Формальные грамматики и языки. Обозначим алфавит символов за σ, а множество конечных строк за $L \subset 2^{\sigma^*}$, где σ^* — это все последовательности конечной длины.

Регулярные выражения. Тривиальные регулярные выражения: (пустое множество) ϵ (пустая строка), $x \in \sigma$. Остальные регулярные выражения определяются рекурсивно относительно них. А именно, мы также разрешаем $A + B$ (последовательная запись), $A|B$ (выбор из двух регулярок), $A*$ (повторение строки из языка A $0, 1, 2, \ldots$ раз) Звездочка еще называется замыканием $Kleene$.

Например, $\{0|1|2|3\} * + \{0|1|2|3\}$ это множество всех непустых строк из символов $\sigma = \{0, 1, 2, 3\}$.

Способы задания языков. Регулярные выражения, ДКА, НКА, ϵ-НКА. Эти способы эквивалентны. Очевидно $L_{\text{ДКА}} \subset L_{\text{НКА}} \subset L_{\epsilon\text{-НКА}}$ Вложенность ϵ-НКА в ДКА можно показать, если рассмотреть алгоритм приведения одного автомата в другой. Например, можно выделить все подмножества вершин НКА как отдельные вершины ДКА (то есть, он будет иметь экспоненциальный размер). Тогда переход по ребру — это то же самое, что взять все вершины подмножества, и объединить переходы по ребрам из них. Еще стоит следить за переходами по пустому символу, потому что вершина сразу задает подмножество вершин, достижимых из нее по ϵ-ребрам.

Осталось показать, что регулярные выражения эквивалентны автоматам. По регулярным выражениям можно достаточно легко строить ϵ-НКА, если отдельно рассмотреть все возможные способы задания языка, и дальше рекурсивно. То есть, $A + B$ выражается, если последовательно записать автоматы A и B, и перегнать терминалы A в стартовую вершину для B. Остальные каким-то похожим образом.

Вот теперь совсем последний шаг — надо показать, что любой ДКА можно задать регулярным выражением. Введем $L_i(v, u)$ — регулярное выражение, которое задает все строки, по которым можно пройти из v в u в ДКА, используя первые i ребер. Изначально это какие-то пустые регулярки. Потом добавляем ребро (x, y, c). Тогда $L_{i+1}(v, u) = L_i(v, u) \cup \{L_i(v, x) + c + \{L_i(y, x) + c\}^* + L_i(y, u)\}$.

Лемма о накачке. Для любого регулярного языка L существует такое n, что для всех строк $|s| \geq n$ верно, $s = uvw$, $|uv| \leq n$, $\forall k \geq 0$ $s' = uv^k w \in L$. Доказательство такое: посмотрим на ДКА и возьмем n, равное числу вершин. Тогда для длинных строк их путь будет содержать цикл. Этот цикл как раз и соответствует строке v. Это значит, что любая достаточно длинная строка имеет циклическую структуру внутри.