# Langages Formels

## TD n°2

Emilie Grienenberger
emilie.grienenberger@lsv.fr

3 février 2021

**Exercise 1 : Arden Lemma**

Let $A, B$ be two languages.

1. Prove that the language $L = A^*B$ is the smallest solution to the equation :

$$X = (A \cdot X) \cup B$$

First, we prove that it is a solution :

$$\begin{aligned}(A \cdot A^*B) \cup B &= A^+B \cup A^0B \\ &= A^*B\end{aligned}$$

Let $S$ be a solution to this equation ; a first observation is that $A^0B = B \subseteq S$. By induction, for every $n \in \mathbb{N}$, $A^nB \subseteq S$. Indeed, if $A^nB \subseteq S$, then $A \cdot A^nB = A^{n+1}B \subseteq S$.

2. Prove that if $\varepsilon \notin A$, then it is the only solution.

By contradiction, let $S$ be a solution of $X = (A \cdot X) \cup B$ such that $\varepsilon \notin A$ and $A^*B \subsetneq S$. Let $w$ be the smallest word in $S$ which is not in $A^*B$.

As a consequence, $w \in (A \cdot S) \cup B$. As $w \notin B$ by hypothesis, $w = w_1 \cdot w_2$ where $w_1 \in A$ and $w_2 \in S$. By hypothesis, $w_2$ cannot be in $A^*B$ so $w_2 \in S \backslash A^*B$. However, $\varepsilon \notin A$ which leads to an absurdity as $w_2$ is strictly shorter than $w$.

**Exercise 2 : Regular identities**

We study identities on regular expressions $r, s, t$. Here, $r = s$ means $\mathcal{L}(r) = \mathcal{L}(s)$.

1. Prove the following identities :

   (a) $(r + s)t = rt + st$

$$\begin{aligned}
& w \in \mathcal{L}((r+s)t) \\
\Leftrightarrow\ & w = w_1 w_2,\ \text{s.t. } w_1 \in \mathcal{L}(r+s),\ w_2 \in \mathcal{L}(t) \\
\Leftrightarrow\ & w = w_1 w_2,\ \text{s.t. } w_1 \in \mathcal{L}(r) \text{ or } w_1 \in \mathcal{L}(s),\ w_2 \in \mathcal{L}(t) \\
\Leftrightarrow\ & w = w_1 w_2,\ \text{s.t. } w_1 \in \mathcal{L}(r),\ w_2 \in \mathcal{L}(t) \text{ or } w_1 \in \mathcal{L}(s),\ w_2 \in \mathcal{L}(t) \\
\Leftrightarrow\ & w \in \mathcal{L}(rt+st)
\end{aligned}$$

(b) $(r^*)^* = r^*$

$$\begin{aligned}
& w \in \mathcal{L}((r^*)^*) \\
\Leftrightarrow\ & w = w_{1,1} \dots w_{1,i_1} \dots w_{n,1} \dots w_{n,i_n},\ i_1, \dots, i_n, n \in \mathbb{N},\ w_{j,\ell} \in \mathcal{L}(r) \\
\Leftrightarrow\ & w = w_1 \dots w_n,\ n \in \mathbb{N}, w_i \in \mathcal{L}(r) \\
\Leftrightarrow\ & w \in \mathcal{L}(r^*)
\end{aligned}$$

(c) $(rs+r)^*r = r(sr+r)^*$

By induction, we prove that $(rs+r)^n r = r(sr+r)^n$.

- Immediate if $n = 0$, i.e. $r = r$.
- $(rs+r)^{n+1}r = (sr+r)(sr+r)^n r = (rs+r)r(sr+r)^n$. We observe that $w \in \mathcal{L}((rs+r)r)$ iff $w = w_1w_2w_3$ or $w = w_1w_3$ where $w_1, w_3 \in \mathcal{L}(r)$ and $w_2 \in \mathcal{L}(s)$ iff $w \in \mathcal{L}(r(sr+r))$.

2. Prove or disprove the following identities :

(a) $(r+s)^* = r^* + s^*$

Counter example : $r = a$ and $s = b$, $aba$ is in only one of these languages.

(b) $(r^*s^*)^* = (r+s)^*$

Expanding or grouping the words $\mathcal{L}(r)$ and $\mathcal{L}(s)$ goes from one expression to the other.

(c) $s(rs+s)^*r = rr^*s(rr^*s)^*$

Counter example : $r = a$ and $s = b$, $ba$ is in only one of these languages.

**Exercise 3 : Antimirov's automaton**

The goal of this exercise is to build an automaton from a regular expression. We will define a *partial derivative* operation $\partial_a(E)$ which corresponds to $a^{-1}\mathcal{L}(E)$ (via interpretation of expressions). Formally, for every expression $E$ and letter

$a$, we define the *set of expressions* $\partial_a(E)$ as follows :

$$\begin{array}{rcl}
\partial_a(\underline{\emptyset}) & = & \emptyset \\
\partial_a(\underline{b}) & = & \begin{cases} \emptyset & \text{if } a \neq b \\ \{\underline{\emptyset}^*\} & \text{else} \end{cases} \\
\partial_a(E+E') & = & \partial_a(E) \cup \partial_a(E') \\
\partial_a(E^*) & = & \partial_a(E) \cdot \{E^*\} \\
\partial_a(E \cdot E') & = & \begin{cases} \partial_a(E) \cdot \{E'\} & \text{if } \varepsilon \notin E \\ (\partial_a(E) \cdot \{E'\}) \cup \partial_a(E') & \text{else} \end{cases}
\end{array}$$

where concatenation is naturally extended over sets of expressions.
We define $\partial_w(E)$ for a word $w$ inductively with $\partial_\varepsilon(E) = \{E\}$ and $\partial_{wa}(E) = \partial_a(\partial_w(E))$, where $\partial_w(S) = \bigcup_{E \in S} \partial_w(E)$ when $S$ is a set of expressions.
Given a set of regular expressions $S$, we denote by $\mathcal{L}(S)$ the set $\bigcup_{E \in S} \mathcal{L}(E)$.

1. Give the partial derivatives of $(ab+b)^*ba$ by $a$ and $b$.

$$\begin{array}{rcl}
\partial_a((ab+b)^*ba) & = & (\partial_a((ab+b)^*) \cdot \{\ ba\ \}) \cup \partial_a(ba) \\
& = & \partial_a(ab+b) \cdot \{\ (ab+b)^*\ \} \cdot \{\ ba\ \} \cup \partial_a(b) \cdot \{\ a\ \} \\
& = & \{\ \partial_a(ab)\ ,\ \partial_a(b)\ \} \cdot \{\ (ab+b)^*ba\ \} \cup (\emptyset \cdot \{\ a\ \}) \\
& = & \{\ \partial_a(a) \cdot \{\ b\ \}\ ,\ \emptyset\ \} \cdot \{\ (ab+b)^*ba\ \} \\
& = & b(ab+b)^*ba
\end{array}$$

$$\begin{array}{rcl}
\partial_b((ab+b)^*ba) & = & (\partial_b((ab+b)^*) \cdot \{\ ba\ \}) \cup \partial_b(ba) \\
& = & \partial_b(ab+b) \cdot \{\ (ab+b)^*\ \} \cdot \{\ ba\ \} \cup \partial_b(b) \cdot \{\ a\ \} \\
& = & \{\ \partial_b(ab)\ ,\ \partial_b(b)\ \} \cdot \{\ (ab+b)^*ba\ \} \cup \{\ a\ \} \\
& = & \{\ \partial_b(a) \cdot \{\ b\ \}\ ,\ \varepsilon\ \} \cdot \{\ (ab+b)^*ba\ \} \\
& = & a + (ab+b)^*ba
\end{array}$$

2. Prove that for every $L, L' \subseteq \Sigma^*$ and $a \in \Sigma$,

$$a^{-1}(L \cup L') = (a^{-1}L) \cup (a^{-1}L')$$
$$a^{-1}L^* = (a^{-1}L) \cdot L^*$$
$$a^{-1}(L \cdot L') = \begin{cases} a^{-1}L \cdot L' & \text{si } \varepsilon \notin L \\ (a^{-1}L \cdot L') \cup (a^{-1}L') & \text{sinon} \end{cases}$$

3. Show that $\mathcal{L}(\partial_w(E)) = w^{-1}\mathcal{L}(E)$.

By induction on $w$, the definition of partial derivatives and the previous result.

4. We define the set of non empty suffixes of a word :

$$\mathrm{Suf}(w) = \{ \ v \in \Sigma^+ \ : \ \exists u, w = uv \ \}$$

Show that for every $w \in \Sigma^+$ :

$$\begin{aligned} \partial_w(E + E') &= \partial_w(E) \cup \partial_w(E') \\ \partial_w(E \cdot E') &\subseteq (\partial_w(E) \cdot E') \cup \bigcup_{v \in \mathrm{Suf}(w)} \partial_v(E') \\ \partial_w(E^*) &\subseteq \bigcup_{v \in \mathrm{Suf}(w)} \partial_v(E) \cdot E^* \end{aligned}$$

By induction on $w$.

5. Let $\|E\|$ be the number of occurences of letters of $\Sigma$ in $E$. Show that the set of partial derivatives different to $E$ has at most $\|E\| + 1$ elements.

By induction on $E$.
For more precision, see `https://core.ac.uk/download/pdf/81113752.pdf` from the end of page 305.

6. Conclude, and apply the construction to the expression $(ab + b)^*ba$.

## Exercise 4 : Closure by morphism

1. Let $h$ be the morphism $h(a) = 01$ and $h(b) = 0$. Give $h(a(a + b)^*)$.

$01(01 + 0)^*$

2. Apply the construction of closure by morphism to this example.
3. Let $h'$ be the morphism $h(0) = ab$, $h(1) = \varepsilon$. Give $h^{-1}(\{ \ abab, \ baba \ \})$.

1*01*01*

4. Apply the construction of closure by inverse morphism to this example.
5. Let $L = (00 \cup 1)^*$, $h(a) = 01$ and $h(b) = 10$. What is $h^{-1}(1001)$ ? $h^{-1}(010110)$ ? $h^{-1}(L)$ ? What is $h(h^{-1}(L))$, and is it related to $L$ ? Apply the construction by inverse morphism to this example.

$\{ \ ba \ \}$, $\{ \ aab \ \}$, $(ba)^*$, $(1001)^* \subsetneq L$.

## Exercise 5 : Characterizing recognizability

We want to show a converse to the pumping lemma. We say that a language $L$ satisfies $P_h$ if for all $uv_1 \dots v_h w$ avec $|v_i| \geq 1$, there exists $0 \leq j < k \leq h$ such that

$$uv_1 \dots v_h w \in L \Leftrightarrow uv_1 \dots v_j v_{k+1} \dots v_h w \in L.$$

The theorem of Ehrenfeucht, Parikh & Rozenberg states that $L$ is rational iff there exists $h$ such that $L$ satisfies $P_h$.

1. Show that if $L$ satisfies $P_h$, then $w^{-1}L$ also does for every word $w \in \Sigma^*$.
2. Let $h \in \mathbb{N}$. We want to show that the number of languages satisfying $P_h$ is finite. We use the following statement of Ramsey's theorem :

   > For every $k$ there is $N$ such that, for every set $E$ of cardinal greater than $N$ and every bipartition $\mathcal{P}$ of $\mathfrak{P}_2(E) = \{ \{e, e'\} : e, e' \in E,\ e \neq e' \}$, there exists a subset $F \subseteq E$ of cardinal $k$ such that $\mathfrak{P}_2(F)$ is contained in one of the classes of $\mathcal{P}$.

   Let $N$ be the natural number given by Ramsey's theorem for $k = h + 1$. Let $L$ and $L'$ be two languages satisfying $P_h$ and coinciding on words of size smaller than $N$. Prove that they coincide on words or size $M \geq N$, by induction on $M$. You may consider, for a word $f = a_1 \ldots a_N t$ of size $M$ (with $a_i \in \Sigma$), the following partition of $\mathfrak{P}_2([0; N])$ :

   $$X_f = \{ (j, k) \ : \ 0 \leq j < k \leq N,\ a_1 \ldots a_j a_{k+1} \ldots a_N t \in L \}$$

   $$Y_f = \mathfrak{P}_2([0; N]) \setminus X_f$$

   Conclude.
3. Conclude that if a language $L$ satisfies $P_h$ for some $h$, then $L$ is regular.

https://www.irif.fr/~carton/Enseignement/Complexite/ENS/Cours/pumping.html

**Exercise 6 : A rational slice ? (open exercise)**

Let $L$ be a rational language over a finite alphabet $\Sigma$. Is $\mathsf{FH}(L) = \{ f \in \Sigma^* : \exists h \in \Sigma^*.\ |h| = |f|,\ fh \in L \}$ rational ?