Words and Languages
(part III)

Prof. Dan A. Simovici

UMB
**Definition**

Let $A, B$ be two alphabets. A **substitution** from $A$ to $B$ is a mapping $s : A \rightarrow \mathcal{P}(B^*)$.

In other words, a substitution from $A$ to $B$ maps every symbol of $A$ to a language over $B$. 
A substitution $s$ from $A$ to $B$ is extended inductively to the mapping $s^* : A^* \rightarrow \mathcal{P}(B^*)$ by

\[
\begin{align*}
s^*(\lambda) & = \{\lambda\} \\
s^*(xa) & = s^*(x)s(a)
\end{align*}
\]

for every word $x \in A^*$.

By taking $x = \lambda$, the second equality gives $s^*(a) = s(a)$, so $s^*$ is indeed an extension of $s$. Further, we can verify by induction on the length of the word $y$ that $s^*(xy) = s^*(x)s^*(y)$ for all words $x, y \in A^*$. 
Example

Let $A = \{a_0, a_1\}$, $B = \{b_0, b_1, b_2\}$. Define the substitution $s$ by $s(a_0) = \{b_1\}^+$ and $s(a_1) = b_0\{b_2\}^+b_0$. The image of the word $a_0a_1a_0$ under $s^*$ is the language $\{b_1\}^+b_0\{b_2\}^+b_0\{b_1\}^+$. 
Further Extending Substitutions

When there is no risk of confusion, we frequently denote the extension of a substitution $s$ simply by $s$ rather than $s^*$. A substitution $s : A \rightarrow \mathcal{P}(B^*)$ can be further extended to a mapping between languages over $A$ and $B$ by taking

$$s(L) = \bigcup \{ s(x) \mid x \in L \}$$

for every language $L \subseteq A^*$. Then, we have the following result:

**Theorem**

For any languages $L_1, L_2$

$$s(L_1 \cup L_2) = s(L_1) \cup s(L_2)$$
$$s(L_1 L_2) = s(L_1)s(L_2)$$
$$s(L_1^*) = (s(L_1))^*$$
Definition

A morphism between $A^*$ and $B^*$ is a mapping $h : A^* \rightarrow B^*$ such that $h(xy) = h(x)h(y)$ for every $x, y \in A^*$.

If $h : A^* \rightarrow B^*$ is a morphism, then $h(\lambda) = \lambda$. Indeed, since $h(x) = h(x\lambda) = h(x)h(\lambda)$ the equality $h(\lambda) = \lambda$ follows immediately.
Definition

A morphism \( h : A^* \rightarrow B^* \) is

- **\( \lambda \)-free** if \( h(x) = \lambda \) implies \( x = \lambda \);
- **fine** if \( h(a) \in B \cup \{\lambda\} \) for every \( a \in A \);
- **very fine** if \( h(a) \in B \) for every \( a \in A \).
If \( h : A^* \rightarrow B^* \) is a morphism and \( K \subseteq B^* \), we denote by \( h^{-1}(K) \) the set \( \{ x \in A^* \mid h(x) \in K \} \). We refer to \( h^{-1}(K) \) as the inverse image of \( K \) under \( h \).
Example

Suppose, for instance, that $A = \{a_0, a_1\}$ and $B = \{b\}$ are two disjoint alphabets and $h : (A \cup B)^* \longrightarrow A^*$ is a morphism such that $h(a) = a$ for every $a \in A$ and $h(b) = \lambda$. Then, we have

$$h^{-1}(\{a_0^n a_1^n \mid n \in \mathbb{N}\}) \cap \{ba_0\}^* \{a_1 b\}^* = \{(ba_0)^n (a_1 b)^n \mid n \in \mathbb{N}\}$$
A morphism $h : A^* \rightarrow B^*$ is completely and uniquely defined by its values on the symbols of $A$.

Morphisms may be regarded as a special case of substitutions, where for each letter in $A$ the corresponding language has only one element.
Definition

Let $A$ be an alphabet and let $G, K$ be two languages over $A$. The shuffle of $G$ and $K$ is the language

$$\text{shuffle}(G, K) = \{x_0y_0x_1y_1\cdots x_{n-1}y_{n-1} \mid x_0x_1\cdots x_{n-1} \in G \text{ and } y_0y_1\cdots y_{n-1} \in K\}.$$
Theorem

There is an alphabet $B$ and there exist three morphisms $g$, $k$, $h$ from $B^*$ to $A^*$ such that $h$ is a very fine morphism, $g$, $k$ are fine morphisms and $\text{shuffle}(G, K) = h(g^{-1}(G) \cap k^{-1}(K))$. 
Proof

Let $B = A \cup A'$, where $A' = \{a' \mid a \in A\}$. Define the morphisms $h, g, k$ by

$h(a) = h(a') = a,$
$g(a) = a, \ g(a') = \lambda,$
$k(a) = \lambda, \text{ and } k(a') = a.$

Let $w \in h(g^{-1}(G) \cap k^{-1}(K))$. There is $y \in g^{-1}(G) \cap k^{-1}(K)$ such that $h(y) = w$. In addition, $g(y) \in G$ and $k(y) \in K$. Note that $g$ erases all primed symbols in $y$ while $k$ erases all non-primed symbols in $y$. Since $y$ is a mix of primed and non-primed symbols we can write

$y = u_0v'_0u_1v'_1\ldots u_{n-1}v'_{n-1},$ where $u_0, u_1, \ldots, u_{n-1}$ are the non-primed infixes of $y$ and $v'_0, v'_1, \ldots, v'_{n-1}$ are the primed infixes of $y$. Therefore, $g(y) = u_0u_1\cdots u_{n-1} \in G$, $k(y) = v_0v_1\cdots v_{n-1} \in K$, and $w = h(y) = u_0v_0u_1v_1\cdots u_{n-1}v_{n-1}$. Thus, $w \in \text{shuffle}(G, K)$. 
Conversely, if \( w \in \text{shuffle}(G, K) \) we can write \( w = u_0 v_0 u_1 v_1 \cdots u_{n-1} v_{n-1} \), where \( u_0 u_1 \cdots u_{n-1} \in G \) and \( v_0 v_1 \cdots v_{n-1} \in K \). Let \( y \) be the word \( y = u_0 v'_0 u_1 v'_1 \cdots u_{n-1} v'_{n-1} \), where \( v'_i \) is the primed version of \( v_i \) for \( 0 \leq i \leq n - 1 \). It is easy to see that \( h(y) = w \) and that

\[
\begin{align*}
g(y) &= u_0 u_1 \cdots u_{n-1} \in G \\
k(y) &= v_0 v_1 \cdots v_{n-1} \in K.
\end{align*}
\]

Therefore, \( y \in g^{-1}(G) \cap k^{-1}(K) \), so \( w \in h(g^{-1}(G) \cap k^{-1}(K)) \), which concludes our argument.