Лекція №2

Транслятори і формальні граматики

Основні визначення

У прикладах для нижчеприведених визначень будемо використовувати граматику для цілого десяткового числа:

1. <число> → <рядок цифр>
2. <рядок цифр> → <цифра>
3. <рядок цифр> → <рядок цифр><цифра>
4. <цифра> → 0
5. <цифра> → 1
6. <цифра> → 2
7. <цифра> → 3
8. <цифра> → 4
9. <цифра> → 5
10. <цифра> → 6
11. <цифра> → 7
12. <цифра> → 8
13. <цифра> → 9

Визначення 1. Рядок $\boldsymbol{\alpha}$ **безпосередньо породжує** рядок $\boldsymbol{\beta}$ $(\boldsymbol{\alpha \Rightarrow \beta})$,
якщо $\boldsymbol{\alpha = \gamma_1 <U> \gamma_2}$, $\boldsymbol{\beta = \gamma_1 \delta \gamma_2}$,
та існує правило $\boldsymbol{<U> \rightarrow \delta}$,
де $\boldsymbol{\alpha, \beta, \delta}, \gamma_1, \gamma_2 \in \mathbf{V}^*$, $\mathbf{<U> \in N}$,
або, можна сказати інакше:
рядок $\boldsymbol{\beta}$ **безпосередньо виводиться** з рядка $\boldsymbol{\alpha}$.

Приклад.
Побудувати число 22 за вищенаведеною граматикою.
Процес побудови:

№ правила	α	Дія	β	Лівий контекст γ_1	Правий контекст γ_2
1	<число>	⇒	< рядок цифр>	ε	ε
3	< рядок цифр >	⇒	< рядок цифр><цифра>	ε	ε
2	< рядок цифр><цифра>	⇒	<цифра><цифра>	ε	<цифра>
6	<цифра><цифра>	⇒	2<цифра>	ε	<цифра>
6	2<цифра>	⇒	22	2	ε

Результуючий ланцюжок виводу матиме вигляд:
<число> $\Rightarrow^1$ <рядок цифр> $\Rightarrow^3$ <рядок цифр><цифра> $\Rightarrow^2$ <цифра><цифра> $\Rightarrow^6$ 2<цифра> $\Rightarrow^6$ 22.

Визначення 2. Кажуть, що рядок α **породжує** рядок β, якщо існує ланцюжок виводів:

$$\alpha = \gamma_0 \Rightarrow \gamma_1 \Rightarrow \gamma_2 \Rightarrow \ldots\ldots \Rightarrow \gamma_n = \beta.$$

Породження позначається $\alpha \Rightarrow^+ \beta$, якщо β виводиться з α більш ніж за один крок.
Породження позначається $\alpha \Rightarrow^* \beta$, якщо може бути одна з двох ситуацій: $\alpha \Rightarrow^+ \beta$ або $\alpha \Rightarrow \beta$.

Наприклад:

1) <число> $\Rightarrow^*$ 22
2) <число><цифра> $\Rightarrow^*$ 22
3) <число><цифра> $\Rightarrow^+$ 22

Для заданої граматики записи 2) і 3) є однаково коректними.

Визначення 3. Сентенцією або реченням граматики G називається рядок, що складається лише з термінальних символів і виводиться з аксіоми граматики.

Визначення 4. **Сентенціальною формою граматики G** з аксіомою S називається будь-який рядок термінальних та/або нетермінальних символів, що виводиться з аксіоми граматики.

Визначення 5. Мовою L граматики G називається множина всіх сентенцій (речень), які можуть бути породжені граматикою G.

Класифікація мов за Хомським.

Ноам Хомський (Аврам Ноум Хомски (Чомски)) (Avram Noam Chomsky) народився у 1928 році у Філадельфії штат Пенсільванія, у єврейській родині. Його батьки — відомий гебраїст, професор Уільям Хомський (William Chomsky, 1896—1977, народився у містечку Купель Волинської губернії) і Елсі Симоновська (народилась у Бобруйську). Фраза з інтерв'ю Хомського (Noam Chomsky): My father came from the Ukraine... — «Мій батько походить з України».

Класифікація формальних мов була вперше викладена в книзі «Синтаксичні структури» у 1957 році. Згідно з цією класифікацією існує чотири типи формальних граматик і, відповідно, формальних мов.

- **Граматика типу 0** має правила вигляду

  $$\alpha \to \beta,$$

  де $\alpha \in V^+$, $\beta \in V^*$.
  До даного типу відносяться всі природні мови.

- **Граматика типу 1** має правила виду

  $$\alpha\langle u\rangle\beta \to \alpha\,\gamma\,\beta,$$

  де $\alpha, \gamma, \beta \in V^*$, $\langle u\rangle \in N$, α – лівий контекст, β – правий контекст.

  Мови, що породжуються граматиками типу 1, називаються **контекстно-залежними мовами** (КЗ-мовами) або **мовами безпосередніх складових** (БС-мовами).

- **Граматика типу 2** має правила вигляду

$$\langle u \rangle \to \alpha,$$

  де $\langle u \rangle \in N$, $\alpha \in V^*$.
  Такі граматики називаються **контекстно-вільними** граматиками, що породжують **контекстно-вільні мови** (КВ-мови).
  До КВ-мов відносяться практично всі мови програмування. Граматики типу 2 використовується для побудови синтаксичних аналізаторів.

- **Граматика типу 3** має правила тільки одного з двох наступних видів:

  перший: $\langle u \rangle \to a$
  $\langle u \rangle \to a \langle v \rangle$
  або
  другий: $\langle u \rangle \to a$
  $\langle u \rangle \to \langle v \rangle a$

  де $a \in T$,$\langle u \rangle \in N$, $\langle v \rangle \in N$.

Граматики типу 3 також називають **регулярними** або **автоматними** граматиками. Мови, які породжуються такими граматиками, називаються **регулярними мовами**.

Визначення деяких властивостей граматик

1. Еквівалентність граматик.
Граматики називаються еквівалентними, якщо вони породжують одну і ту ж саму мову.

Наведені нижче граматики 1 та 2 є еквівалентними:

Граматика 1:

1. <рядок> → <рядок символів>
2. <рядок символів> → a
3. <рядок символів> → <рядок символів>a

Граматика 2:

1. <рядок> → <рядок символів>
2. <рядок символів> → <символ>
3. <рядок символів> → <рядок символів><символ>
4. <символ> → a

2. Однозначність граматики.
Граматика називається однозначною, якщо для будь-якого речення, породжуваного цією граматикою, всі можливі схеми його виводу приводять до одного і того ж дерева виводу.

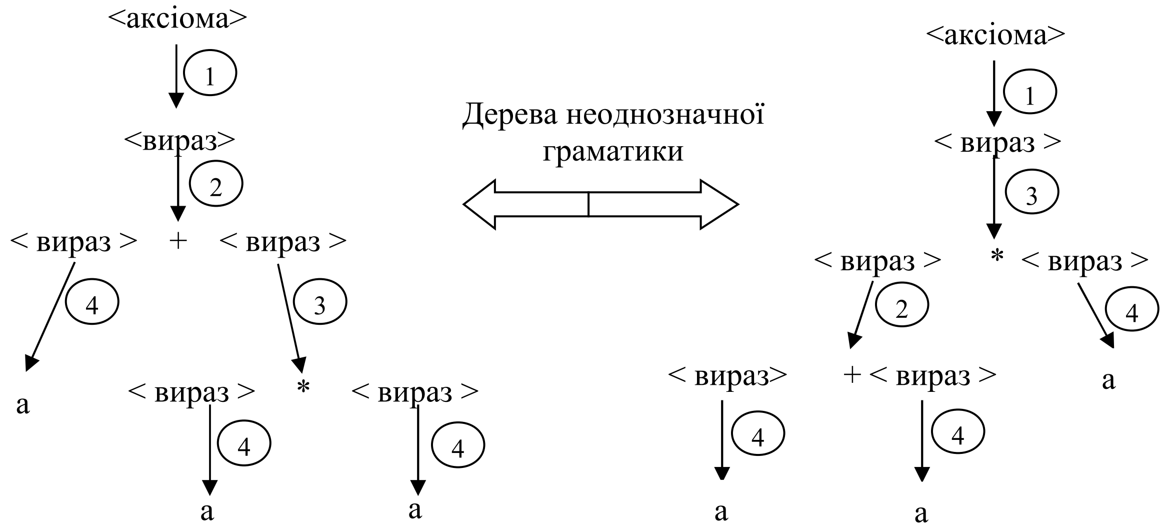
3. Неоднозначність граматики.
Граматика називається неоднозначною, якщо для одного і того ж речення, породжуваного граматикою, існує декілька неспівпадаючих дерев виводу.

Наприклад, для граматики

1. <аксіома> → <вираз>
2. <вираз> → <вираз>+<вираз>
3. <вираз> → <вираз>*<вираз>
4. <вираз> → a

можливі наступні різні дерева виводу:

a+a*a



Якщо цю граматику переписати інакше, то отримаємо однозначну граматику вигляду

1. <аксіома> → <вираз>
2. <вираз> → <терм>
3. <вираз> → <вираз>+<терм>
4. <терм> → <множник>
5. <терм> → <терм>*<множник>
6. <множник> → a

Примітка. *В термінології формальних граматик під терміном «терм» розуміють вираз, що складається лише з множників.*

Деякі властивості КВ-граматик

1. *Будь-яку ε-вільну КВ-граматику можна привести до* ***нормальної форми Грейбах*** *з правилами вигляду*:

<x> $\rightarrow b\alpha$,

де $b\in T$, $\alpha\in N^*$, *тобто* α *- рядок нетерміналів, можливо порожній.*

2. *Будь-яку ε-вільну КВ-граматику можна привести до* ***нормальної форми Хомського*** *з правилами вигляду:*

1. *<x> → <A><B>,*
2. *<x> → b,*

де <x>,<A>,<B>∈N; b∈T.

Приклад. Привести до нормальної форми Хомського граматику

1. <U> → <A><b><Z><Y>
2. <b> → b

Введемо правило: <U1> → <A><b><Z>,
тоді: <U> → <U1><Y>.
Введемо правило: <U2> → <A><b>,
тоді: <U1> → <U2><Z>.

Тобто замість граматики

1. <U> → <A><b><Z><Y>
2. <b> → b

отримаємо граматику у нормальній формі Хомського:

1. <U> → <U1><Y>
2. <U1> → <U2><Z>
3. <U2> → <A><b>
4. <b> → b.

3. *Для будь-якої КВ-мови L з граматикою G існує ε-вільна КВ-граматика G', така що*

$$L(G') = L(G) \setminus \{ \varepsilon \}.$$