1. Транслятор – это программа или техническое устройство, выполняющее трансляцию.

Токен – это пара, состоящая из класса токена и лексемы.

Класс токена –это классификация лексем в соответствии с их ролью в программе.

Лексема – это последовательность символов, образующих слово, которое соответствует классу токена.

Алфавит – это конечное множество символов.

Строка над алфавитом – это последовательность символов, взятых из этого языка.

Формальный язык – это любое счетное множество строк над некоторым фиксированным алфавитом.

Регулярное выражение – это выражение, составленное из множеств символов и языковыми операциями над ними – объединение, конкатенация и итерация.

Регулярные выражения описывают регулярные языки.

Конечный автомат – математическая абстракция, модель устройства, имеющее вход и выход и находящееся в одном из нескольких состояний.

НКА – не накладываются ограничения на дуги.

ДКА – из некоторого состояния по некоторому символу только одна дуга, нет эпсилон-переходов. N состояний НКА = 2^N-1 состояний ДКА.

Эпсилон-переход – это переход, который можно совершить без чтения входного символа.

Эпсилон-замыкание для состояния s – это множество всех состояний T, которые можно достичь по эпсилон-переходам из состояния s.

Регулярный язык – это множество слов, которые может определить некоторый конечный автомат.

Формальная грамматика — способ описания формального языка, то есть выделения некоторого подмножества из множества всех слов некоторого конечного алфавита. (правила задания языка из символов алфавита)

Терминал – это базовые символы, из которых формируются строки. (лексемы языка)

Имя токена = терминал. If, else, (, )

Нетерминалы – это синтаксические переменные, которые обозначают множества строк. N. stmt -> if (expr) stmt else stmt. stmt, expr – нетерминалы.

Продукции – это способ, которым терминалы и нетерминалы могут объединяться для создания строк. (правила подстановки)

Цель синтаксического анализа — доказать существование дерева грамматического разбора в контекстно-свободной грамматике G для входной цепочки s, состоящей из терминальных символов.

Дерево разбора – результат работы парсера, наглядно показывает, как стартовый символ грамматики порождает строку языка.

Связь между лексическим анализом и КС-грамматикой заключается в том, что лексер помогает разбить входной текст на лексемы, которые затем используются синтаксическим анализатором для построения структуры синтаксиса, описанной в КС-грамматике.

Порождение – последовательность решений о применении продукции.