**Ще кілька визначень**

У попередніх прикладах ми вже користувалися такими термінами, як «виведення», «мова, яка породжується граматикою». Тепер дамо формальні визначення для цих та інших понять.

Нехай є граматика G = (Т, N, P, S) і ланцюжок α1 складена з терміналів і нетерміналів граматики G, причому α1 представима у вигляді

**α1 = γ1αγ2**

де **γ1**,**γ2**-ланцюжки терміналів і нетерміналів граматики G, **α**-непустий ланцюжок терміналів та нетерміналів граматики G: **γ1,γ2**∈ (T∪N)\*; **α** ∈ (T∪N)+.

Нехай також серед множини правил Р граматики G є правило **α→β**. Тоді підчіпка **α** ланцюжка **α1** може бути замінена ланцюжком **β**, в результаті чого буде отримано ланцюжок **α2 = γ1βγ2**

У цьому випадку кажуть, що ланцюжок **α2** безпосередньо виводиться (породжується) з ланцюжка **α1** у граматиці G, що записується так:

**α1** $⇒$ **α2**, или просто **α1** $⇒$ **α2**,

якщо використовувана граматика очевидна.

Якщо є послідовність ланцюжків **α1**, **α2**,..., **αn** (n > 1),таких що

**α1** $⇒$ **α2**, … $⇒$ **αn**, (\*)

то кажуть, що ланцюжок **αn** нетривіально виводиться з **α1** в граматиці G (виводиться за один або більше кроків), що позначається так:

**α1** $⇒$ **αn** или просто **α1** $⇒$ **αn**.

Послідовність ланцюжків **α1, α2,...,αn** у разі називається виведенням ланцюжка **αn** з **α1** в граматиці G.

Надалі виеденням називатимемо також запис, подібний (\*). Використовується також запис

**α1** $⇒$ **αn** или **α1** $⇒$ **αn**.

що означає, що ланцюжок **αn** виводиться з **α1** в граматиці G (виводиться за нуль або більше кроків), що слід розуміти так, що **αn** або збігається з **α1**, або

**α1** $⇒$ **αn**

***Сентенційною формою*** граматики G називається ланцюжок, що виводиться з початкового нетерміналу граматики G.

Ланцюжок **α** є сентенційною формою граматики G, якщо

 S $⇒$ **α**.

 **G**

***Сентенцією*** (від sentence-пропозиція) граматики G називається сентенційна форма, що складається тільки з термінальних символів. Мова, що породжується граматикою, є множина її сентенцій. Можна сказати, що сентенції граматики — це речення мови, яку вона породжує.

**Дерево виведення**

У подальшому розгляді передбачається, що ми маємо справу з граматиками, всі правила яких у своїй лівій частині містять єдиний нетермінал. Саме такі граматики, які називаються контекстно-вільними, становлять для нас найбільший практичний інтерес. Розглянемо граматику

G6: S→AB (1)

 А→аА|а (2)

 B→bB|b (3)

Виконаємо виведення ланцюжків у цій граматиці. Спочатку використовуємо правило (1):

S⇒AB

Супроводимо процес виведення побудовою дерева (рис. 1). Коренем дерева буде вершина, що відповідає початковому нетерміналу S. Дочірніми вершинами кореня будуть вершини А і B, що відповідають правій частині першого застосованого правила (рис.1, а). Вершини А і В йдуть у дереві зліва направо у тому порядку, що у правилі (1): зліва — A, праворуч — В.



Рис.1 Побудова дерева виведення.

Продовжуючи висновок, ми можемо вибрати як правило для нетерміналу А – правило (2), так і правило для В – правило (3). Використовуємо спочатку першу частину правила (2) (А → аА):

S⇒AB⇒aAB

і продовжимо побудову дерева (рис. 1, б). Тепер виконаємо підстановку замість нетерміналу В ланцюжка bВ за правилом (3):

S ⇒ АВ ⇒ аАВ ⇒ аАbВ ,

додавши до наявної вершини В дочірні вершини b і B (рис. 1, в). Ще раз застосуємо правило А → аА і щоб отримати ланцюжок, що складається тільки з терміналів, виконаємо підстановки за правилами A →а і В → b. Після цього висновок набуде наступного вигляду:

S⇒АВ⇒аАВ⇒аАbВ⇒ааАbВ⇒aaabB⇒ aaabb,

а дерево, що вийшло, показано на

рис. 1, г. Дерево, що вийшло, називається деревом виведення, деревом розбору або синтаксичним деревом. Його корінь – початковий символ граматики, внутрішні вершини – нетермінали, листя дерева (кінцеві вершини) – термінали.

Обхід листя дерева ліворуч дає ланцюжок терміналів, виведений з початкового символу граматики (сентенцію). Неважко помітити, що збудоване нами дерево відповідатиме й іншим виведенням ланцюжка aaabb у граматиці G6. Ось один із них:

S⇒АВ⇒АbВ⇒аАbВ⇒aAbb⇒aaAbb⇒ aaabb.

Розгляд дерева замість виведення дозволяє ігнорувати порядок застосування правил, якщо він не є важливим.

**Завдання розбору**

Завдання аналізу полягає у відновленні дерева виведення для заданої сентенції. Розбір - це побудова виведення для заздалегідь заданого ланцюжка. Іншими словами, розбір - це то саме виведення, простежене у зворотному порядку. Послідовність сентенційних форм, що призводить до ланцюжка терміналів (сентенції, речення мови, що породжується граматикою), визначає структуру цього ланцюжка. Дерево виведення представляє структуру ланцюжка наочніше незалежно від послідовності застосування правил. Результатом розв'язання задачі розбору у разі, якщо вдалося відновити дерево для заданого термінального ланцюжка, є виявлення структури цього ланцюжка.

**Побудоване дерево називається деревом розбору.**

Успішне відновлення дерева розбору для заданого ланцюжка означає, що цей ланцюжок є правильним реченням мови, що породжується граматикою.

Навпаки, якщо для деякого ланцюжка терміналів дерево розбору в даній граматиці побудувати неможливо, це означає, що ланцюжок не належить мові, що породжується граматикою.

Для чого треба вирішувати завдання розбору

Розбір (англійською - parsing) називають також розпізнаванням або синтаксичним аналізом. Синтаксичний аналіз має дві мети - з'ясування належності ланцюжка мови та виявлення її структури. Робота будь-якого транслятора заснована на розпізнаванні структури речень мови, що транслюється.

Синтаксичний аналіз - обов'язкова фаза у роботі компіляторів та інтерпретаторів мов програмування.

***Синтаксичний аналізатор*** - це частина транслятора, що становить його основу. Тільки розпізнаючи структуру вхідної програми, визначаючи наявність або відсутність окремих її частин — описів, операторів, виразів— транслятор може виконати роботу з перекладу програми іншою мовою.

Часто транслятори у явному вигляді будують дерево програми, яке представляється внутрішніми динамічними структурами даних транслятора, а потім використовується для формування еквівалентної вихідної програми.

Якщо під час розпізнавання дерево виведення не будується, воно є неявно, відбиваючись у послідовності виконуваних синтаксичним аналізатором дій. Розгляду способів розв'язання задачі аналізу — синтаксичного аналізу буде присвячено більшість наших занять.

**Доміно Де Ремера**

Де Ремер (De Remer F. L.) запропонував наочну інтерпретацію завдання розбору, представивши її як гру у своєрідне доміно.

Граючий має в своєму розпорядженні «кістки» доміно декількох типів. Типів стільки, скільки правил у граматиці. Кожне правило дає один тип платівки. Типи доміно для граматики G6 показано на

рис. 2. Вважається, що «кістяшок» кожного типу є скільки потрібно.



Рис.2 Доміно Де Ромера для граматики G6

Верхня частина кожного доміно відповідає лівій частині правила граматики, нижня – правій. Верхня та нижні пластинки з'єднані «гумовими» нитками. Пластинки можна приставляти одна до одної плоскими сторонами півкола, якщо на них записані однакові символи. Фігури доміно не можна перевертати, і не можна змінювати порядок проходження символів (перехрещувати нитки).

На початку гри у верхній частині поля міститься півколо, звернене опуклістю вгору, в якому записаний початковий нетермінал граматики. У нижній частині ігрового поля в півколах, звернених плоскою частиною вгору, розміщуються термінальні символи ланцюжка, що розпізнається. На рис. 3 показано початкова конфігурація гри для ланцюжка aaabb.

Рис.3 Початок гри

Мета полягає в тому, щоб з'єднати за допомогою наявних фігур символи термінального ланцюжка та початковий нетермінал. Отримана конфігурація доміно для ланцюжка aaabb та граматики G6 (набір доміно на рис. 2) показано на рис. 4.

 Рис.4 Дерево розбору, побудоване за допомогою доміно Де Ремера