Factorization of prefix-closed subsets of words

Véronique Bruyère
Faculté des Sciences
15, avenue Maistriau
B-7000-Mons (Belgium)

A set of words is said to be prefix-closed if it contains the left factors of its elements. We study the factorizations of finite prefix-closed sets of words P of the following form

$$P = R.S$$

where the product is unambiguous (i.e. any word $p \in RS$ has only one factorization $p=rs$ with $r \in R$ and $s \in S$).
Such factorizations appear in the study of non-biprefix and asynchronous codes [3].

A subset of words, distinct of {1}, is a prefix code if no element of X is a proper left factor of another word in X. X is a maximal prefix code if it is a prefix code and if it is not included in another prefix code over the same alphabet.
There exists a 1 to 1 correspondence between the set of finite maximal prefix codes and the one of nonempty finite prefix-closed sets [1].
Let Y and Z be prefix codes respectively over the alphabets B and A, such that the elements of Z are in bijection with the letters of the words in Y. These codes can be composed to obtain a new prefix code X over A: X is constructed from Y, where each letter of the words in Y are replaced by the associated word in Z.

By means of the 1 to 1 correspondence and the composition operation over codes, it is not very hard to construct an unambiguous product which is finite and prefix-closed: the composition of two finite maximal prefix codes gives rise to a new finite maximal prefix code X of which the associated prefix-closed set P is equal to an unambiguous product P=RS with S itself prefix-closed.
The converse of this result is generally false, except if the second factor of P=RS is prefix-closed: if a product RS is unambiguous, finite and prefix-closed, with S also prefix-closed, then the finite maximal prefix code corresponding to RS can be obtained by the composition of two finite maximal prefix codes.

We then study a particular family of unambiguous products RS which are finite and prefix-closed, where the

second factor S is no more prefix-closed. We suppose that the words that lack S to be prefix-closed are all left factor of a power of a word t, where t is the smallest of these words.
We use a result of Krasner [2] to obtain the complete characterization of such products. In his paper, Krasner gives the form of the polynomials with positive real coefficients, which factorize

$$1 + X + X^2 + \ldots + X^n$$

This particular family of prefix-closed products include the examples considered by D. Perrin in the construction of non-biprefix, asynchronous codes.
We then establish suffisant conditions under which these products RS can be factorized into R'S' with S' prefix-closed.

So, for two particular families of products RS, the form of the two factors R and S has been completely described. However, there exist examples of products that do not belong to these families.
It seems that general unambiguous factorizations of finite prefix-closed sets is still unknown.

[1] Berstel J., Perrin D., "Theory of codes", Academic Press, 1985.

[2] Krasner M., Ranulac B., "Sur une propriété des polynômes de la division du cercle", C.R. Acad. Sc. Paris 240, 1937, 397-399.

[3] Perrin D., "Codes asynchrones", Bull. Soc. Math. France 105, 1977, 385-404.