

## AMBIGUITY

**Definition 1.** For a language  $L$  over  $\Sigma$ , define

$$\text{suff}(L) = \{p \in \Sigma^+ \mid \exists u \in L : up \in L\}$$

$$\text{pref}(L) = \{p \in \Sigma^+ \mid \exists v \in L : pv \in L\}$$

Note that  $\text{suff}(L)$  and  $\text{pref}(L)$  do not contain  $\epsilon$ . Note also that we only take prefixes and suffixes that can be expanded to a word in  $L$  with another word in  $L$ , not any word in  $\Sigma^*$ .

The set  $\text{suff}(L)$  is identical to the left quotient  $L^{-1}L$  of  $L$  by itself. Similarly,  $\text{pref}(L) = LL^{-1}$ .

**Definition 2.** Two languages  $L_1$  and  $L_2$  are *unambiguously concatenable*, written  $L_1 \cdot^! L_2$ , iff for every  $u_1, v_1 \in L_1$  and  $u_2, v_2 \in L_2$ , if  $u_1 u_2 = v_1 v_2$  then  $u_1 = v_1$  and  $u_2 = v_2$ .

**Lemma 1.** *The languages  $L_1$  and  $L_2$  are unambiguously concatenable if and only if  $\text{suff}(L_1) \cap \text{pref}(L_2) = \emptyset$ .*

*Proof.* Let  $L_1$  and  $L_2$  not unambiguously concatenable, i.e. there exist  $u_1, v_1$  in  $L_1$  and  $u_2, v_2$  in  $L_2$  such that  $u_1 u_2 = v_1 v_2$  and  $u_1 \neq v_1$  and therefore  $u_2 \neq v_2$ . Assume  $u_1$  is strictly shorter than  $v_1$ , therefore  $v_1 = u_1 p, p \neq \epsilon$ . With that  $v_1 v_2 = u_1 p v_2$  and  $u_2 = p v_2$  and  $p \in \text{suff}(L_1) \cap \text{pref}(L_2)$

Let  $p \in \text{suff}(L_1) \cap \text{pref}(L_2)$ . That implies that there are  $u \in L_1$  and  $v \in L_2$  such that  $up \in L_1$  and  $pv \in L_2$ . The word  $upv$  can be split as  $up \cdot v$  and as  $u \cdot pv$  and therefore  $L_1$  and  $L_2$  are not unambiguously concatenable.  $\square$

To ease notation we write  $L_{\setminus \epsilon}$  for  $L \setminus \{\epsilon\}$ .

**Definition 3.** A language  $L$  is *unambiguously iterable*, written  $L^{!*}$ , iff for every  $u_1, \dots, u_m, v_1, \dots, v_n \in L_{\setminus \epsilon}$  with  $u_1 \cdots u_m = v_1 \cdots v_n$ ,  $m = n$  and  $u_i = v_i$ .

It is very important that we exclude  $\epsilon$  in the definition of unambiguous iteration, since otherwise *any* language  $L$  that contains  $\epsilon$  is trivially not unambiguously iterable, even though that presents no problems for our purposes, since we never split  $\epsilon$  into more than one word.

**Lemma 2.** *The language  $L$  is unambiguously iterable if and only if  $L_{\setminus \epsilon}$  and  $L^*_{\setminus \epsilon}$  are unambiguously concatenable.*

*Proof.* Let  $L$  be unambiguously iterable, and assume that there is  $u_1 v_1 = u_2 v_2$  with  $u_i \in L_{\setminus \epsilon}$  and  $v_i \in L^*_{\setminus \epsilon}$  and  $u_1 \neq u_2$ . This contradicts  $L^{!*}$  since  $L_{\setminus \epsilon} \cdot L^*_{\setminus \epsilon} \subseteq L^*$ .

Let  $L_{\setminus \epsilon} \cdot^! L^*_{\setminus \epsilon}$  and assume there are  $u_1, \dots, u_m, v_1, \dots, v_n \in L_{\setminus \epsilon}$  with  $u_1 \cdots u_m = v_1 \cdots v_n$ , but there is an  $i \leq \min(m, n)$  with  $u_i \neq v_i$ . We can assume that  $u_1 \neq v_1$ <sup>1</sup>, and therefore  $\min(m, n) \geq 2$ . Since  $u_1$  and  $v_1$  are in  $L_{\setminus \epsilon}$ , we have an ambiguous split of a word from  $L_{\setminus \epsilon} \cdot L^*_{\setminus \epsilon}$ , contradicting the initial assumption.  $\square$

**Example.** Using regular expression notation, the language  $L = (a|b|c|abc)$  is unambiguously concatenable with itself, but not unambiguously iterable.  $L^* = (a|b|c)^*$  and the word  $abcabc$  can be split into  $abc \cdot abc$  and  $a \cdot bcabc$ . This example shows that  $L \cdot^! L$  does not imply  $L^{!*}$ , but as the previous lemma showed,  $L_{\setminus \epsilon} \cdot^! L^*_{\setminus \epsilon}$  does.

If you have  $P = \text{suff}(L_1) \cap \text{pref}(L_2)$ , how do you generate a word that is ambiguous?

---

<sup>1</sup>if  $u_1 = v_1$ , we simply use  $u_2 \cdots u_m$  and  $v_2 \cdots v_n$  for this argument