Formal Languages

Philippe de Groote

2018-2019
3 Regular expressions and regular languages
- Definition
- Some algebraic properties
- From regular expressions to FSA
- From FSA to type-3 grammars
- From type-3 grammars to regular expressions
The *set of regular expressions* over an alphabet $\Sigma$ is inductively defined as follows:

- 0 is a regular expression;
- 1 is a regular expression;
- every symbol $a \in \Sigma$ is a regular expression;
- if $\alpha$ is a regular expression so is $\alpha^*$;
- if $\alpha$ and $\beta$ are regular expressions so is $(\alpha \cdot \beta)$;
- if $\alpha$ and $\beta$ are regular expressions so is $(\alpha + \beta)$;
The *set of regular expressions* over an alphabet $\Sigma$ is inductively defined as follows:

- $0$ is a regular expression;
- $1$ is a regular expression;
- every symbol $a \in \Sigma$ is a regular expression;
- if $\alpha$ is a regular expression so is $\alpha^*$;
- if $\alpha$ and $\beta$ are regular expressions so is $(\alpha \cdot \beta)$;
- if $\alpha$ and $\beta$ are regular expressions so is $(\alpha + \beta)$;

We write $\text{rexp}(\Sigma)$ for the set of regular expressions over $\Sigma$. 
Language defined by a regular expressions:

- \( L(0) = \emptyset \);
- \( L(1) = \{ \epsilon \} \);
- \( L(a) = \{a\} \) for every \( a \in \Sigma \);
- \( L(\alpha^*) = L(\alpha)^* \);
- \( L(\alpha \cdot \beta) = L(\alpha) \cdot L(\beta) \);
- \( L(\alpha + \beta) = L(\alpha) \cup L(\beta) \).
Some algebraic properties

\[(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)\]
\[\alpha + 0 = \alpha\]
\[0 + \alpha = \alpha\]
\[\alpha + \beta = \beta + \alpha\]
\[\alpha + \alpha = \alpha\]

\[(\alpha \cdot \beta) \cdot \gamma = \alpha \cdot (\beta \cdot \gamma)\]
\[\alpha \cdot 1 = \alpha\]
\[1 \cdot \alpha = \alpha\]
\[\alpha \cdot 0 = 0\]
\[0 \cdot \alpha = 0\]

\[\alpha \cdot (\beta + \gamma) = \alpha \cdot \beta + \alpha \cdot \gamma\]
\[(\alpha + \beta) \cdot \gamma = \alpha \cdot \gamma + \beta \cdot \gamma\]
Some algebraic properties

\[ 0^* = 1 \]
\[ 1^* = 1 \]
\[ (\alpha^*)^* = \alpha^* \]
\[ 1 + \alpha \cdot (\alpha^*) = \alpha^* \]
\[ 1 + \alpha^* \cdot \alpha = \alpha^* \]
Automaton accepting $L(0)$
Automaton accepting $L(0)$

Automaton accepting $L(1)$:
Automaton accepting $L(0)$:

Automaton accepting $L(1)$:

Automaton accepting $L(a)$:
Assuming we have an automaton accepting $L(\alpha)$:
Assuming we have an automaton accepting $L(\alpha)$:

```
\[ \begin{array}{c}
 \text{\hspace{.5cm} \hspace{.5cm}} \\
 \quad \alpha \\
 \end{array} \]
```

Automaton accepting $L(\alpha^*)$:

```
\[ \begin{array}{c}
 \text{\hspace{.5cm} \hspace{.5cm}} \\
 \quad \epsilon \\
 \quad \text{\hspace{.5cm} \hspace{.5cm}} \\
 \quad \alpha \\
 \quad \epsilon \\
 \end{array} \]
```
Assuming we have automata accepting $L(\alpha)$ and $L(\beta)$:
Assuming we have automata accepting $L(\alpha)$ and $L(\beta)$:

Automaton accepting $L(\alpha \cdot \beta)$:
Assuming we have automata accepting $L(\alpha)$ and $L(\beta)$:
Assuming we have automata accepting $L(\alpha)$ and $L(\beta)$:

Automaton accepting $L(\alpha + \beta)$:
Let $A = \langle Q, \Sigma, \delta, q_0, F \rangle$ be an DFSA. Define a type-3 grammar $G = \langle N, \Sigma_G, P, S \rangle$ as follows:

- $N = Q$
- $\Sigma_G = \Sigma$
- $P = \{ A \rightarrow aB : \delta(A, a) = B \} \cup \{ A \rightarrow \epsilon : A \in F \}$
- $S = q_0$
Let $A = \langle Q, \Sigma, \delta, q_0, F \rangle$ be an DFSA. Define a type-3 grammar $G = \langle N, \Sigma_G, P, S \rangle$ as follows:

- $N = Q$
- $\Sigma_G = \Sigma$
- $P = \{ A \rightarrow aB : \delta(A, a) = B \} \cup \{ A \rightarrow \epsilon : A \in F \}$
- $S = q_0$

Proposition $L(A) = L(G)$. 
**PROOF:**
PROOF:

We prove by induction on the length of $\alpha$ that $A \Rightarrow^* \alpha$ if and only if $\hat{\delta}(A, \alpha) \in F$. 
PROOF:

We prove by induction on the length of $\alpha$ that $A \Rightarrow^* \alpha$ if and only if $
\hat{\delta}(A, \alpha) \in F$.

Basis:
PROOF:

We prove by induction on the length of $\alpha$ that $A \Rightarrow^* \alpha$ if and only if $\hat{\delta}(A, \alpha) \in F$.

Basis:

$A \Rightarrow^* \epsilon$ iff $A \Rightarrow \epsilon$
**PROOF:**

We prove by induction on the length of $\alpha$ that $A \Rightarrow^* \alpha$ if and only if $\hat{\delta}(A, \alpha) \in F$.

**Basis:**

$$A \Rightarrow^* \epsilon \text{ iff } A \Rightarrow \epsilon$$
$$\text{iff } (A \rightarrow \epsilon) \in P$$
PROOF:

We prove by induction on the length of $\alpha$ that $A \Rightarrow^* \alpha$ if and only if $\hat{\delta}(A, \alpha) \in F$.

Basis:

$$A \Rightarrow^* \epsilon \iff A \Rightarrow \epsilon$$
$$\iff (A \rightarrow \epsilon) \in P$$
$$\iff A \in F$$
**PROOF:**

We prove by induction on the length of $\alpha$ that $A \Rightarrow^* \alpha$ if and only if $\hat{\delta}(A, \alpha) \in F$.

**Basis:**

\[
A \Rightarrow^* \epsilon \iff A \Rightarrow \epsilon \\
\iff (A \rightarrow \epsilon) \in P \\
\iff A \in F \\
\iff \hat{\delta}(A, \epsilon) \in F
\]
From FSA to type-3 grammars

**Induction:**

\[ A \Rightarrow^* a\alpha' \iff A \Rightarrow aB \Rightarrow^* a\alpha', \text{ for some } (A \rightarrow aB) \in P \]

\[ A \Rightarrow^* a\alpha' \iff \delta(A,a) = B \text{ and } B \Rightarrow^* \alpha' \iff \delta(A,a) = B \text{ and } \hat{\delta}(B,\alpha') \in F \text{ by induction hypothesis} \]

\[ \hat{\delta}(A,a\alpha') \in F \iff \hat{\delta}(\delta(A,a),\alpha') \in F \iff \hat{\delta}(A,a\alpha) \in F \]
Induction:

\[ A \Rightarrow^* a\alpha' \iff A \Rightarrow aB \Rightarrow^* a\alpha', \text{ for some } (A \rightarrow aB) \in P \]
From FSA to type-3 grammars

**Induction:**

\[ A \Rightarrow^* a\alpha' \text{ iff } A \Rightarrow aB \Rightarrow^* a\alpha', \text{ for some } (A \rightarrow aB) \in P \]

iff \((A \rightarrow aB) \in P\) and \(B \Rightarrow^* \alpha'\)
From FSA to type-3 grammars

Induction:

\[ A \Rightarrow^* a\alpha' \iff A \Rightarrow aB \Rightarrow^* a\alpha', \text{ for some } (A \rightarrow aB) \in P \]

iff \((A \rightarrow aB) \in P \) and \( B \Rightarrow^* \alpha' \)

iff \( \delta(A, a) = B \) and \( B \Rightarrow^* \alpha' \)
Induction:

\[ A \Rightarrow^* a\alpha' \text{ iff } A \Rightarrow aB \Rightarrow^* a\alpha', \text{ for some } (A \rightarrow aB) \in P \]

iff \( (A \rightarrow aB) \in P \) and \( B \Rightarrow^* \alpha' \)

iff \( \delta(A, a) = B \) and \( B \Rightarrow^* \alpha' \)

iff \( \delta(A, a) = B \) and \( \hat{\delta}(B, \alpha') \in F \)

by induction hypothesis
Induction:

\[ A \Rightarrow^* a\alpha' \text{ iff } A \Rightarrow aB \Rightarrow^* a\alpha', \text{ for some } (A \rightarrow aB) \in P \]
\[ \text{iff } (A \rightarrow aB) \in P \text{ and } B \Rightarrow^* \alpha' \]
\[ \text{iff } \delta(A, a) = B \text{ and } B \Rightarrow^* \alpha' \]
\[ \text{iff } \delta(A, a) = B \text{ and } \hat{\delta}(B, \alpha') \in F \]
\[ \text{by induction hypothesis} \]
\[ \text{iff } \hat{\delta}(\delta(A, a), \alpha') \in F \]
Induction:

\[ A \Rightarrow^* a\alpha' \text{ iff } A \Rightarrow aB \Rightarrow^* a\alpha', \text{ for some } (A \rightarrow aB) \in P \]

iff \( (A \rightarrow aB) \in P \) and \( B \Rightarrow^* \alpha' \)

iff \( \delta(A, a) = B \) and \( B \Rightarrow^* \alpha' \)

iff \( \delta(A, a) = B \) and \( \hat{\delta}(B, \alpha') \in F \)

by induction hypothesis

iff \( \hat{\delta}(\delta(A, a), \alpha') \in F \)

iff \( \delta(A, a\alpha') \in F \)
Consider $\alpha \in \text{rexp}(\Sigma)^*$. One defines $L(\alpha)$ as follows:

- $L(\epsilon) = \{\epsilon\}$
- $L(e\alpha') = L(e) \cdot L(\alpha')$

We consider type-3 grammars whose set of terminal symbols is the set of regular expressions over some alphabet $\Sigma$: 

$$G = \langle N, \text{rexp}(\Sigma), P, S \rangle$$
Consider $\alpha \in \text{rexp}(\Sigma)^\ast$. One defines $L(\alpha)$ as follows:

- $L(\epsilon) = \{\epsilon\}$
- $L(e\alpha') = L(e) \cdot L(\alpha')$

We consider type-3 grammars whose set of terminal symbols is the set of regular expressions over some alphabet $\Sigma$:

$$G = \langle N, \text{rexp}(\Sigma), P, S \rangle$$

For such grammars, one may define:

$$L_E(G) = \bigcup_{e \in L(G)} L(e)$$
Example:

Example:

G =

S → (a + b) S
S → (c · d)

L(G) =

(c · d), (a + b)(c · d), (a + b)(a + b)(c · d), ...

Remark:

Since \( \Sigma \subseteq \text{rexp}(\Sigma) \), every grammar over \( \Sigma \) may be seen as a grammar over \( \text{rexp}(\Sigma) \), with \( L(E(G)) = L(G) \).
Example:

\[ G = \begin{cases} 
S \rightarrow (a+b)S \\
S \rightarrow (c \cdot d) 
\end{cases} \]
From type-3 grammars to regular expressions

Example:

\[ G = \begin{cases} 
S \to (a+b) S \\
S \to (c\cdot d) 
\end{cases} \]

\[ L(G) = \{ (c\cdot d), (a+b)(c\cdot d), (a+b)(a+b)(c\cdot d), (a+b)(a+b)(a+b)(c\cdot d), \ldots \} \]
From type-3 grammars to regular expressions

Example:

\[ G = \begin{cases} 
S \rightarrow (a+b) S \\
S \rightarrow (c \cdot d) 
\end{cases} \]

\[ L(G) = \{ (c \cdot d), (a+b)(c \cdot d), (a+b)(a+b)(c \cdot d), (a+b)(a+b)(a+b)(c \cdot d), \ldots \} \]

\[ L_E(G) = \{ cd, acd, bcd, aacd, abcd, bacd, bbcd, aaacd, aabcd, abacd, \ldots \} \]
From type-3 grammars to regular expressions

Example:

\[ G = \begin{cases} 
S \to (a+b) S \\
S \to (c \cdot d)
\end{cases} \]

\[ L(G') = \{ (c \cdot d), (a+b)(c \cdot d), (a+b)(a+b)(c \cdot d), (a+b)(a+b)(a+b)(c \cdot d), \ldots \} \]

\[ L_E(G) = \{ cd, acd, bcd, aacd, abcd, bacd, bbcd, aaacd, aabcd, abacd, \ldots \} \]

Remark:

Since \( \Sigma \subseteq \text{rexp}(\Sigma) \), every grammar over \( \Sigma \) may be seen as a grammar over \( \text{rexp}(\Sigma) \), with \( L_E(G) = L(G) \).
Elimination of non-recursive non-terminal symbols
Elimination of non-recursive non-terminal symbols

Given a type-3 grammar $G$, one says that a rule is recursive if it is of the form $A \rightarrow aA$. A non-terminal symbol $A$ is said to be recursive in case there is at least one recursive rule whose lefthand side is $A$. 
Elimination of non-recursive non-terminal symbols

Given a type-3 grammar $G$, one says that a rule is recursive if it is of the form $A \to aA$. A non-terminal symbol $A$ is said to be recursive in case there is at least one recursive rule whose lefthand side is $A$.

Let $G_1 = \langle N_1, \text{rexp}(\Sigma), P_1, S_1 \rangle$ be a type-3 grammar, and let $A \in N_1$ be a non-recursive non-terminal symbol different from $S$. 

Philippe de Groote

Formal Languages

2018-2019 16 / 28
Elimination of non-recursive non-terminal symbols

Given a type-3 grammar $G$, one says that a rule is recursive if it is of the form $A \rightarrow aA$. A non-terminal symbol $A$ is said to be recursive in case there is at least one recursive rule whose lefthand side is $A$.

Let $G_1 = \langle N_1, \text{rexp}(\Sigma), P_1, S_1 \rangle$ be a type-3 grammar, and let $A \in N_1$ be a non-recursive non-terminal symbol different from $S$.

Let $P_A = \{ A \rightarrow e_0 B_0, \ldots, A \rightarrow e_{m-1} B_{m-1}, A \rightarrow f_0, \ldots, A \rightarrow f_{n-1} \} \subset P_1$ be the set of all the production rules whose lefthand side is $A$. 
Elimination of non-recursive non-terminal symbols

Given a type-3 grammar $G$, one says that a rule is recursive if it is of the form $A \rightarrow aA$. A non-terminal symbol $A$ is said to be recursive in case there is at least one recursive rule whose lefthand side is $A$.

Let $G_1 = \langle N_1, \text{rexp}(\Sigma), P_1, S_1 \rangle$ be a type-3 grammar, and let $A \in N_1$ be a non-recursive non-terminal symbol different from $S$.

Let $P_A = \{ A \rightarrow e_0 B_0, \ldots, A \rightarrow e_{m-1} B_{m-1}, A \rightarrow f_0, \ldots, A \rightarrow f_{n-1} \} \subset P_1$ be the set of all the production rules whose lefthand side is $A$.

Let $Q_A = \{ C_0 \rightarrow a_0 A, \ldots, C_{l-1} \rightarrow a_{l-1} A \} \subset P_1$ be the set of all the production rules whose righthand side contains $A$. 
Elimination of non-recursive non-terminal symbols

Given a type-3 grammar \( G \), one says that a rule is recursive if it is of the form \( A \rightarrow aA \). A non-terminal symbol \( A \) is said to be recursive in case there is at least one recursive rule whose lefthand side is \( A \).

Let \( G_1 = \langle N_1, \text{rexp}(\Sigma), P_1, S_1 \rangle \) be a type-3 grammar, and let \( A \in N_1 \) be a non-recursive non-terminal symbol different from \( S \).

Let \( P_A = \{ A \rightarrow e_0 B_0, \ldots, A \rightarrow e_{m-1} B_{m-1}, A \rightarrow f_0, \ldots, A \rightarrow f_{n-1} \} \subset P_1 \) be the set of all the production rules whose lefthand side is \( A \).

Let \( Q_A = \{ C_0 \rightarrow a_0 A, \ldots, C_{l-1} \rightarrow a_{l-1} A \} \subset P_1 \) be the set of all the production rules whose righthand side contains \( A \).

Define \( R_A = \bigcup_{i \in l} \left( (\bigcup_{j \in m} \{ C_i \rightarrow (a_i \cdot e_j B_j) \} \right) \cup (\bigcup_{j \in n} \{ C_i \rightarrow (a_i \cdot f_j) \}) \))
One defines a new grammar $G_2 = \langle N_2, \text{rexp}(\Sigma), P_2, S_2 \rangle$ as follows:

- $N_2 = N_1 \setminus \{A\}$
- $P_2 = (P \setminus (P_A \cup Q_A)) \cup R_A$
- $S_2 = S_1$
One defines a new grammar $G_2 = \langle N_2, \text{rexp}(\Sigma), P_2, S_2 \rangle$ as follows:

- $N_2 = N_1 \setminus \{A\}$
- $P_2 = (P \setminus (P_A \cup Q_A)) \cup R_A$
- $S_2 = S_1$

Proposition $L_E(G_1) = L_E(G_2)$. 
PROOF:

We prove that $L(E(G_1)) \subset L(E(G_2))$ and $L(E(G_2)) \subset L(E(G_1))$.

**PART 1:**

$L(E(G_1)) \subset L(E(G_2))$

Let us write $\Rightarrow_1$ and $\Rightarrow_2$ for the generation relations of $G_1$ and $G_2$, respectively. We prove that for every $B \in N_{G_2}$ and every $\alpha_1 \in \text{rexp}(\Sigma)^*$ such that $B \Rightarrow_1^* \alpha_1$, there exists $\alpha_2 \in \text{rexp}(\Sigma)^*$ such that $B \Rightarrow_2^* \alpha_2$ and $L(\alpha_1) = L(\alpha_2)$. The proof proceeds by induction on the number of occurrences of rules from $Q_A$ that appear in the derivation $B \Rightarrow_1^* \alpha_1$.

**Basis:**

There is no occurrence of any rule from $Q_A$ in the derivation $B \Rightarrow_1^* \alpha_1$. Then, there is no occurrence of any rule from $P_A$ either. Consequently, the derivation $B \Rightarrow_1^* \alpha_1$ is also a derivation of $G_2$, and we take $\alpha_2 = \alpha_1$. 

Philippe de Groote

Formal Languages

2018-2019 18 / 28
**PROOF:**

We prove that $L_E(G_1) \subset L_E(G_2)$ and $L_E(G_2) \subset L_E(G_1)$. 
\textit{PROOF}: \\
We prove that $L_E(G_1) \subset L_E(G_2)$ and $L_E(G_2) \subset L_E(G_1)$. \\

\textbf{PART 1:} $L_E(G_1) \subset L_E(G_2)$
**PROOF:**

We prove that $L_E(G_1) \subset L_E(G_2)$ and $L_E(G_2) \subset L_E(G_1)$.

**PART 1: $L_E(G_1) \subset L_E(G_2)$**

Let us write $\Rightarrow_1$ and $\Rightarrow_2$ for the generation relations of $G_1$ and $G_2$, respectively. We prove that for every $B \in N_2$ and every $\alpha_1 \in \text{rexp}(\Sigma)^*$ such that $B \Rightarrow_1^* \alpha_1$, there exists $\alpha_2 \in \text{rexp}(\Sigma)^*$ such that $B \Rightarrow_2^* \alpha_2$ and $L(\alpha_1) = L(\alpha_2)$. The proof proceed by induction on the number of occurences of rules from $Q_A$ that appear in the derivation $B \Rightarrow_1^* \alpha_1$. 
**From type-3 grammars to regular expressions**

**PROOF:**

We prove that $L_E(G_1) \subset L_E(G_2)$ and $L_E(G_2) \subset L_E(G_1)$.

**PART 1:** $L_E(G_1) \subset L_E(G_2)$

Let us write $\Rightarrow_1$ and $\Rightarrow_2$ for the generation relations of $G_1$ and $G_2$, respectively. We prove that for every $B \in N_2$ and every $\alpha_1 \in \text{rexp}(\Sigma)^*$ such that $B \Rightarrow_1^* \alpha_1$, there exists $\alpha_2 \in \text{rexp}(\Sigma)^*$ such that $B \Rightarrow_2^* \alpha_2$ and $L(\alpha_1) = L(\alpha_2)$. The proof proceed by induction on the number of occurrences of rules from $Q_A$ that appear in the derivation $B \Rightarrow_1^* \alpha_1$.

**Basis:**
PROOF:

We prove that \( L_E(G_1) \subset L_E(G_2) \) and \( L_E(G_2) \subset L_E(G_1) \).

PART 1: \( L_E(G_1) \subset L_E(G_2) \)

Let us write \( \Rightarrow_1 \) and \( \Rightarrow_2 \) for the generation relations of \( G_1 \) and \( G_2 \), respectively. We prove that for every \( B \in N_2 \) and every \( \alpha_1 \in \text{rexp}(\Sigma)^* \) such that \( B \Rightarrow_1^* \alpha_1 \), there exists \( \alpha_2 \in \text{rexp}(\Sigma)^* \) such that \( B \Rightarrow_2^* \alpha_2 \) and \( L(\alpha_1) = L(\alpha_2) \). The proof proceed by induction on the number of occurences of rules from \( Q_A \) that appear in the derivation \( B \Rightarrow_1^* \alpha_1 \).

Basis:

There is no occurence of any rule from \( Q_A \) in the derivation \( B \Rightarrow_1^* \alpha_1 \). Then, there is no occurrence of any rule from \( P_A \) either. Consequently, the derivation \( B \Rightarrow_1^* \alpha_1 \) is also a derivation of \( G_2 \), and we take \( \alpha_2 = \alpha_1 \).
Regular expressions and regular languages

From type-3 grammars to regular expressions

**Induction:**

- If there is at least one occurrence of a rule from $Q_A$ in the derivation $B \Rightarrow \ast \alpha_1$, it must obey one of the two following forms:
  
  $$(1) \quad B \Rightarrow \ast \beta C_i \Rightarrow \beta a_i A \Rightarrow \beta a_i e_j B_j \Rightarrow \ast \gamma_1$$

  $$(2) \quad B \Rightarrow \ast \beta C_i \Rightarrow \beta a_i A \Rightarrow \beta a_i f_j B_j \Rightarrow \ast \gamma_2$$

  where the occurrence of $(C_i \rightarrow a_i A) \in Q_A$ is the leftmost occurrence of a rule from $Q_A$.

- Consequently, $B \Rightarrow \ast \beta C_i$ is also a derivation of $G$.

In the first case, we have $\alpha_1 = \beta a_i e_j \gamma_1$ and $B_j \Rightarrow \ast \gamma_1$. By induction hypothesis, there exists $\gamma_2 \in \text{rexp}(\Sigma)^\ast$ such that $B_j \Rightarrow \ast \gamma_2$ and $L(\gamma_1) = L(\gamma_2)$.

Hence:

$$B \Rightarrow \ast \beta C_i \Rightarrow \beta (a_i \cdot e_j) B_j \Rightarrow \ast \gamma_2$$

Then, we take $\alpha_2 = \beta (a_i \cdot e_j) \gamma_2$.

Indeed $L(\alpha_1) = L(\beta a_i e_j \gamma_1) = L(\beta) L(a_i) L(e_j) L(\gamma_1) = L(\beta) L(a_i \cdot e_j) L(\gamma_2) = L(\alpha_2)$. 

- Philippe de Groote
Induction:

If there is at least one occurrence of a rule from $Q_A$ in the derivation $B \Rightarrow_1^* \alpha_1$, it must obey one of the two following forms:

(1) $B \Rightarrow_1 * \beta C_i \Rightarrow_1 \beta a_i A \Rightarrow_1 \beta a_i e_j B_j \Rightarrow_1 * \beta a_i e_j \gamma_1$

(2) $B \Rightarrow_1 * \beta C_i \Rightarrow_1 \beta a_i A \Rightarrow_1 \beta a_i f_j$

where the occurrence of $(C_i \rightarrow a_i A) \in Q_A$ is the leftmost occurrence of a rule from $Q_A$. Consequently, $B \Rightarrow_1^* \beta C_i$ is also a derivation of $G_2$. 

**Induction:**

If there is at least one occurrence of a rule from $Q_A$ in the derivation $B \Rightarrow_1^* \alpha_1$, it must obey one of the two following forms:

1. $B \Rightarrow_1^* \beta C_i \Rightarrow_1 \beta a_i A \Rightarrow_1 \beta a_i e_j B_j \Rightarrow_1^* \beta a_i e_j \gamma_1$
2. $B \Rightarrow_1^* \beta C_i \Rightarrow_1 \beta a_i A \Rightarrow_1 \beta a_i f_j$

where the occurrence of $(C_i \rightarrow a_i A) \in Q_A$ is the leftmost occurrence of a rule from $Q_A$. Consequently, $B \Rightarrow_1^* \beta C_i$ is also a derivation of $G_2$.

In the first case, we have $\alpha_1 = \beta a_i e_j \gamma_1$ and $B_j \Rightarrow_1^* \gamma_1$. By induction hypothesis, there exists $\gamma_2 \in \text{rexp}(\Sigma)^*$ such that $B_j \Rightarrow_2^* \gamma_2$ and $L(\gamma_1) = L(\gamma_2)$. Hence:

$$B \Rightarrow_2^* \beta C_i \Rightarrow_2 \beta (a_i \cdot e_j) B_j \Rightarrow_2^* \beta (a_i \cdot e_j) \gamma_2$$

Then, we take $\alpha_2 = \beta (a_i \cdot e_j) \gamma_2$. Indeed $L(\alpha_1) = L(\beta a_i e_j \gamma_1) = L(\beta) L(a_i) L(e_j) L(\gamma_1) = L(\beta) L(a_i \cdot e_j) L(\gamma_2) = L(\beta (a_i \cdot e_j) \gamma_2) = L(\alpha_2)$
In the second case, we have $\alpha_1 = \beta a_i f_j$. Then, we take $\alpha_2 = \beta(a_i \cdot f_j)$.

Indeed, we have that

$$B \Rightarrow^*_2 \beta C_i \Rightarrow^*_2 \beta(a_i \cdot f_j)$$

and that $L(\beta a_i f_j) = L(\beta(a_i \cdot f_j))$. 
In the second case, we have $\alpha_1 = \beta a_i f_j$. Then, we take $\alpha_2 = \beta(a_i \cdot f_j)$

Indeed, we have that

$$B \Rightarrow^* \beta_2 C_i \Rightarrow_2 \beta(a_i \cdot f_j)$$

and that $L(\beta a_i f_j) = L(\beta(a_i \cdot f_j))$.

PART 2: $L_E(G_2) \subset L_E(G_1)$
In the second case, we have $\alpha_1 = \beta a_i f_j$. Then, we take $\alpha_2 = \beta (a_i \cdot f_j)$.

Indeed, we have that

$$B \Rightarrow^* \beta C_i \Rightarrow^2 \beta (a_i \cdot f_j)$$

and that $L(\beta a_i f_j) = L(\beta (a_i \cdot f_j))$.

PART 2: $L_E(G_2) \subseteq L_E(G_1)$

We prove that for every $B \in N_2$ and every $\alpha_2 \in \text{rexp}(\Sigma)^*$ such that $B \Rightarrow^*_2 \alpha_2$, there exists $\alpha_1 \in \text{rexp}(\Sigma)^*$ such that $B \Rightarrow^*_1 \alpha_1$ and $L(\alpha_2) = L(\alpha_1)$. The proof, which proceed by induction on the number of occurrences of rules from $R_A$ that appear in the derivation $B \Rightarrow^*_2 \alpha_2$, is similar to the proof of Part 1.
Elimination of recursive rules
Elimination of recursive rules

Let $G_1 = \langle N_1, \text{rexp}(\Sigma), P_1, S_1 \rangle$ be a type-3 grammar, and let $A \in N_1$ be a recursive non-terminal symbol different from $S$. 
Elimination of recursive rules

Let $G_1 = \langle N_1, \text{rexp}(\Sigma), P_1, S_1 \rangle$ be a type-3 grammar, and let $A \in N_1$ be a recursive non-terminal symbol different from $S$.

Let $P_A = \{ A \rightarrow e_0 B_0, \ldots, A \rightarrow e_{l-1} B_{l-1}, A \rightarrow f_0, \ldots, A \rightarrow f_{m-1} \} \subset P_1$ be the set of all the non-recursive production rules whose lefthand side is $A$. 
Elimination of recursive rules

Let $G_1 = \langle N_1, \text{rexp}(\Sigma), P_1, S_1 \rangle$ be a type-3 grammar, and let $A \in N_1$ be a recursive non-terminal symbol different from $S$.

Let $P_A = \lbrace A \rightarrow e_0 B_0, \ldots, A \rightarrow e_{l-1} B_{l-1}, A \rightarrow f_0, \ldots, A \rightarrow f_{m-1} \rbrace \subset P_1$ be the set of all the non-recursive production rules whose lefthand side is $A$.

Let $Q_A = \lbrace A \rightarrow g_0 A, \ldots, A \rightarrow g_{n-1} A \rbrace \subset P_1$ be the set of all the recursive production rules whose lefthand side is $A$. 
Elimination of recursive rules

Let $G_1 = \langle N_1, \text{rexp}(\Sigma), P_1, S_1 \rangle$ be a type-3 grammar, and let $A \in N_1$ be a recursive non-terminal symbol different from $S$.

Let $P_A = \{A \rightarrow e_0 B_0, \ldots, A \rightarrow e_{l-1} B_{l-1}, A \rightarrow f_0, \ldots, A \rightarrow f_{m-1}\} \subset P_1$ be the set of all the non-recursive production rules whose lefthand side is $A$.

Let $Q_A = \{A \rightarrow g_0 A, \ldots, A \rightarrow g_{n-1} A\} \subset P_1$ be the set of all the recursive production rules whose lefthand side is $A$.

Define $R_A = (\bigcup_{i \in l} \{A \rightarrow ((g_0 + \cdots + g_{n-1})^* e_i) B_i\}) \cup (\bigcup_{i \in m} \{A \rightarrow ((g_0 + \cdots + g_{n-1})^* f_i)\})$
One defines a new grammar $G_2 = \langle N_2, \text{rexp}(\Sigma), P_2, S_2 \rangle$ as follows:

- $N_2 = N_1$
- $P_2 = (P \setminus (P_A \cup Q_A)) \cup R_A$
- $S_2 = S_1$
One defines a new grammar \( G_2 = \langle N_2, \text{rexp}(\Sigma), P_2, S_2 \rangle \) as follows:

- \( N_2 = N_1 \)
- \( P_2 = (P \setminus (P_A \cup Q_A)) \cup R_A \)
- \( S_2 = S_1 \)

Proposition \( L_E(G_1) = L_E(G_2) \).
PROOF:

\[ L(G_1) \subset L(G_2) \]

We prove that for every \( B \in \mathbb{N}_1 \) and every \( \alpha_1 \in \text{rexp}(\Sigma)^* \) such that \( B \Rightarrow^* \alpha_1 \), there exists \( \alpha_2 \in \text{rexp}(\Sigma)^* \) such that \( B \Rightarrow^* \alpha_2 \) and \( L(\alpha_1) \subset L(\alpha_2) \). The proof proceed by induction on the number of occurrences of rules from \( P \) that appear in the derivation \( B \Rightarrow^* \alpha_1 \).

Basis:

There is no occurrence of any rule from \( P \) in the derivation \( B \Rightarrow^* \alpha_1 \). Then, there is no occurrence of any rule from \( Q \) either. Consequently, the derivation \( B \Rightarrow^* \alpha_1 \) is also a derivation of \( G_2 \), and we take \( \alpha_2 = \alpha_1 \).
PROOF:

PART 1: \( L_E(G_1) \subset L_E(G_2) \)
**PROOF:**

**PART 1:** $L_E(G_1) \subset L_E(G_2)$

We prove that for every $B \in N_1$ and every $\alpha_1 \in \text{rexp}(\Sigma)^*$ such that $B \Rightarrow_1^* \alpha_1$, there exists $\alpha_2 \in \text{rexp}(\Sigma)^*$ such that $B \Rightarrow_2^* \alpha_2$ and $L(\alpha_1) \subset L(\alpha_2)$. The proof proceed by induction on the number of occurences of rules from $P_A$ that appear in the derivation $B \Rightarrow_1^* \alpha_1$. 

**Basis:**

There is no occurrence of any rule from $P_A$ in the derivation $B \Rightarrow_1^* \alpha_1$. Then, there is no occurrence of any rule from $Q_A$ either. Consequently, the derivation $B \Rightarrow_1^* \alpha_1$ is also a derivation of $G_2$, and we take $\alpha_2 = \alpha_1$. 

Philippe de Groote
From type-3 grammars to regular expressions

**PROOF:**

**PART 1:** $L_E(G_1) \subset L_E(G_2)$

We prove that for every $B \in N_1$ and every $\alpha_1 \in \text{rexp}(\Sigma)^*$ such that $B \Rightarrow_1^* \alpha_1$, there exists $\alpha_2 \in \text{rexp}(\Sigma)^*$ such that $B \Rightarrow_2^* \alpha_2$ and $L(\alpha_1) \subset L(\alpha_2)$. The proof proceed by induction on the number of occurrences of rules from $P_A$ that appear in the derivation $B \Rightarrow_1^* \alpha_1$.

**Basis:**
**PROOF:**

**PART 1:** \( L_E(G_1) \subset L_E(G_2) \)

We prove that for every \( B \in N_1 \) and every \( \alpha_1 \in \text{rexp}(\Sigma)^* \) such that \( B \Rightarrow_1^* \alpha_1 \), there exists \( \alpha_2 \in \text{rexp}(\Sigma)^* \) such that \( B \Rightarrow_2^* \alpha_2 \) and \( L(\alpha_1) \subset L(\alpha_2) \). The proof proceed by induction on the number of occurrences of rules from \( P_A \) that appear in the derivation \( B \Rightarrow_1^* \alpha_1 \).

**Basis:**

There is no occurrence of any rule from \( P_A \) in the derivation \( B \Rightarrow_1^* \alpha_1 \). Then, there is no occurrence of any rule from \( Q_A \) either. Consequently, the derivation \( B \Rightarrow_1^* \alpha_1 \) is also a derivation of \( G_2 \), and we take \( \alpha_2 = \alpha_1 \).
Induction:

If there is at least one occurrence of a rule from $P_A$ in the derivation $B \Rightarrow \ast \alpha_1$, it must obey one of the two following forms:

1. $B \Rightarrow \ast \beta \alpha \Rightarrow \ast \beta g_{i_0} \alpha \Rightarrow \ast \beta g_{i_0} \cdots \Rightarrow \ast \beta g_{i_k-1} \alpha \Rightarrow \ast \beta g_{i_0} \cdots g_{i_k} e_i B_i \Rightarrow \ast \beta g_{i_0} \cdots g_{i_k} e_i \gamma_1$

2. $B \Rightarrow \ast \beta \alpha \Rightarrow \ast \beta g_{i_0} \alpha \Rightarrow \ast \beta g_{i_0} \cdots \Rightarrow \ast \beta g_{i_k-1} \alpha \Rightarrow \ast \beta g_{i_0} \cdots g_{i_k} f_i$ where the occurrence of $(A \rightarrow e_i B_i) \in P_A$ (respectively, $(A \rightarrow f_i) \in P_A$) is the leftmost occurrence of a rule from $P_A$ (respectively, $Q_A$).

Consequently, $B \Rightarrow \ast \beta \alpha$ is also a derivation of $G_2$. 
From type-3 grammars to regular expressions

**Induction:**

If there is at least one occurrence of a rule from $P_A$ in the derivation $B \Rightarrow_1^* \alpha_1$, it must obey one of the two following forms:

1. $B \Rightarrow_1^* \beta A \Rightarrow_1 \beta g_{i_0} A \Rightarrow_1 \cdots \Rightarrow_1 \beta g_{i_0} \cdots g_{i_{k-1}} A$
   $\Rightarrow_1 \beta g_{i_0} \cdots g_{i_{k-1}} e_i B_i \Rightarrow_1 \beta g_{i_0} \cdots g_{i_{k-1}} e_i \gamma_1$

2. $B \Rightarrow_1^* \beta A \Rightarrow_1 \beta g_{i_0} A \Rightarrow_1 \cdots \Rightarrow_1 \beta g_{i_0} \cdots g_{i_{k-1}} A$
   $\Rightarrow_1 \beta g_{i_0} \cdots g_{i_{k-1}} f_i$

where the occurrence of $(A \rightarrow e_i B_i) \in P_A$ (respectively, $(A \rightarrow f_i) \in P_A$) is the leftmost occurrence of a rule from $P_A$, and the occurrence of $(A \rightarrow g_{i_0} A) \in Q_A$ is the leftmost occurrence of a rule from $Q_A$.

Consequently, $B \Rightarrow_1^* \beta A$ is also a derivation of $G_2$. 
From type-3 grammars to regular expressions

In the first case, we have \( \alpha_1 = \beta g_{i_0} \cdots g_{i_{k-1}} e_i \gamma_1 \) and \( B_i \Rightarrow_1^* \gamma_1 \). By induction hypothesis, there exists \( \gamma_2 \in \text{rexp}(\Sigma)^* \) such that \( B_i \Rightarrow_2^* \gamma_2 \) and \( L(\gamma_1) \subset L(\gamma_2) \). Hence:

\[
B \Rightarrow_2^* \beta A \Rightarrow_2^* \beta((g_0 + \cdots + g_{n-1})^* e_i)B_i \Rightarrow_2^* \beta((g_0 + \cdots + g_{n-1})^* e_i) \gamma_2
\]

Then, we take

\( \alpha_2 = \beta((g_0 + \cdots + g_{n-1})^* e_i) \gamma_2 \)

Indeed

\( L(\alpha_1) \subset L(\alpha_2) \)

because

\( L(g_{i_0} \cdots g_{i_{k-1}}) \subset L((g_0 + \cdots + g_{n-1})^*) \) and \( L(\alpha_1) \subset L(\alpha_2) \)
From type-3 grammars to regular expressions

In the first case, we have \( \alpha_1 = \beta g_{i_0} \cdots g_{i_{k-1}} e_i \gamma_1 \) and \( B_i \Rightarrow_1^* \gamma_1 \). By induction hypothesis, there exists \( \gamma_2 \in \text{rexp}(\Sigma)^* \) such that \( B_i \Rightarrow_2^* \gamma_2 \) and \( L(\gamma_1) \subset L(\gamma_2) \). Hence:

\[
B \Rightarrow_2^* \beta A \Rightarrow_2^* \beta((g_0 + \cdots + g_{n-1})^* e_i) B_i \Rightarrow_2^* \beta((g_0 + \cdots + g_{n-1})^* e_i) \gamma_2
\]

Then, we take

\[
\alpha_2 = \beta((g_0 + \cdots + g_{n-1})^* e_i) \gamma_2
\]

Indeed

\[
L(\alpha_1) \subset L(\alpha_2)
\]

because

\[
L(g_{i_0} \cdots g_{i_{k-1}}) \subset L((g_0 + \cdots + g_{n-1})^*) \quad \text{and} \quad L(\alpha_1) \subset L(\alpha_2)
\]

Similarly, in the second case, we have

\[
B \Rightarrow_2^* \beta A \Rightarrow_2^* \beta((g_0 + \cdots + g_{n-1})^* f_i)
\]

And, we take

\[
\alpha_2 = \beta((g_0 + \cdots + g_{n-1})^* f_i)
\]
We prove that for every $B \in N_2$, every $\alpha_2 \in \text{rexp}(\Sigma)^*$ such that $B \Rightarrow \alpha_2$, and every $\omega \in L(\alpha_2)$, there exists $\alpha_1 \in \text{rexp}(\Sigma)^*$ such that $B \Rightarrow \alpha_1$ and $\omega \in L(\alpha_1)$.

The proof proceeds by induction on the number of occurrences of rules from $RA$ that appear in the derivation $B \Rightarrow \alpha_2$.

**Basis:**
There is no occurrence of any rule from $RA$ in the derivation $B \Rightarrow \alpha_2$.

Consequently, the derivation $B \Rightarrow \alpha_2$ is also a derivation of $G_1$, and we take $\alpha_1 = \alpha_2$. 

**Part 2:** $L_E(G_2) \subset L_E(G_1)$
PART 2: \( L_E(G_2) \subset L_E(G_1) \)

We prove that for every \( B \in N_2 \), every \( \alpha_2 \in \text{rexp}(\Sigma)^* \) such that \( B \Rightarrow^*_2 \alpha_2 \), and every \( \omega \in L(\alpha_2) \), there exists \( \alpha_1 \in \text{rexp}(\Sigma)^* \) such that \( B \Rightarrow^*_1 \alpha_1 \) and \( \omega \in L(\alpha_1) \). The proof proceed by induction on the number of occurrences of rules from \( R_A \) that appear in the derivation \( B \Rightarrow^*_2 \alpha_2 \).
PART 2: \( L_E(G_2) \subset L_E(G_1) \)

We prove that for every \( B \in N_2 \), every \( \alpha_2 \in \text{rexp}(\Sigma)^* \) such that \( B \Rightarrow^*_2 \alpha_2 \), and every \( \omega \in L(\alpha_2) \), there exists \( \alpha_1 \in \text{rexp}(\Sigma)^* \) such that \( B \Rightarrow^*_1 \alpha_1 \) and \( \omega \in L(\alpha_1) \). The proof proceed by induction on the number of occurrences of rules from \( R_A \) that appear in the derivation \( B \Rightarrow^*_2 \alpha_2 \).

Basis:
PART 2: $L_E(G_2) \subset L_E(G_1)$

We prove that for every $B \in N_2$, every $\alpha_2 \in \text{rexp}(\Sigma)^*$ such that $B \Rightarrow^*_2 \alpha_2$, and every $\omega \in L(\alpha_2)$, there exists $\alpha_1 \in \text{rexp}(\Sigma)^*$ such that $B \Rightarrow^*_1 \alpha_1$ and $\omega \in L(\alpha_1)$. The proof proceed by induction on the number of occurrences of rules from $R_A$ that appear in the derivation $B \Rightarrow^*_2 \alpha_2$.

Basis:

There is no occurrence of any rule from $R_A$ in the derivation $B \Rightarrow^*_2 \alpha_2$. Consequently, the derivation $B \Rightarrow^*_2 \alpha_2$ is also a derivation of $G_1$, and we take $\alpha_1 = \alpha_2$. 
Induction:

If there is at least one occurrence of a rule from $R_A$ in the derivation $B \Rightarrow \ast 2\alpha$, it must obey one of the two following forms:

(1) $B \Rightarrow \ast 2\beta A \Rightarrow \ast \gamma$

where the occurrence of $(A \rightarrow (\ast \cdot e_i)) \in R_A$ (respectively, $(A \rightarrow (\ast \cdot f_i)) \in R_A$) is the leftmost occurrence of a rule from $R_A$. Consequently, $B \Rightarrow \ast 2\beta A$ is also a derivation of $G_1$. 
Induction:

If there is at least one occurrence of a rule from $R_A$ in the derivation $B \Rightarrow_2^* \alpha_2$, it must obey one of the two following forms:

\begin{align*}
(1) & \quad B \Rightarrow_2^* \beta A \Rightarrow_2 \beta((g_0 + \cdots + g_{n-1})^* \cdot e_i) B_i \Rightarrow_2 \beta((g_0 + \cdots + g_{n-1})^* \cdot e_i) \gamma_2 \\
(2) & \quad B \Rightarrow_2^* \beta A \Rightarrow_2 \beta((g_0 + \cdots + g_{n-1})^* \cdot f_i)
\end{align*}

where the occurrence of $(A \rightarrow ((g_0 + \cdots + g_{n-1})^* \cdot e_i) B_i) \in R_A$
(respectively, $(A \rightarrow ((g_0 + \cdots + g_{n-1})^* \cdot f_i)) \in R_A$) is the leftmost occurrence of a rule from $R_A$. Consequently, $B \Rightarrow_2^* \beta A$ is also a derivation of $G_1$. 
In the first case, we have $\alpha_2 = \beta((g_0 + \cdots + g_{n-1})^* e_i)\gamma_2$ and $B_i \Rightarrow^*_2 \gamma_2$. Now, let $\omega \in L(\alpha_2)$. It must obey the following form:

$$\omega = \omega_1 g_{i_0} \cdots g_{i_{k-1}} \omega_2 \omega_3$$

where:

- $\omega_1 \in L(\beta)$;
- the sequence of $g_i$’s is possibly empty;
- $\omega_2 \in L(e_i)$;
- $\omega_3 \in L(\gamma_2)$.

By induction hypothesis, there exists $\gamma_1 \in \text{rexp}(\Sigma)^*$ such that $B_i \Rightarrow^*_1 \gamma_1$ and $\omega_3 \in L(\gamma_1)$. Then, we take

$$\alpha_1 = \beta g_{i_0} \cdots g_{i_{k-1}} e_i \gamma_2$$

Indeed

$$B \Rightarrow^*_1 \beta A \Rightarrow^*_1 \beta g_{i_0} A \Rightarrow^*_1 \cdots \Rightarrow^*_1 \beta g_{i_0} \cdots g_{i_{k-1}} A$$

$$\Rightarrow^*_1 \beta g_{i_0} \cdots g_{i_{k-1}} e_i B_i \Rightarrow^*_1 \beta g_{i_0} \cdots g_{i_{k-1}} e_i \gamma_1$$
From type-3 grammars to regular expressions

In the first case, we have $\alpha_2 = \beta((g_0 + \cdots + g_{n-1})^* e_i)\gamma_2$ and $B_i \Rightarrow_2^* \gamma_2$. Now, let $\omega \in L(\alpha_2)$. It must obey the following form:

$$\omega = \omega_1 g_{i_0} \cdots g_{i_{k-1}} \omega_2 \omega_3$$

where:
- $\omega_1 \in L(\beta)$;
- the sequence of $g_i$’s is possibly empty;
- $\omega_2 \in L(e_i)$;
- $\omega_3 \in L(\gamma_2)$.

By induction hypothesis, there exists $\gamma_1 \in \text{rexp}(\Sigma)^*$ such that $B_i \Rightarrow_1^* \gamma_1$ and $\omega_3 \in L(\gamma_1)$. Then, we take

$$\alpha_1 = \beta g_{i_0} \cdots g_{i_{k-1}} e_i \gamma_2$$

Indeed

$$B \Rightarrow_1^* \beta A \Rightarrow_1^* \beta g_{i_0} A \Rightarrow_1 \cdots \Rightarrow_1 \beta g_{i_0} \cdots g_{i_{k-1}} A$$

$$\Rightarrow_1 \beta g_{i_0} \cdots g_{i_{k-1}} e_i B_i \Rightarrow_1 \beta g_{i_0} \cdots g_{i_{k-1}} e_i \gamma_1$$

The second case, is similar.