знаків *potion*, службовців для вираження визначаються нетермінальних знаків проблемних мов. У квадратних дужках вказані частини продукцій, які можуть бути опущені, а альтернативні праві частини правил граматики розділені вертикальною лінією.

Опис cognition квазінатуральної мови складається з описів сутностей його предметної області essences. Сутностям присвоюється ім'я notion нетермінального поняття, яке визначається проблемною мовою.

Зміст intension нетермінального поняття notion складається з пропозицій sentence, які служать для вираження синтаксису syntax пропозицій, що виражають це поняття в тексті.

Синтаксис пропозиції `syntax` виражається послідовністю елементів `item`: понять `notion` і лексем `lexeme`. Лексема є термінальним поняттям визначається мови і задається у вигляді послідовністю термінальних знаків `terminals`. Для вираження лексем використовуються як терм `term`, так і безлічі термів, що задаються на мові регулярних виразів у вигляді шаблонів `pattern`.

Після будь-якого елемента `item` може бути зазначено застосування аксіоми `axiom`, яке виражається текстом, можливо порожнім, укладеними в квадратні дужки. Парні круглі дужки використовуються для виділення визначених нетермінальних знаків `token` проблемного мови, а фігурні - як роздільники пропозицій `sentence`.

ПРОБЛЕМНІ МОВИ

Проблемна мова - це контекстно-вільна мова, призначена для опису своєї семантики, предметної області та розв'язуваних у ній завдань. Для опису семантики пропозицій *sentence* проблемної мови будемо використовувати конструкцію *semantic*, яку визначимо множиною прагматик *pragmatic* з іменами *aspect* і текстом *text* на вже певній до цього проблемній підмові (рис. 5).

sentence → syntax semantic
semantics → pragmatic [semantic]
pragmatic → [aspect] '{' [text] '}'
text → terminals | [aspect] '{' [text] '}'

Рис. 5. Правила опису семантики

Таким чином, для кожної пропозиції є можливість опису однієї або декількох іменованих семантик. Для вказівки на ту чи іншу семантичну інтерпретацію деякого фрагмента тексту *text* використовуються фігурні дужки, перед якими вказується ім'я відповідної прагматики *aspect*.

ІНТЕРПРЕТАЦІЯ

Для опису семантики квазінатуральної мови було використано відображення його рядків в рядки цієї ж мови. Однак не вирішеною залишилася завдання інтерпретації квазінатуральної мови в об'єкти реального світу, яка можлива як за допомогою розумового апарату людини, так і за допомогою технічного пристрою (машини).

На рис. 7 показаний механізм машинної інтерпретація рядків квазінатуральної мови. В цьому випадку система команд цільової машини реалізує безліч первинних семантичних категорій, а сама інтерпретація здійснюється обчислювальним механізмом машини, який виконує послідовності команд, що становлять імперативи розпізнаних в тексті пропозицій квазінатуральної мови.

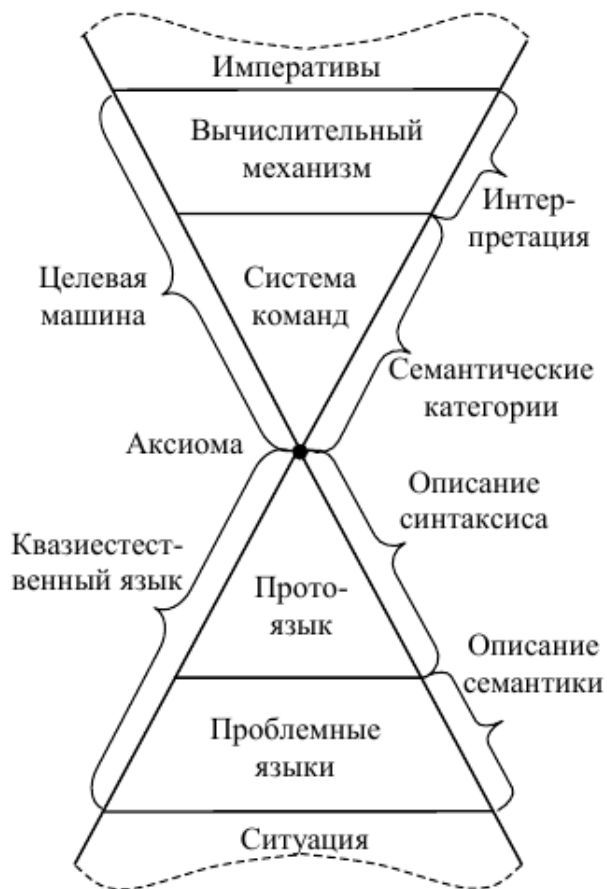


Рис. 7. Машинная интерпретация

Для створення таких імперативів використовується аксіома. Для запису в імператив пропозиції деякої команди застосовуються «порожні» квадратні дужки. «Порожні» квадратні дужки, будучи наведені після деякого елемента пропозиції, викликають запис значення цього елемента в тіло імперативу.

Приклад 10. Опишемо семантику мови числення висловів для його машинної інтерпретації (рис. 8). Перші дві пропозиції висловлюють «пусте» поняття і ставлять виразні засоби для формування імперативів. Перше речення реалізує запис в тіло імперативу байта, що задається двома шістнадцятковими цифрами.

```

() ()
  '# "[0-9A-F][0-9A-F]" [] {}'
  '< ".+" [] '>' {}
() Boolean ()
  'false'
    { #B8 #00 #00 #00 #00 #50 }
    asm { <mov eax, 0; push eax;> }
  'true'
    { #B8 #FF #FF #FF #FF #50 }
    asm { <mov eax, -1; push eax;> }
  '(' Boolean ')' {}
  'not' Boolean
    { #58 #F7 #D0 #50 }
    asm { <pop eax; not eax; push eax;> }
  Boolean 'and' Boolean
    { #58 #5A #0B #C2 #50 }
    asm { <pop eax; pop edx;
          and eax, edx; push eax;> }
  Boolean [a] 'or' Boolean [b]
    { not (not a and not b) }

```

Рис. 8. Мова числення висловлювань

Друга пропозиція здійснює запис в тіло імперативу мнемонічного позначення команди цільової машини або рядки цільової мови програмування. В останньому випадку мається на увазі, що обчислювальний механізм перед виконанням імперативу викликає інтерпретатор (асемблер) або цільову систему програмування для отримання відповідної послідовності команд для виконання.

У прикладі визначено дві семантики. Неіменованого семантика може виконуватися на «голому» процесорі, а семантика з ім'ям `asm` вимагає використання обчислювального механізму з попередніми асемблюванні імперативів. Слід звернути увагу на останнє речення аналізованого мови. Для визначення семантики цієї пропозиції використані виразні засоби описаного до цього под'язика обчислення висловлювань.

У підсумку отримуємо, що текст, що описує семантику пропозицій квазінатуральної мови, перетворюється (компілюється) в імператив - деяку послідовність дій, яку цільова машина виконує щоразу, коли при граматичному розборі вхідного тексту розпізнається ця пропозицію. За своєю суттю імперативи - одиниці виклику цільової машини, в той час як синтаксис пропозицій - опис структури таких викликів.

СЕМАНТИЧНА ІНДУКЦІЯ

Суть запропонованого підходу до опису семантики контекстно-вільних мов полягає в використанні методу математичної індукції.

Базою індукції в цьому випадку виступають первинні семантичні категорії, які безпосередньо реалізуються цільовою обчислювальною платформою і оголошуються перед використанням, для чого служить аксіома - єдине необхідне і достатнє правило, включається в граматику всіх визначених мов.

Припущенням індукції служать все раніше певні семантичні категорії (всі раніше визначені нетермінальні поняття мови).

Сукупність правил виведення граматики, що виражає один і той же нетермінальних поняття, використовується для позначення в тексті семантичної категорії, пов'язаної з цим поняттям. Індуктивний перехід здійснюється в процесі додавання нового правила в граматику визначається мови і опису семантики цього правила текстом, що виражається на вже певному до цього підмові квазінатуральної мови.

Висновком індукції є визначення нової або уточнення (доопределение) вже існуючої семантичної категорії.

Таким чином, в процесі опису квазінатуральної мови створюється і семантична мова, необхідна для опису його семантики.