Journal of Automata, Languages and Combinatorics **7** (2002) 1, 143–157

REGULAR NESTS

HUEI-JAN SHYR
Department of Applied Mathematics, National Chung-Hsing University
Taichung, Taiwan 402
e-mail: hjshyr@flower.amath.nchu.edu.tw

ABSTRACT

A language over a finite alphabet X is regular if it can be accepted by a finite automaton. It is known that the family of all regular languages $\mathcal{L}_3$ over X is closed under catenation. This paper is a study of the regular property of the catenation of two languages. We call a family of languages $\mathcal{L}$ a regular nest if the catenation of any two languages in $\mathcal{L}$ is regular. We construct some natural regular nests which contain non-regular languages. We also investigate some languages which catenate with any non-empty languages on the right (left) can never be regular. Such languages we call them absolutely right (left) non-regular languages. We give an example of absolutely right non-regular language which is not an absolutely left non-regular. We show that for $X = \{a, b, c\}$, the well known context-sensitive language $\{a^n b^n c^n \mid n \geq 1\}$, and the case when $|X| = 2$, Dyck language D and the balanced language H are absolutely right non-regular languages.

Keywords: Bifix codes, regular nests, absolutely non-regular languages

1. Introduction and Preliminaries

In this paper, we let X^* be the free monoid generated by a finite alphabet X containing more than one letter. Every element of X^* is called a *word* and every subset of X^* is a *language*. Let 1 be the empty word and let $X^+ = X^* \setminus \{1\}$. For a word $u \in X^*$ we let $\lg(u)$ be the length of the word, that is, the number of the letters occurred in the word u. The catenation of two languages $A, B \subseteq X^*$ is the set $AB = \{xy \mid x \in A,\ y \in B\}$. In formal language theory we call a language $L \subseteq X^*$ a *regular language* if it can be accepted by a finite automaton (see [3]). The family of regular languages play a very important role in formal language theory. One of the important properties of the family of regular languages is closed under the catenation. By using the family of bifix codes we are able to construct a new family of languages in which the catenation of any two languages in the family is regular. Now we call a non-empty language $L \subset X^+$ a *prefix code* (*suffix code*) if $L \cap LX^+ = \emptyset$ ($L \cap X^+L = \emptyset$). By a *bifix code* L we mean that L is a prefix code and also a suffix code. We will consider $\{1\}$ as a bifix code (see [6]). For a language $A \subseteq X^+$, let $X_A^+ = X^+ \setminus A$. We are interested in the language X_A^+, where A is a prefix code, a suffix code or a bifix code. If L is a singleton, $L = \{u\}$, we represent X_L^+ by X_u^+. The regular property of the catenation