1991

Codes Andn-ary Relations

Shyr-shen Yu

Codes and n-ary Relations

by

S. S. Yu

Department of Computer Science

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Graduate Studies
The University of Western Ontario
London, Ontario

September, 1990

ISBN 0-315-64274-2

ABSTRACT

The aim of this thesis is to develop a general mechanism for the construction of codes and to extract general properties of classes of codes. This mechanism makes it unnecessary to study various classes of codes separately — at least to some extent — by different constructions and properties.

To achieve this goal, the mechanism of characterizing classes of languages by binary relations is studied. Some general properties related to binary relations and languages are obtained. Moreover, three new classes of codes, n-shuffle codes, solid codes, and intercodes are constructed. Solid codes and intercodes have the synchronous decoding property which is very useful in the design of circuits of coders and decoders.

The studies of codes, n-codes, and intercodes indicate that these three classes of codes cannot be characterized by binary relations. We introduce a more general mechanism, that is, to characterize classes of languages by finitary relations. This mechanism can be used to characterize more classes of languages, such as the classes of n-codes and intercodes. Sometimes, it is difficult to show inclusion relations between classes of codes and hierarchy properties of classes of codes. Results derived in this thesis provide a mechanism which can simplify this task.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Professor H. Jürgensen, for his help in completing this thesis. Through his extensive knowledge and very useful suggestions, he has extended my comprehension of the subject and taught me how to describe a theory with depth and clarity. I would also like to express my deep gratitude to my supervisor, Professor G. Thierrin, for his invaluable guidance, support, and encouragement throughout my studies at The University of Western Ontario.

Special thanks are due to Dr. H.J. Shyr for his guidance. With his inspiring teaching and meticulous training he laid a solid groundwork for my research. Many of my ideas came from doing research with him. I also want to thank Professor Tom Head, Professor A.M. Dawes, Professor A. Szilard and Professor S. Yu for their rewarding corrections and suggestions.

Finally, I wish to express my sincere gratitude to my dear wife and parents for their patience and help throughout my studies.

TABLE OF CONTENTS

CHAPTER 1

Preliminaries

1.1 Introduction

A language is a *code* if every finite string can be decomposed into words taken from this language in at most one way. This *unique decipherability* of codes is very important in the application in information communication systems. Several code testing methods can be found in [Sal1]. While code testing is difficult in general, there are many abstract properties of languages that define classes of codes directly. Many classes of codes, such as uniform, infix, bifix, prefix, suffix codes, etc., have been constructed in this way (see [Ber2], [Sal1] and [Shy1] for examples).*

To develop a universal mechanism for characterizing classes of codes and extracting common properties of codes is a very important issue in the theory of codes and it is the aim of this thesis. This work provides a general tool to study common properties of codes and to avoid studying different classes of codes by different constructions and properties. To achieve this goal, the mechanism of characterizing classes of languages by binary relations is studied. Moreover, the relationship between different classes of codes, and constructions and properties of some new classes of codes are investigated.

In 1952, Higman investigated the connections between languages and sets which are independent with respect to some binary relations ([Hig1]). The characterizations of several classes of codes, such as hypercodes, uniform codes, infix codes, bifix codes, prefix codes and suffix codes, as the independent sets with respect to binary relations can be found in [Hai1], [Jul1], [Shy4], [Shy5], [Thi2] and [Val1].

* In this introduction we give a general survey of our results in the context of related research. Precise definitions of terms used are found in the literature or in Section 1.2.

Shyr provided a survey of this early work in [Shy1]. After that, the connections between independent sets and codes have been extensively investigated (see [Day1], [Ito2], [Ito3], [Ito4], [Jür1], [Jür4], [Jür5], [Pet1] and [Thi3]).

In most of that work, partial orderings provide characterizations of certain classes of codes. For instance, in [Jür1], a binary relation $\leq_{pi}$ on X^* is defined by:

$$u \leq_{pi} v \iff u = v \text{ or } v = xuy \text{ for some } x \in X^* \text{ and } y \in X^+.$$

It has been shown that the binary relation $\leq_{pi}$ is a partial order and the class of p-infix codes is equivalent to the $\leq_{pi}$-independent sets.

However, it was also found that some classes of codes, like the classes of outfix codes ([Jür5]) and n-shuffle codes ([Thi3]), cannot be characterized as independent sets with respect to any partial order. Instead of partial orders, strict binary relations are considered (see [Day1], [Ito3], [Jür5], [Shy4] and [Thi3] for examples). Furthermore, in our research we have found that the length-preserving condition of strict binary relations is not necessary in several cases. Only the reflexivity and symmetry conditions will affect the relationship between binary relations and their independent sets. Instead of considering strict binary relations, we consider reflexive and symmetric closures of binary relations. The consideration of reflexive and symmetric closures of binary relations provides a fundamental step for a further generalization to develop a universal mechanism.

A counterexample for the binary relation mechanism is the class of n-codes. n-codes have been introduced by Shyr and Thierrin in [Shy4]. Ito, Jürgensen, Shyr and Thierrin further investigated the hierarchy of n-codes and codes in 1987 and 1989 ([Ito2], [Ito3] and [Ito4]). The hierarchy of n-codes is as follows:

$$\mathcal{C} \subseteq \ldots \subseteq \mathcal{C}_n \subseteq \mathcal{C}_{n-1} \subseteq \ldots \subseteq \mathcal{C}_2 \subseteq \mathcal{C}_1$$

where $\mathcal{C}$ and $\mathcal{C}_n$ are classes of codes and n-codes, respectively. It has been shown that n-codes for $n > 2$ cannot be expressed as sets being independent with respect to any binary relation.

In our recent research, we construct another counterexample, the classes of intercodes ([Jür2] and [Shy7]). A language is an *intercode of index* m if every sentence consisting of m words taken from this language can only be a left subsentence or a right subsentence of other sentences consisting of $m+1$ words of this language. The hierarchy of intercodes is as follows:

$$\mathcal{I}_1 \subseteq \mathcal{I}_2 \subseteq \ldots \subseteq \mathcal{I}_m \subseteq \mathcal{I}_{m+1} \subseteq \ldots \subseteq \mathcal{I}$$

where $\mathcal{I}$ and $\mathcal{I}_m$ are classes of intercodes and intercodes of index m, respectively. Again, intercodes cannot be expressed as sets which are independent with respect to any binary relations.

These results provide the motivation to consider a more general mechanism to cover classes of n-codes and intercodes. The mechanism, expressing the classes of languages as the independent sets with respect to finitary relations, is introduced in [Jür2]. This finitary relation mechanism succeeds in expressing classes of n-codes and intercodes as the independent sets with respect to some finitary relations.

Two important results are the gap theorem and the inclusion theorem. The gap theorem provides a powerful tool to determine the impossibility of characterizing certain classes of languages by finitary relations. In 1975, Shyr and Thierrin showed that there is no strict binary relation such that the class of all independent sets is exactly the class of all codes ([Shy4]). This result now can be modified and shown by the gap theorem as that there is no finitary relation such that the class of all independent sets is exactly the class of all codes. The inclusion theorem provides a tool to show the hierarchy of classes of codes by the hierarchy of finitary relations. Sometimes, to show the hierarchy of relations is much easier than to show the hierarchy of codes directly.

A diagram of the relationship between certain various classes of codes and n-codes is provided in [Ito4]. This diagram is shown as Figure 1 below. Three classes of codes, n-shuffle codes, solid codes, and intercodes, which do not appear in the diagram of Figure 1, are constructed and studied in our research. These additional classes of codes are shown in the redrawn diagram in Figure 2 in further below. In addition to the main subject for developing a universal mechanism to characterize classes of codes, we also study constructions and properties of n-shuffle codes, solid codes, and intercodes.

The infix order is contained in the embedding order and between these two relations there is an infinite hierarchy of reflexive and antisymmetric binary relations called the n-shuffle relations (see [Thi3]). These relations fill the gap between the infix order and the embedding order. n-shuffle relations are binary relations but not partial orders. It is shown that the n-shuffle relations are not compatible for $n \geq 1$ and not transitive for $n \geq 2$. Moreover, the transitive closure of each of them is the embedding order. The corresponding independent sets form the classes of n-shuffle codes. The classes of n-shuffle codes are submonoids of the free monoid of prefix codes. But they are not free submonoids.

The unique decipherability condition does not address the case when a word does not have a decomposition over L. For instance, such a case may arise when the word in question, that is, the encoded message received over a noisy channel, contains inserted noise symbols or has some symbols missing. To some extent, the definition of solid codes attempts to guarantee unique decipherability in such cases too.

Solid codes were introduced in [Shy6] and studied further in [Jür1] and [Rei1]. Intuitively speaking, a set L of words is a solid code if and only if every word has a unique factorization into words in L and words which are not in X^*LX^*.

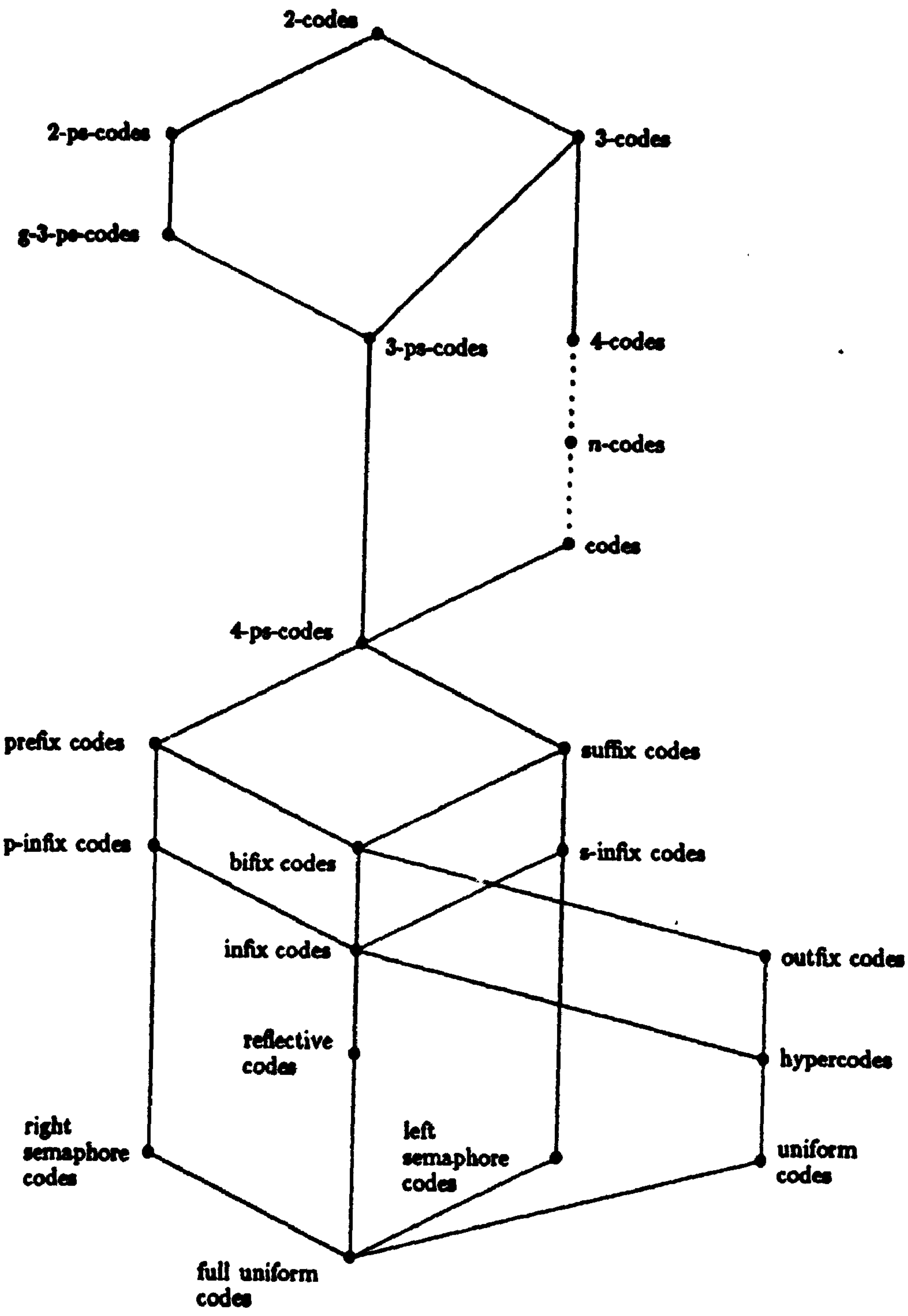

Figure 1. Relation between families of codes or n-codes: Lines indicate (known) proper inclusion. Dotted lines indicate hierarchies. This diagram does not indicate intersections and unions of families.

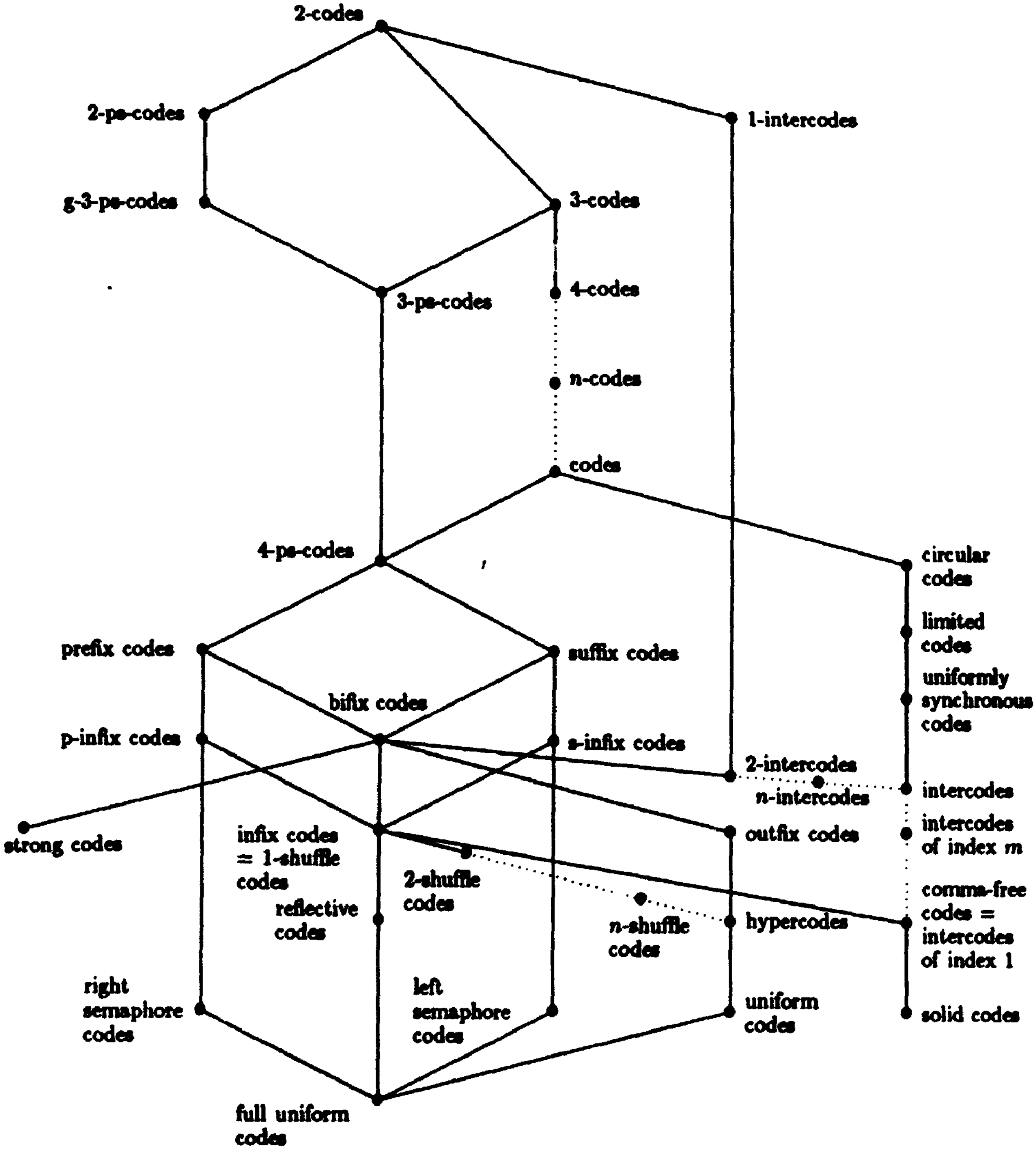

Figure 2. Relation between families of codes or n-codes: Lines indicate (known) proper inclusion. Dotted lines indicate hierarchies. This diagram does not indicate intersections and unions of families.

Solid codes may be useful as cryptographic techniques. Using a solid code L as the coding method, in the transmission of a sequence $y_1, y_2, \ldots, y_n, \ldots$ (with $y_i \in L$), we choose an arbitrary sequence of words $x_1, x_2, \ldots, x_n, x_{n+1}, \ldots$ such that $y_i \notin E(x_j)$ for all y_i and x_j where $E(x) = \{u \mid u \in X^+ \text{ with } u \leq_i x\}$ for $x \in X^+$. Then the sequence $x_1 y_1 x_2 y_2 \ldots x_n y_n x_{n+1} \ldots$ can be decoded correctly. This encryption method assumes that L is known to the sender and the receiver, and unknown to everybody else.

Some properties of solid codes with cardinality less than 3 were first studied in [Shy6]. Let X be an alphabet and for $x \in X^+$, let $P(x)$, $S(x)$ and $E(x)$ denote the sets of proper prefixes of x, of proper suffixes of x, and of all subwords of x, respectively. Then, the following characterization is a result of [Shy6]. For $u, v \in X^+$, $\{u, v\}$ is a solid code if and only if $P(z_1) \cap S(z_2) = \emptyset$ for any $z_1, z_2 \in \{u, v\}$ and $u \in E(v)$ or $v \in E(u)$ implies $u = v$. From this characterization, the class of solid codes forms a subclass of the class of infix codes. The general case of solid codes of arbitrary cardinality is characterized as follows: A language L is a solid code if and only if every subset of L with at most two elements is a solid code ([Jür3]).

Since the publication in 1957 of the paper "Codes without commas" ([Cri1]), comma-free codes have been extensively investigated (see [Ber2], [Gol1], [Gol2], [Hsi1] and [Shy1] for examples). A comma-free code L has the property that for a message x in L^+, that is $x = u_1 u_2 \ldots u_n$ for $u_i \in L$, if we decode this message and find a factor y of x in L, then y is one term of u_i. Comma-free codes have a very important property, that is, the synchronization delay of a comma-free code is just one. A word in a message consisting of words taken from a comma-free code can be identified when the last symbol of this word is received.

In [Gol1], comma-free codes were defined as bifix codes with the synchronization delay equal to one. In [Sch3], Schützenberger introduced the notion of limited codes. A characterization of comma-free codes has been given by using limited codes (see [Ber2]). Recently, in [Hsi1], Hsieh, Hsu and Shyr gave a characterization of comma-free codes by using the intersection of sets: A language $L \subseteq X^+$ is a comma-free code if and only if $L^2 \cap X^+LX^+ = \emptyset$.

The intercodes introduced in [Shy7] and already mentioned above are a class of generalized comma-free codes. A language $L \subseteq X^+$ is an *intercode* if and only if $L^{m+1} \cap X^+L^mX^+ = \emptyset$ for some $m \geq 1$. The class of intercodes is a subclass of the bifix codes. Some properties and operations of intercodes and the relationship between intercodes and comma-free codes are also studied in [Shy7] and [Yu1]. It has been shown that every intercode is limited and circular. In particular, for an infix code L, L is an intercode if and only if L is a (p, q)-limited code. A thorough study of the hierarchy of intercodes and n-intercodes can be found in [Jür2]. The diagram in Figure 3 below illustrates this hierarchy.

The maximality and the decidability of several classes of codes are also studied in this thesis. For some properties of maximal (prefix) codes, see [Ber2], [Cha1] and [Per1]. It is not difficult to show that every code is contained in a maximal code. In [Szi1], we also have the result that every prefix code is contained in a maximal prefix code. In [Shy5], Shyr and Thierrin showed that a maximal code which is also a hypercode is a maximal prefix code. Some properties of the maximality of hypercodes have also been investigated in [Shy5]. For the properties of the maximality of bifix codes see [Sch5]. In [Sch4], Schützenberger showed that every semaphore code is a maximal code. Some properties of maximal n-codes have been studied in [Ito2]. In [Guo1], some properties of maximal infix codes have been mentioned. In [Ito4], some properties of the maximal p-infix codes, s-infix codes, etc., have been studied. The maximality of solid codes has been studied in [Jür3]. Some problems of the maximality of comma-free codes have been investigated in [Jig1] and [Gol2]. However several problems about the maximality of comma-free codes and all the

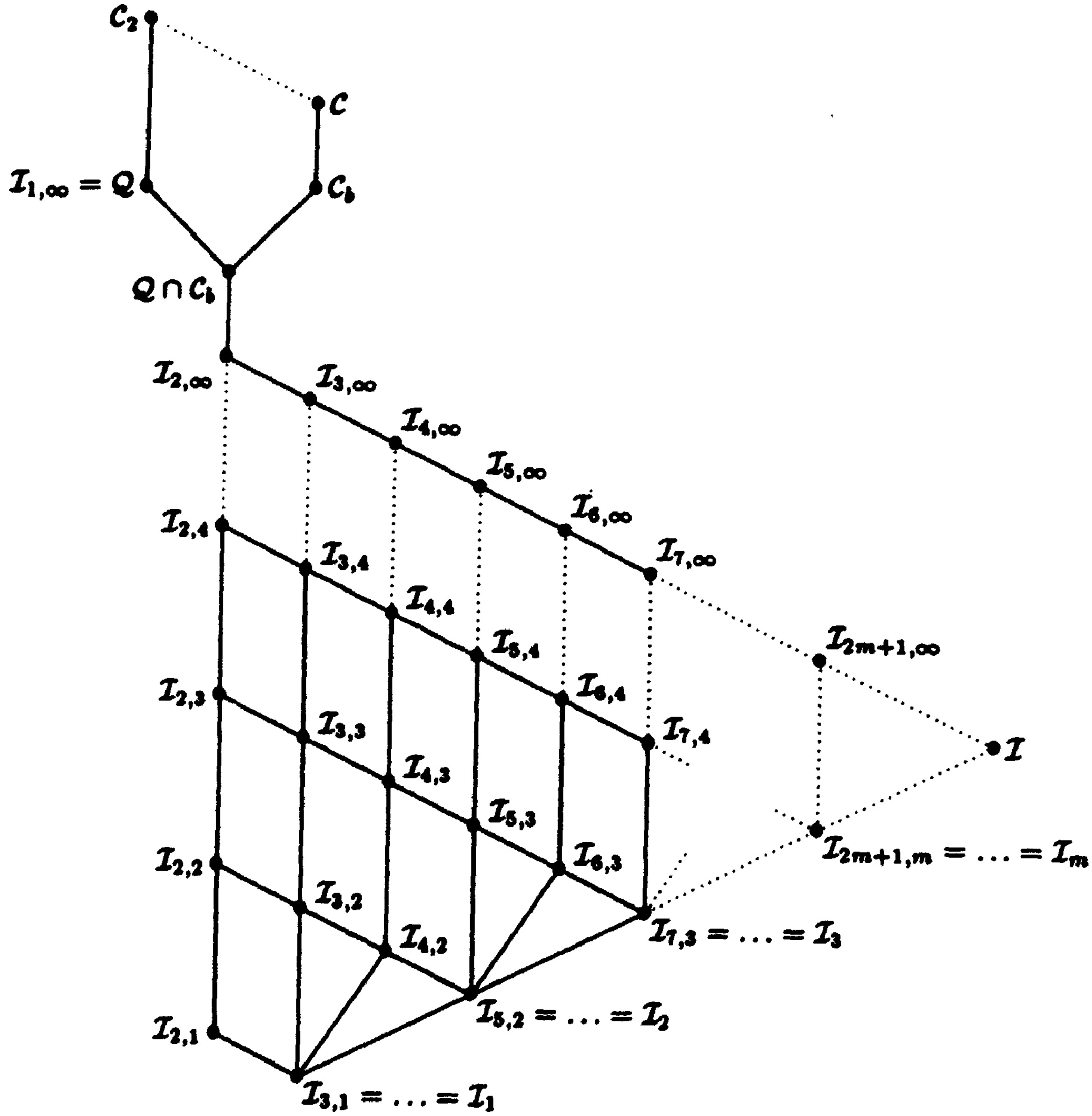

Figure 3. The intercode hierarchy

problems about the maximality of intercodes are still open.

In [Ito4], Ito, Jürgensen, Shyr and Thierrin showed that whether a regular language is in the class of right semaphore codes, p-infix codes, s-infix codes, infix codes or outfix codes is decidable. The decidability problem of solid codes has been studied in [Jür3]. The decidability problems of intercodes are still open.

The organization of this thesis is as follows:

Chapter 1 is the preface including the introduction, fundamental definitions, notation which is used in this thesis, and a summary of the main results of this thesis.

Chapter 2 contains several subtopics of the relationship between some classes of codes and binary relations. In Section 2.1, we investigate the connections between binary relations and the independent sets and extract some fundamental properties of binary relations and classes of related languages. p-infix codes can be characterized as the independent sets with respect to a binary relation $\leq_{pi}$. Some properties of p-infix codes are studied in Section 2.2. The purpose of Section 2.3 is to study the hierarchy of n-shuffle relations. n-shuffle codes are characterized as the independent sets with respect to n-shuffle relations. In Section 2.3, some properties of n-shuffle relations and of the syntactic monoid of n-shuffle codes are also investigated.

Chapter 3 deals with the properties of solid codes. The definitions and notations can be found in Section 3.1. In Section 3.1, we give a characterization of solid codes which is decidable for regular languages. In Section 3.2, we establish closure and non-closure properties of solid codes. The maximality of solid codes is investigated in Section 3.3. Section 3.4 mentions some results of further investigations of solid codes (ref. [Shy6] and [Rei1]).

Chapter 4 is devoted to the study of intercodes. First, in Section 4.1, we give the following characterization of intercodes: A language is an intercode if and only if it is a synchronously decipherable code. In particular, an infix code is an intercode if and only if it is a limited code. n-intercodes are studied in Section 4.2. In Section 4.3, we show that the class of comma-free codes is exactly the class of all intercodes of index 1. This result provides an abstract mathematical means to study the properties of comma-free codes without checking the so-called comma-free dictionary to impose restrictions on the set of words. We also characterize the class of comma-free codes as the class of infix (p, q)-limited codes with $p + q = 3$. In Section 4.4, some properties of 2-comma-free codes are investigated. The same closure and non-closure properties of the class of solid codes are established of the class of 2-comma-free codes. A thorough study of the hierarchy of intercodes and n-intercodes by using the hierarchy of certain finitary relations is later provided in Section 5.4.

Chapter 5 deals with the mechanism for characterizing classes of languages as sets being independent with respect to certain finitary relations. Some fundamental properties of finitary relations and independent sets are investigated in Section 5.1. The gap theorem is shown in this section. In Section 5.2, we discuss how the earlier treatment of binary relations and independent sets fits into the new framework. The binary relations given before are redefined and listed in Table 5.1 in Section 5.2. This shows that the generalization from binary relations to finitary relations is possible in such a way that all results concerning binary relations of the earlier work are preserved. To show that the new mechanism is more powerful, a binary relation ω_σ is defined such that the class of ω_σ-independent sets is the class of solid codes. The purpose of Section 5.3 is to discuss the relationship between classes of sets being independent with respect to finitary relations of different arities. In Section 5.4, results obtained in the previous sections of this chapter are applied to the cases of n-codes, n-intercodes, and n-ps-languages. We strengthen the hierarchy results obtained in [Ito2] and add to those results obtained in [Shy7]. This chapter completes the goal of this thesis.

Chapter 6 is the conclusion of this thesis. In this thesis we discover general concepts, constructions, and results which not only extend the hierarchy of codes and n-codes, but also generalize results of binary relations obtained before.

Given the results obtained in this thesis, it is now clear that it is not so much relations, but rather dependence systems that ought to be considered for defining classes of codes. These connections are being investigated in [Jür6].

1.2 Fundamental Definitions and Notations

In this section we give some fundamental definitions and notations. Items not defined in this section or in the following chapters can be found in the books [Ber2], [Gin1], [Grä1], [Hop1], [Lal1], and [Woo1] which we use as standard references.

1.2.1 Semigroups

An *alphabet* X is a nonempty, finite set. In this thesis we always assume that $2 \leq |X|$. Any finite string $x = x_1x_2\cdots x_n$ where $x_i \in X$, $i = 1, 2, \ldots, n$, is called a *word*. The empty word will be denoted by 1. Let X^* be the set of all words over X and let $X^+ = X^* \setminus \{1\}$. For any word $x \in X^*$, the length of x is denoted by $|x|$. In particular, $|1| = 0$. Any subset L of X^* is called a *language* over X and the cardinality of L is denoted by $|L|$.

The *multiplication of two words* x and y of X^* is defined as the juxtaposition of x and y, that is, the *catenation*. For any two sets A and B, the *catenation* of A and B is the set

$$AB = \{xy \mid x \in A,\ y \in B\}.$$

In the sequel, $A^0 = \{1\}$ and, inductively, $A^n = A^{n-1}A = AA^{n-1}$.

A nonempty subset S of a monoid M is called a *base* of M if and only if

$$a_1a_2\cdots a_n = b_1b_2\cdots b_m, \text{ for } a_i, b_j \in S \text{ implies } n = m \text{ and } a_i = b_i,\ i = 1, 2, \ldots, n.$$

A monoid M is *free* if and only there exists a base S such that $S^* = M$. To test whether or not a monoid is free is difficult. But there is an easier method to test whether a submonoid of a free monoid is free or not.

For any subsets A and B of M, we define

$$A^{-1}B = \{x \in M \mid Ax \cap B \neq \emptyset\};$$

$$BA^{-1} = \{x \in M \mid xA \cap B \neq \emptyset\}.$$

Theorem 1.1 ([Sch2]) *Let M be a free monoid and let S be a submonoid of M. Then S is free if and only if $S^{-1}S \cap SS^{-1} \subseteq S$.*

A word $u \in X^+$ is said to be *primitive* if it is not a power of another word, that is, u is primitive if and only if $u = f^n$ with $f \in X^+$ implies $n = 1$. Let Q be the set of all primitive words. Let $\mathcal{Q}$ denote the family of non-empty subsets of Q.

It is well-known that, for every word $u \neq 1$, there exists a unique primitive word f and a unique integer $k \geq 1$ such that $u = f^k$ ([Lyn1]). For any word u let $\sqrt{u}$ be this unique primitive word such that u is a power of $\sqrt{u}$. And, for any $L \subseteq X^*$, let $\sqrt{L} = \{\sqrt{u} \mid u \in L\}$. A language L is a 2-code if and only if $L \in \mathcal{Q}$ (see [Shy4]).

A word $u \in X^+$ is said to be *unbordered* if $u \in vX^+ \cap X^+v$ implies $v = 1$. Of course, an unbordered word is primitive. But a primitive word need not be unbordered. For example, let $X = \{a, b\}$. Then aba is primitive but not unbordered.

1.2.2 Codes and n-Codes

For a language $L \subseteq X^*$ and a word $x \in L^*$, an *L-factorization of x* is a sequence $(x_1, x_2, \ldots, x_n)$ of words in L for some $n \geq 0$ such that $x = x_1x_2\ldots x_n$. A nonempty language $L \subseteq X^+$ is called a *code* if every word $x \in L^+$ has a unique L-factorization, that is, $x_1x_2\ldots x_m = y_1y_2\ldots y_n$ with $x_i, y_j \in L$ for all i and j implies $n = m$ and $x_i = y_i$ for all i.

A language $L \subseteq X^+$ is called an *n-code* if every subset of L with at most n elements is a code (see [Shy4]). It is trivial that every $(n+1)$-code is an n-code and every code L with $|L| = m$ is an n-code for $n \leq m$. Thus we have the following hierarchy of n-codes:

$$\mathcal{C} \subseteq \ldots \subseteq \mathcal{C}_n \subseteq \mathcal{C}_{n-1} \subseteq \ldots \subseteq \mathcal{C}_2 \subseteq \mathcal{C}_1$$

Since $\{u\}$ is a code for every word $u \in X^+$, $\mathcal{C}_1$ is equivalent to $2^{X^+} \setminus \{\emptyset\}$, that is every non-empty language in X^+ is an 1-code. Moreover, it was proved in [Ito2] and [Shy4] that the hierarchy of n-codes is strict, that is, there exist $(n-1)$-codes which are not n-codes. Of course an n-code need not be a code. For example, let $X = \{a, b\}$ and let $L = \{a, b, ab\}$. Then L is a 2-code but not a code.

Many classes of codes have been studied. We give some definitions.

Definition: 1.2 *Let X be an alphabet and let $L \subseteq X^+$. Then L is*

(1) a prefix code if $L \cap LX^+ = \emptyset$;

(2) a suffix code if $L \cap X^+L = \emptyset$;

(3) a bifix code if it is both a prefix code and a suffix code;

(4) an infix code if for every $x \in L$, $X^*xX^* \cap L = \{x\}$;

(5) a uniform code if for every $x, y \in L$, $|x| = |y|$;

(6) a full uniform code if $L = X^n$ for some $n \geq 1$;

(7) an outfix code if for all $x, y, z \in X^*$, $xy \in L$ and $xzy \in L$ imply that $z = 1$;

(8) a right semaphore code if $X^*L \subseteq LX^*$ and L is a prefix code;

(9) a left semaphore code if $LX^* \subseteq X^*L$ and L is a suffix code;

(10) a p-infix code if for all $x, y, z \in X^*$, $z \in L$ and $xzy \in L$ imply $y = 1$;

(11) an s-infix code if for all $x, y, z \in X^*$, $z \in L$ and $xzy \in L$ imply $x = 1$.

The class of all codes is denoted by $\mathcal{C}$. For any fixed alphabet X, if P is a property labelling codes or n-codes, then $\mathcal{C}_P$ is the class of codes or n-codes so labelled over X. Usually we use the initial of the name of a class of codes or n-codes to name the property of this class of codes or n-codes. For examples: $\mathcal{C}_p$, $\mathcal{C}_s$, $\mathcal{C}_{pi}$, $\mathcal{C}_{rs}$ are classes of prefix, suffix, p-infix, and right semaphore codes, respectively. In Table 1.1 below, we list some classes of codes and n-codes. The table also shows the property name and a binary relation defining that class of codes or n-codes.

Let S be an arbitrary non-empty set. A family $\mathcal{L}$ of subsets of S is said to be ***non-trivial*** if $\emptyset \notin \mathcal{L}$. For a language L, let Fin(L) be the set of non-empty, finite sublanguages of L. A family $\mathcal{L}$ of languages is said to be Fin-*determined* if it is non-trivial and

$$L \in \mathcal{L} \iff \text{Fin}(L) \subseteq \mathcal{L}$$

holds true.

Let $\mathcal{L}$ be a non-trivial class of languages over an alphabet X. $\mathcal{L}$ is *strictly non-trivial* if $\mathcal{L}$ satisfies the following additional condition:

$$\text{for all } u \in X^*,\ \{u\} \in \mathcal{L}.$$

The class of all codes and every subclass of codes given in Definition 1.2 are strictly non-trivial and Fin-determined.

A language $L \in \mathcal{L}$ is a *maximal element of* $\mathcal{L}$ if and only if for any $x \in X^+ \setminus L$, $L \cup \{x\}$ is not in $\mathcal{L}$, that is, L is a maximal element of $\mathcal{L}$ if and only if L is not properly contained in any other element of $\mathcal{L}$.

Two special subclasses of n-codes, n-ps-codes and g-3-ps-codes, are introduced in [Ito3]. A language $L \subseteq X^+$ is called an *n-prefix-suffix code*, or *n-ps-code* for short, if every subset of L with at most n elements is a prefix code or a suffix code. A language $L \subseteq X^+$ is called a *g-3-ps-code* if

(1) for any $u \neq v \in L$, $u \notin vX^+$ or $u \notin X^+v$,

(2) for any distinct $u, v, w \in L$, $u \notin vX^+$ or $u \notin X^+w$.

The classes of n-ps-codes and g-3-ps-codes are denoted by $\mathcal{PS}_n$ and $\mathcal{GPS}$, respectively.

languages	property	relation
uniform codes	u	$w \le_u v \iff w = v \lor \lvert w\rvert < \lvert v\rvert$
hypercodes	h	$w \le_h v \iff \exists n, \exists x_1, \ldots, x_n \in X^* : w = x_1 \cdots x_n \land v \in X^* x_1 X^* \cdots x_n X^*$
n-shuffle codes	sh_n	$w \omega_{sh_n} v \iff \exists x_1, \ldots, x_n \in X^* : w = x_1 \cdots x_n \land v \in X^* x_1 X^* \cdots x_n X^*$
prefix codes	p	$w \le_p v \iff v \in wX^*$
suffix codes	s	$w \le_s v \iff v \in X^* w$
bifix codes	b	$\omega_b = \le_p \cup \le_s$
2-codes	c	$w \le_c v \iff \exists x : v = wx = xw$
infix codes	i	$w \le_i v \iff v \in X^* w X^*$
p-infix codes	pi	$w \le_{pi} v \iff w = v \lor v \in X^* w X^+$
s-infix codes	si	$w \le_{si} v \iff w = v \lor v \in X^+ w X^*$
outfix codes	o	$w \omega_o v \iff \exists w_1, w_2 : w = w_1 w_2 \land v \in w_1 X^* w_2$
2-ps-codes	d	$\le_d = \le_p \cap \le_s$

Table 1.1. Some classes of codes and n-codes, and their class labels and the associated independence relations.

1.2.3 Relations

For a set S, let

$$[S]^n = \underbrace{S \times \ldots \times S}_{n \text{ times}}$$

be the n-fold Cartesian product of S. An *n-ary relation* ω on S is a subset of $[S]^n$.

A *strict binary relation* ω on X^* is defined as follows: for all $u, v \in X^*$,

(1) $u\omega u$ and $1\omega u$,

(2) $u\omega v$ implies $|u| \leq |v|$,

(3) $u\omega v$ and $|u| = |v|$ imply $u = v$.

A language L is called an *independent set with respect to a strict binary relation* ω if for all $u, v \in L$, $u\omega v$ implies that $u = v$. Some strict binary relations and the related independent sets are also listed in Table 1.1.

If a binary relation $\leq$ defined on a set S is reflexive, antisymmetric and transitive, then $\leq$ is called a *partial order* of S. A *strict partial order* is a strict binary relation which is also a partial order. A partial order $\leq$ is said to be a *total order on S* if for any $u, v \in S$, one has $u \leq v$ or $v \leq u$. A total order $\leq$ is called the *standard total order on X^** if for any $u, v \in X^*$, $u < v$ when $|u| < |v|$ and $u \leq v$ is the inherited lexicographic order from a given order on X when $|u| = |v|$. By the standard total order $\leq$ on X^*, we enumerate the totally ordered set $(X^*, \leq)$ as $\{x_0 < x_1 < x_2 < \ldots < x_m < \ldots\}$. For every $x \in X^*$, let the index of x be $\#(x) = m$, if $x = x_m$. We note that $x_0 = 1$. In the sequel, if $\leq$ is any total order on X^*, and if $A = \{x_1 < x_2 < \ldots\}$, $B = \{y_1 < y_2 < \ldots\}$ are two infinite languages over X, then following [Shy2], we define the *ordered catenation* of A and B to be the set $A \triangle B = \{x_i y_i \mid i = 1, 2, \ldots\}$ and for $i \geq 2$, let $A^{(i)} = A^{(i-1)} \triangle A = A \triangle A^{(i-1)}$.

1.3 Summary of Results

Now we summarize the main results of this thesis:

In Chapter 2, we investigate binary relations, p-infix codes and n-shuffle codes. The class of all nonempty, ω-independent sets is denoted by $\mathcal{L}_\omega$.

Let $w, v \in X^+$. We define the binary relation $\leq_{pi}$ as follows:

$$w \leq_{pi} v \text{ if and only if } (w = v \text{ or } v = xwy \text{ for some } x \in X^* \text{ and } y \in X^+).$$

- The relation $\leq_{pi}$ is a partial order on X^+.
- The class of all non-empty, $\leq_{pi}$-independent sets is the class of p-infix codes.

For every positive integer n, the binary relation ω_{sh_n} is defined on X^* by:

$$u\omega_{sh_n}v \iff u = u_1u_2\cdots u_n,\ v = v_0u_1v_1\cdots u_nv_n,\ \text{where } u_i, v_j \in X^*.$$

The relation ω_{sh_n} is reflexive and antisymmetic and it is called the n-shuffle relation.

- The relation ω_{sh_1} is transitive, but for $n \geq 2$, ω_{sh_n} is not transitive.

A non-empty set $L \subseteq X^+$ which is ω_{sh_n}-independent is a code, called an n-shuffle code.

- Every n-shuffle code is an m-shuffle code for $m \leq n$ and a prefix code.

We consider the relationship between binary relations and the independent sets and derive the following results:

- Let $\omega_1, \omega_2 \subseteq [X^*]^2$. Then $\operatorname{symm}\omega_1 = \operatorname{symm}\omega_2 \Rightarrow \mathcal{L}_{\omega_1} = \mathcal{L}_{\omega_2}$ where $\operatorname{symm}\omega_i$ denotes the symmetric closure of ω_i.

Let $\mathfrak{S}$ denote the set of all symmetric binary relations on X^*.

- Let $\omega_1, \omega_2 \in \mathfrak{S}$. Then $\omega_1 \subseteq \omega_2 \Rightarrow \mathcal{L}_{\omega_2} \subseteq \mathcal{L}_{\omega_1}$.

  If ω_2 is reflexive then also $\mathcal{L}_{\omega_2} \subseteq \mathcal{L}_{\omega_1} \Rightarrow \omega_1 \subseteq \omega_2$.

  Hence, if both ω_1 and ω_2 are reflexive, then $\omega_1 = \omega_2 \iff \mathcal{L}_{\omega_1} = \mathcal{L}_{\omega_2}$.

Because of this result, in the sequel, we need to consider reflexive and symmetric binary relations only.

Let $\mathfrak{RS}$ denote the set of reflexive and symmetric binary relations on X^*. For any strictly non-trivial, Fin-determined family $\mathcal{L}$ of languages over X, let the relation $\omega_{\mathcal{L}}$ be defined as

$$\omega_{\mathcal{L}} = \{(u,v) \mid u = v \text{ or } \{u,v\} \notin \mathcal{L}\}.$$

We derive the following result:

- Let $\mathcal{L}$ be a strictly non-trivial, Fin-determined family of languages. If $\mathcal{L} = \mathcal{L}_{\overline{\omega}}$ for some binary relation $\overline{\omega}$ then $\text{symm ref}\,\overline{\omega} = \omega_{\mathcal{L}}$.
- Let $\mathcal{L}$, $\mathcal{L}_1$ be two strictly non-trivial families of languages such that $\mathcal{L} \subsetneq \mathcal{L}_1 \subsetneq \mathcal{L}_{\omega_{\mathcal{L}}}$. Then there is no binary relation ω such that $\mathcal{L}_1 = \mathcal{L}_\omega$.
- For $n = 3, 4, \ldots$ there is no binary relation ω on X^* with $\mathcal{L}_\omega = \mathcal{C}_n$ or $\mathcal{L}_\omega = \mathcal{PS}_n$.

In Chapter 3, we consider a subclass of infix codes, the class of solid codes. The fact that a singleton is not necessarily a solid code provides the impetus to consider relations without the reflexive condition and to redefine the independent sets.

- A language L over X is a solid code if and only if every two words $u, v \in L$ satisfy the following conditions:

(1) $P(u) \cap S(v) = \emptyset$;

(2) If $u \neq v$ then $u \notin E(v)$ and $v \notin E(u)$ where $P(w) = \{u \mid u \in X^+ : \ w \in uX^+\}$, $S(w) = \{u \mid u \in X^+ : \ w \in X^+u\}$ and $E(w) = \{u \mid u \in X^+ : \ w \in X^*uX^*\}$ for any $w \in X^*$.

- For any word $u \in X^+$, $\{u\}$ is a solid code if and only if u is a primitive word.

In Chapter 4, we investigate the properties of intercodes. We have the following results:

- A language $L \subseteq X^+$ such that $L^{m+1} \cap X^+L^mX^+ = \emptyset$ for some $m \geq 1$ is a bifix code. A language $L \subseteq X^+$ satisfying this property is called an intercode of index m.

- Let $\mathcal{I}_m$ be the class of all intercodes of index m. Then we have the following strict hierarchy:

$$\mathcal{I}_1 \subsetneq \mathcal{I}_2 \subsetneq \ldots \subsetneq \mathcal{I}_m \subsetneq \ldots \subsetneq \mathcal{Q}.$$

- Let $L \subseteq X^+$. Then L is an intercode of index m if and only if L is a $(2m+1)$-intercode of index m.

These results provide examples of the fact that some classes of codes can only be characterized as the independent sets with respect to some n-ary relations for $n \geq 3$. The mechanism to express classes of languages as the independent sets with respect to n-ary relations is studied in Chapter 5.

Let S be an arbitrary set and let $\mathfrak{R}^{(n)}$ be the set of all n-ary relations on S. Let $\omega \in \mathfrak{R}^{(n)}$. For an n-tuple $x = (x_1, x_2, \ldots, x_n) \in [S]^n$, let

$$\operatorname{cont} x = \{y \mid y \in S,\ \exists i : x_i = y\}.$$

A set $L \subseteq S$ is said to be ***ω-independent*** if

$$x \in \omega \Rightarrow \operatorname{cont} x \not\subseteq L.$$

Let $\mathcal{L}_\omega$ denote the class of all nonempty, ω-independent sets and let

$$\mathcal{L}^{(n)} = \{\mathcal{L} \mid \exists \omega \subseteq [S]^n : \mathcal{L} = \mathcal{L}_\omega\}.$$

ω is said to be ***symmetric*** if it has the following property: For $x, y \in [S]^n$, if $x \in \omega$ and $\operatorname{cont} x = \operatorname{cont} y$ then $y \in \omega$. ω is said to be ***upward symmetric*** if it has the following property: For $x, y \in [S]^n$, if $x \in \omega$ and $\operatorname{cont} x \subseteq \operatorname{cont} y$ then $y \in \omega$. Let $\operatorname{symm} \omega$ and $\operatorname{upsymm} \omega$ denote the symmetric and the upward symmetric closures of ω, respectively. For any n-ary relation ω on S, let $\overline{\omega} = \operatorname{upsymm} \omega$. One has the following theorem:

- **Inclusion Theorem:** Let ω_1, ω_2 be n-ary relations on S. One has

$$\overline{\omega}_1 \subseteq \overline{\omega}_2 \iff \mathcal{L}_{\omega_2} \subseteq \mathcal{L}_{\omega_1}.$$

This theorem can be used to show the inclusion relationship between classes of codes using this relationship between finitary relations.

For any non-trival family $\mathcal{L}$ of subsets of S and any $n \in \mathbb{N}$, we consider the n-ary relation

$$\omega_{\mathcal{L}}^{(n)} = \{x \mid x \in [S]^n, \operatorname{cont} x \notin \mathcal{L}\}.$$

- Let $\mathcal{L}$ be a non-trivial family of subsets of S. If $\mathcal{L} = \mathcal{L}_{\tilde{\omega}}$ for some n-ary relation $\tilde{\omega}$ then

$$\operatorname{upsymm} \tilde{\omega} = \omega_{\mathcal{L}}^{(n)}.$$

- **Gap Theorem:** Let $n \in \mathbb{N}$ and $\mathcal{L}, \mathcal{L}_1$ be two strictly non-trivial families of subsets of S such that

$$\mathcal{L} \subseteq \mathcal{L}_1 \subsetneq \mathcal{L}_{\omega_{\mathcal{L}}^{(n)}}.$$

  Then there is no n-ary relation ω such that $\mathcal{L}_1 = \mathcal{L}_\omega$.

The gap theorem provides a powerful criterion to determine whether a given family of sets can be characterized as families of independent sets with respect to n-ary relations. For example, using the gap theorem, we show the following:

- For $n \in \mathbb{N}$ let ω_n be the n-ary relation on X^+ given by

$$x \in \omega_n \iff \operatorname{cont} x \notin \mathcal{C}$$

  for $x \in [X^+]^n$. The following properties obtain:

  (1) $\mathcal{C}_n = \mathcal{L}_{\omega_n}$ for all n.

  (2) There is no finitary relation ω such that $\mathcal{C} = \mathcal{L}_\omega$.

  (3) For all $m \in \mathbb{N}$, there is no n-ary relation ω with $\mathcal{L}_\omega = \mathcal{C}_m$ and $n < m$.

Let $\mathcal{I}_{n,m}$ denote the family of n-intercodes of index m over X. Consider the n-ary relation $\varrho_{n,m}$ on X^+ given by the following condition:

$$x \in \varrho_{n,m} \iff (\operatorname{cont} x)^{m+1} \cap X^+(\operatorname{cont} x)^m X^+ \neq \emptyset$$

for all $x \in [X^+]^n$. Clearly, $\mathcal{L}_{\varrho_{n,m}} = \mathcal{I}_{n,m}$.

- Let $k, n, m \in \mathbb{N}$ with $n \leq 2m+1$. If $k < n$ and $\omega \subseteq [X^+]^k$ then $\mathcal{I}_{n,m} \neq \mathcal{L}_\omega$.
- There is no binary relation ω such that $\mathcal{PS}_n = \mathcal{L}_\omega$ for $n = 3, 4$.

The results of this thesis provide a clear outline of the hierarchy of codes and n-codes and a powerful tool for the investigation of codes. They constitute a foundation for further research in this area.

CHAPTER 2

Codes and Binary Relations

2.1 Binary Relations and Their Independent Sets

This section deals with the relationships between binary relations and their independent sets. It generalizes results of [Shy1] concerning strict partial orders to arbitrary relations.

Definition 2.1 *Let ω be a binary relation on X^*. A language L is ω-independent if for any $u, v \in L$, $u\omega v$ implies $u = v$.*

Let $\mathcal{L}_\omega$ denote the class of all non-empty, ω-independent languages. For a binary relation ω, let $\mathrm{ref}\,\omega$, $\mathrm{symm}\,\omega$, and $\mathrm{trans}\,\omega$ denote the reflexive, symmetric, and transitive closures of ω, respectively. When we consider the connection between binary relations and their independent sets, the following lemma allows us to restrict our attention to symmetric binary relations.

Lemma 2.2 *Let $\omega_1, \omega_2 \subseteq X^* \times X^*$ be such that* $\mathrm{symm}\,\omega_1 = \mathrm{symm}\,\omega_2$. *Then* $\mathcal{L}_{\omega_1} = \mathcal{L}_{\omega_2}$.

Proof: Consider $L \in \mathcal{L}_{\omega_1}$ and suppose that $L \notin \mathcal{L}_{\omega_2}$. Then $\neg u\omega_1 v$ for all $u, v \in L$ with $u \neq v$. On the other hand, there exist distinct words u and v in L with $u\omega_2 v$ or $v\omega_2 u$. The fact that the symmetric closures of ω_1 and ω_2 coincide implies that $u\omega_1 v$ or $v\omega_1 u$, a contradiction! This proves $\mathcal{L}_{\omega_1} \subseteq \mathcal{L}_{\omega_2}$. The converse is proved analogously. □

Let $\mathfrak{S}$ denote the set of all symmetric binary relations on X^*.

Lemma 2.3 *Let ω_1, $\omega_2 \in \mathfrak{S}$. Then*

$$\omega_1 \subseteq \omega_2 \Rightarrow \mathcal{L}_{\omega_2} \subseteq \mathcal{L}_{\omega_1}.$$

If ω_2 is reflexive then also

$$\mathcal{L}_{\omega_2} \subseteq \mathcal{L}_{\omega_1} \Rightarrow \omega_1 \subseteq \omega_2.$$

Hence, if both ω_1 and ω_2 are reflexive, then

$$\omega_1 = \omega_2 \iff \mathcal{L}_{\omega_1} = \mathcal{L}_{\omega_2}.$$

Proof: Let $\omega_1 \subseteq \omega_2$. Consider $L \in \mathcal{L}_{\omega_2}$ and assume that $L \notin \mathcal{L}_{\omega_1}$. Then there are $u, v \in L$, $u \neq v$, such that $u\omega_1 v$. But then $u\omega_2 v$ by $\omega_1 \subseteq \omega_2$. This implies $L \notin \mathcal{L}_{\omega_2}$, a contradiction! Conversely, let ω_2 be reflexive, $\mathcal{L}_{\omega_2} \subseteq \mathcal{L}_{\omega_1}$, and assume that u, v exist such that $u\omega_1 v$ and $\neg u\omega_2 v$. Then $u \neq v$, as ω_2 is reflexive. Thus, $\{u, v\} \in \mathcal{L}_{\omega_2}$, but $\{u, v\} \notin \mathcal{L}_{\omega_1}$, a contradiction! □

Lemma 2.4 *Let $\omega_1 \in \mathfrak{S}$ be any symmetric binary relations on X^* such that ω_2 is the reflexive closure of ω_1. Then*

$$\mathcal{L}_{\omega_1} = \mathcal{L}_{\omega_2}.$$

Proof: Clearly $\omega_1 \subseteq \omega_2$ and $\omega_2 \in \mathfrak{S}$. By the first part of Lemma 2.3, one has $\mathcal{L}_{\omega_2} \subseteq \mathcal{L}_{\omega_1}$. Now consider $L \in \mathcal{L}_{\omega_1}$ and assume that $L \notin \mathcal{L}_{\omega_2}$. Then there are $u, v \in L$, $u \neq v$, such that $u\omega_2 v$. But $u \neq v$ implies that $u\omega_1 v$ as ω_2 is the reflexive closure of ω_1, hence $L \notin \mathcal{L}_{\omega_1}$, a contradiction! □

Because of this result we need to consider reflexive and symmetric binary relations only in the sequel. Let $\mathfrak{RS}$ denote the set of reflexive and symmetric binary relations on X^* and let $\mathfrak{L}$ denote the set $\{\mathcal{L} \mid \exists \omega \subseteq X^* \times X^* : \mathcal{L} = \mathcal{L}_\omega\}$. The preceding results immediately imply the following statement.

Theorem 2.5 *The mapping $\omega \mapsto \mathcal{L}_\omega$ is an antitonic bijection of $\mathfrak{RS}$ onto $\mathfrak{L}$.*

A $\cup$-semilattice is said to be $\cup$-complete if it is closed under unions taken over arbitrary index sets. Similarly, one defines $\cap$-complete $\cap$-semilattices. A lattice is complete if it is both $\cup$-complete and $\cap$-complete. Note that $(\mathfrak{RS}, \cup, \cap)$ is a complete lattice. Hence, the mapping $\omega \mapsto \mathcal{L}_\omega$ induces a lattice structure on $\mathfrak{L}$ via $\mathcal{L}_{\omega_1} \vee \mathcal{L}_{\omega_2} = \mathcal{L}_{\omega_1 \cap \omega_2}$ and $\mathcal{L}_{\omega_1} \wedge \mathcal{L}_{\omega_2} = \mathcal{L}_{\omega_1 \cup \omega_2}$ such that this mapping is an isomorphism of $(\mathfrak{RS}, \cup, \cap)$ onto $(\mathfrak{L}, \wedge, \vee)$.

For any strictly non-trivial family $\mathcal{L}$ of languages over X, let

$$\Omega(\mathcal{L}) = \{\omega \mid \omega \in \mathfrak{RS}, \mathcal{L} \subseteq \mathcal{L}_\omega\}$$

and

$$\omega(\mathcal{L}) = \{\omega \mid \omega \in \mathfrak{RS}, \mathcal{L}_\omega \subseteq \mathcal{L}\}.$$

$\Omega(\mathcal{L})$ always contains at least the equality relation. The strict non-triviality of $\mathcal{L}$ implies that $\omega(\mathcal{L})$ is non-empty; it contains at least the universal relation. We now turn to investigating the structure of the sets $\Omega(\mathcal{L})$ and $\omega(\mathcal{L})$.

Theorem 2.6 *Let $\mathcal{L}$ be a strictly non-trivial family of languages and let $\omega_1, \omega_2 \in \mathfrak{RS}$.*

(1) Let $\omega_0 \in \Omega(\mathcal{L})$. If $\omega_1 \subseteq \omega_0$ then $\omega_1 \in \Omega(\mathcal{L})$. Moreover, $\omega_0 \cap \omega_2 \in \Omega(\mathcal{L})$ and $\Omega(\mathcal{L})$ is a $\cap$-complete $\cap$-semilattice with the equality relation as its minimum.

(2) $\Omega(\mathcal{L})$ is a $\cup$-complete $\cup$-semilattice. Hence $\Omega(\mathcal{L})$ has a maximum.

(3) Let $\omega_0 \in \omega(\mathcal{L})$. If $\omega_0 \subseteq \omega_1$, then $\omega_1 \in \omega(\mathcal{L})$. Moreover, $\omega_0 \cup \omega_2 \in \omega(\mathcal{L})$, and $\omega(\mathcal{L})$ is a $\cup$-complete $\cup$-semilattice with the universal relation as its maximum.

Proof: For (1), let $\omega_1 \subseteq \omega_0 \in \Omega(\mathcal{L})$. Then $\mathcal{L} \subseteq \mathcal{L}_{\omega_0} \subseteq \mathcal{L}_{\omega_1}$, thus $\omega_1 \in \Omega(\mathcal{L})$. As $\omega_0 \cap \omega_2 \subseteq \omega_0$ one has $\omega_0 \cap \omega_2 \in \Omega(\mathcal{L})$. Let $\{\omega_i \mid i \in I\}$ be any family of relations with $\omega_i \in \Omega(\mathcal{L})$ for $i \in I$. Then

$$\omega = \bigcap_{i \in I} \omega_i \subseteq \omega_j$$

for every $j \in I$. Hence $\omega \in \Omega(\mathcal{L})$. As $\Omega(\mathcal{L})$ contains the equality relation, it follows that $\bigcap_{\omega\in\Omega(\mathcal{L})}\omega = \{(u,u) \mid u \in X^*\}$.

For (2), consider a family $\{\omega_i \mid i \in I\}$ of relations $\omega_i \in \Omega(\mathcal{L})$ for $i \in I$. Let

$$\omega = \bigcup_{i\in I}\omega_i.$$

We show that $\omega \in \Omega(\mathcal{L})$, that is, that $\mathcal{L} \subseteq \mathcal{L}_\omega$. Assume the contrary. Then there is $L \in \mathcal{L}$ such that L is not ω-independent. Therefore, for some $u, v \in L$, $u \neq v$, we have $u\omega v$. Hence, there exists an $i \in I$ such that $u\omega_i v$. But then $L \notin \mathcal{L}_{\omega_i}$, that is, $\omega_i \notin \Omega(\mathcal{L})$, a contradiction! Clearly, the relation

$$\bigcup_{\omega\in\Omega(\mathcal{L})}\omega$$

is the maximum of $\Omega(\mathcal{L})$.

For (3), let $\omega_0 \subseteq \omega_1$ and $\omega_0 \in \omega(\mathcal{L})$. Then $\mathcal{L}_{\omega_1} \subseteq \mathcal{L}_{\omega_0} \subseteq \mathcal{L}$, hence $\omega_1 \in \omega(\mathcal{L})$. As $\omega_0 \subseteq \omega_0 \cup \omega_2$, one has $\omega_0 \cup \omega_2 \subseteq \omega(\mathcal{L})$. Let $\{\omega_i \mid i \in I\}$ be any family with $\omega_i \in \omega(\mathcal{L})$ for $i \in I$. Then

$$\omega_j \subseteq \omega = \bigcup_{i\in I}\omega_i$$

for every $j \in I$. Hence $\omega \in \omega(\mathcal{L})$. As $\omega(\mathcal{L})$ contains the universal relation, it follows that

$$\bigcup_{\omega\in\omega(\mathcal{L})}\omega$$

is the universal relation. □

Further below we show that $\omega(\mathcal{L})$ is not closed under intersections in general. It will turn out that $\omega(\mathcal{L})$ is closed under intersections if and only if $\mathcal{L} = \mathcal{L}_\omega$ for some binary relation ω.

Theorem 2.7 *Let $\mathcal{L}$ be a strictly non-trivial, Fin-determined family of languages over X. Then the following properties hold true:*

(1) The relation

$$\omega_{\mathcal{L}} = \{(u,v) \mid u = v \text{ or } \{u,v\} \notin \mathcal{L}\}$$

is the maximum of $\Omega(\mathcal{L})$.

(2) For every $\omega_0 \in \omega(\mathcal{L})$ there is a minimal element $\omega_\infty \in \omega(\mathcal{L})$ such that $\omega_\infty \subseteq \omega_0$.

Proof: Let ω be the maximum of $\Omega(\mathcal{L})$. Both ω and $\omega_{\mathcal{L}}$ are reflexive. Hence in order to prove that $\omega = \omega_{\mathcal{L}}$ it is sufficient to show that $\mathcal{L}_\omega = \mathcal{L}_{\omega_{\mathcal{L}}}$. Consider $L \in \mathcal{L}_\omega$ and suppose that $L \notin \mathcal{L}_{\omega_{\mathcal{L}}}$. Then $u\omega_{\mathcal{L}}v$ for some $u,v \in L$, $u \neq v$. By the definition of $\omega_{\mathcal{L}}$ this implies that $\{u,v\} \notin \mathcal{L}$. Now consider the relation

$$\widetilde{\omega} = \omega \cup \{(u,v),(v,u)\}.$$

Then $L' \in \mathcal{L}_\omega \setminus \mathcal{L}_{\widetilde{\omega}}$ if and only if $\{u,v\} \subseteq L'$. But for such L' one has $L' \notin \mathcal{L}$ as $\mathcal{L}$ is Fin-determined. Therefore, $\widetilde{\omega} \in \Omega(\mathcal{L})$. However, ω is a proper subset of $\widetilde{\omega}$ contradicting the maximality of ω. This shows $L \in \mathcal{L}_{\omega_{\mathcal{L}}}$, that is $\mathcal{L}_\omega \subseteq \mathcal{L}_{\omega_{\mathcal{L}}}$.

Conversely, let $L \in \mathcal{L}_{\omega_{\mathcal{L}}}$. Then $\{u,v\} \in \mathcal{L}$ for all $u,v \in L$, $u \neq v$. Therefore, $\neg u\hat{\omega}v$ for every $\hat{\omega} \in \Omega(\mathcal{L})$. In particular, $\neg u\omega v$, that is $L \in \mathcal{L}_\omega$. This proves $\mathcal{L}_{\omega_{\mathcal{L}}} \subseteq \mathcal{L}_\omega \subseteq \mathcal{L}_{\omega_{\mathcal{L}}}$, that is, $\omega_{\mathcal{L}}$ is the maximum of $\Omega(\mathcal{L})$.

To prove the second statement, assume that ω_0 is not minimal—otherwise nothing needs to be proved. Consider a decreasing chain

$$\omega_0 \supseteq \omega_1 \supseteq \omega_2 \supseteq \ldots$$

in $\omega(\mathcal{L})$ and let $\omega = \bigcap_{i \geq 0} \omega_i$. Let $L \in \mathcal{L}_\omega$. We show that $L \in \mathcal{L}$. As $\mathcal{L}$ is Fin-determined, it is sufficient to show that $L' \in \mathcal{L}$ for every $L' \in \mathrm{Fin}(L)$.

Let $L' \in \mathrm{Fin}(L)$. For every $u,v \in L'$, $u \neq v$, there is an index $i(u,v)$ such that $\neg u\omega_{i(u,v)}v$. Let

$$j = \max\{i(u,v) \mid u,v \in L',\ u \neq v\}.$$

As L' is finite, j exists. Then $L' \in \mathcal{L}_{\omega_j} \subseteq \mathcal{L}$. This implies $L \in \mathcal{L}$ and therefore $\omega \in \omega(\mathcal{L})$. Now the existence of ω_∞ follows by Zorn's lemma. □

An intuitive explanation of the first part of Theorem 2.7 is as follows: Suppose $\mathcal{L}$ is defined by some property P of languages, that is, $\mathcal{L} = \mathcal{L}_P$. A language L is in $\mathcal{L}_{\omega_\mathcal{L}}$ if and only if every subset of L of cardinality 2 is in $\mathcal{L}$, that is, has property P. Thus, in the cases of codes and prefix codes one has

$$\omega_{\mathcal{C}_p} = \text{symm} \leq_p \quad \text{and} \quad \mathcal{L}_{\omega_{\mathcal{C}_p}} = \mathcal{C}_p$$

while

$$\mathcal{C} \subsetneq \mathcal{L}_{\omega_\mathcal{C}},$$

for instance.

Corollary 2.8 ([Shy1]) *Let $\mathcal{L} = \mathcal{C}$, the class of codes. Then* symm $\leq_c = \max \Omega(\mathcal{C})$.

Proof: $\leq_c$ characterizes the class of 2-codes [Ito2], that is, $\mathcal{L}_{\leq_c}$ is the class of non-empty languages $L \subseteq X^+$ such that for every $u, v \in L$, $u \neq v$, the set $\{u, v\}$ is a code. Therefore, symm $\leq_c = \omega_\mathcal{C}$. □

Corollary 2.9 *Consider $\mathcal{C}_p \cup \mathcal{C}_s$ with $\mathcal{C}_p$ and $\mathcal{C}_s$ the classes of prefix codes and suffix codes, respectively. Then* symm $\leq_d = \max \Omega(\mathcal{C}_p \cup \mathcal{C}_s)$.

Proof: By definition, symm $\leq_d = \{(u, v) \mid u = v \quad \text{or} \quad \{u, v\} \notin \mathcal{C}_p \cup \mathcal{C}_s\}$. □

While $\Omega(\mathcal{L})$ always has a minimum and a maximum, $\omega(\mathcal{L})$ is only guaranteed to have a maximum. It turns out that it has a minimum if and only if $\mathcal{L}$ coincides with the class of independent sets with respect to some binary relation.

Theorem 2.10 *Let $\mathcal{L}$ be a strictly non-trivial family of languages.*

(1) If $\mathcal{L} = \mathcal{L}_{\tilde{\omega}}$ for some binary relation $\tilde{\omega}$ then $\omega(\mathcal{L})$ has a minimum ω and

$$\omega = \text{symm ref}\, \tilde{\omega} = \max \Omega(\mathcal{L}).$$

(2) If $\mathcal{L}$ is Fin-determined and $\omega(\mathcal{L})$ has a minimum ω then $\mathcal{L} = \mathcal{L}_\omega$ and $\omega = \max \Omega(\mathcal{L})$.

Proof: Suppose that $\mathcal{L} = \mathcal{L}_{\tilde{\omega}}$. Without loss of generality we assume that $\tilde{\omega}$ is symmetric and reflexive. Then $\tilde{\omega} \in \omega(\mathcal{L})$. Suppose there is $\omega_1 \in \omega(\mathcal{L})$, such that ω_1 is a proper subset of $\tilde{\omega}$. Then $\mathcal{L} = \mathcal{L}_{\tilde{\omega}}$ is a proper subset of $\mathcal{L}_{\omega_1}$ contradicting the fact that $\omega_1 \in \omega(\mathcal{L})$. Hence, $\tilde{\omega}$ is minimal. Now suppose that there is another minimal element $\hat{\omega} \in \omega(\mathcal{L})$. Then $\mathcal{L}_{\hat{\omega}} \subseteq \mathcal{L} = \mathcal{L}_{\tilde{\omega}}$. Hence, $\tilde{\omega} \subseteq \hat{\omega}$ and, therefore, $\tilde{\omega} = \hat{\omega}$. Thus $\tilde{\omega}$ is the minimum of $\omega(\mathcal{L})$. Clearly also $\tilde{\omega} \in \Omega(\mathcal{L})$. Hence $\tilde{\omega} \subseteq \max \Omega(\mathcal{L})$ and $\mathcal{L}_{\max \Omega(\mathcal{L})} \subseteq \mathcal{L}_{\tilde{\omega}} = \mathcal{L}$, that is, $\mathcal{L}_{\max \Omega(\mathcal{L})} = \mathcal{L}_{\tilde{\omega}}$. This proves part (1).

For part (2), we only need to show that $\mathcal{L} = \mathcal{L}_\omega$. By the definition of $\omega(\mathcal{L})$, one has $\mathcal{L}_\omega \subseteq \mathcal{L}$. Suppose there is a language $L \in \mathcal{L} \setminus \mathcal{L}_\omega$. Then $\{u, v\}$ is ω-dependent for some u and v, $u \neq v$, while $\{u, v\} \in \mathcal{L}$ as $\mathcal{L}$ is Fin-determined. Let

$$\tilde{\omega} = \{(x, y) \mid x, y \in X^* : (x, y) \notin \{(u, v), (v, u)\}\}.$$

Then $\tilde{\omega} \in \omega(\mathcal{L})$ as

$$\mathcal{L}_{\tilde{\omega}} = \{\{u, v\}\} \cup \{\{x\} \mid x \in X^+\}.$$

Let $\hat{\omega}$ be a minimal element of $\omega(\mathcal{L})$ with $\hat{\omega} \subseteq \tilde{\omega}$. The existence of $\hat{\omega}$ follows from the fact that $\mathcal{L}$ is Fin-determined. Then $\neg u\hat{\omega}v$ and $u\omega v$, that is, $\hat{\omega} \neq \omega$, contradicting the fact that ω is unique. Thus $\mathcal{L} = \mathcal{L}_\omega$. $\square$

An explanation of the first part of Theorem 2.10 is as follows: The independent sets with respect to a strict binary relation are exactly the same as the independent sets with respect to the symmetric and reflexive closure of this binary relation. Hence, all the early work and results concerning strict binary relations can be preserved in the consideration of symmetric and reflexive binary relations.

The results obtained so far suffice to extend the non-characterizability proofs of [Shy1] and [Jür5].

Theorem 2.11 (Gap Theorem) *Let $\mathcal{L}$ be any strictly non-trivial family of languages over X with $C_p \cup C_s \subseteq \mathcal{L} \subseteq C$. Then there is no binary relation ω on X^* such that $\mathcal{L} = \mathcal{L}_\omega$.*

Proof: The proof is based on a generalization of the idea used in the proof of [Ito2], Theorem 3.3. Let ω_p, ω_s be the symmetric and reflexive binary relations defining C_p and C_s, respectively, that is, the symmetric closures of $\leq_p$ and of $\leq_s$. Assume that ω is a binary relation such that $\mathcal{L}$ is the class of all ω-independent languages, that is, $\mathcal{L} = \mathcal{L}_\omega$. Without loss of generality we may assume that $\omega \in \mathfrak{RS}$. Then $C_p \cup C_s \subseteq \mathcal{L}$ implies $\omega \subseteq \omega_p \cap \omega_s$. For $a, b \in X$, $a \neq b$, consider $L = \{a, b, ab\}$. Then L is not a code, hence $L \notin \mathcal{L}$ as $\mathcal{L} \subseteq C$. We show that L is ω-independent. Let $u, v \in L$, $u \neq v$, $u\omega v$. Then by $\omega \subseteq \omega_p \cap \omega_s$ there are words $x, y \in X^*$ such that $v = ux = yu$. But this is not possible for any choice of u and v in L. Thus L is ω-independent, that is, $L \in \mathcal{L}_\omega$, a contradiction! □

Corollary 2.12 *There is no binary relation ω on X^* such that the class C of codes over X is the class of ω-independent languages over X.*

Recall that the class $C_p \cup C_s$ coincides with the class $\mathcal{PS}_4$ of 4-ps-codes [Ito3]; for a general definition of n-ps-codes see further below.

Corollary 2.13 *There is no binary relation ω on X^* such that the class $\mathcal{PS}_4$ of 4-ps-codes over X is the class of ω-independent languages over X.*

Corollary 2.14 *The class $\mathcal{L}_{\omega_p \cap \omega_s}$ is not a class of codes.*

We now develop a more general method for proving non-characterizability as independent sets of binary relations. As a special case we then obtain a result of [Ito2] stating that n-codes cannot be characterized as independent sets of binary relations. We start with three useful lemmata. Let $\mathfrak{F}$ denote the class of all strictly non-trivial families of languages.

Lemma 2.15 *The class $\mathfrak{F}$ of all strictly non-trivial families of languages over X is a complete lattice. The family of all singleton languages is the minimum and the set $2^{X^+} \setminus \{\emptyset\}$ is the maximum of $\mathfrak{F}$.*

Proof: Consider a family $\{\mathcal{L}_i \mid i \in I\}$ of strictly non-trivial families of languages over X where I is some index set. Clearly, both

$$\mathcal{L} = \bigcap_{i \in I} \mathcal{L}_i \quad \text{and} \quad \mathcal{L}' = \bigcup_{i \in I} \mathcal{L}_i$$

are strictly non-trivial families of languages. □

Lemma 2.16 *Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be two strictly non-trivial families of languages over X with $\mathcal{L}_1 \subseteq \mathcal{L}_2$. Then $\max \Omega(\mathcal{L}_2) \subseteq \max \Omega(\mathcal{L}_1)$.*

Proof: Let $\omega_1 = \max \Omega(\mathcal{L}_1)$ and $\omega_2 \in \Omega(\mathcal{L}_2)$. Then $\mathcal{L}_1 \subseteq \mathcal{L}_2$ implies $\Omega(\mathcal{L}_2) \subseteq \Omega(\mathcal{L}_1)$, hence $\max \Omega(\mathcal{L}_2) \subseteq \max \Omega(\mathcal{L}_1)$. □

Recall that a ***closure operator*** on a lattice is an idempotent, monotonic mapping f of the lattice into itself such that $x \leq f(x)$ for all lattice elements x. Generalizing our notation, for any $\mathcal{L} \in \mathfrak{F}$ let

$$\omega_{\mathcal{L}} = \max \Omega(\mathcal{L}).$$

In Theorem 2.7 a characterization of $\omega_{\mathcal{L}}$ was given for the case of $\mathcal{L}$ being Fin-determined.

Lemma 2.17 *The mapping*

$$\mathcal{L} \mapsto \mathcal{L}_{\omega_{\mathcal{L}}}$$

is a closure operator on $\mathfrak{F}$.

Proof: Consider $\mathcal{L}_1, \mathcal{L}_2 \in \mathfrak{F}$ with $\mathcal{L}_1 \subseteq \mathcal{L}_2$. By Lemma 2.16 one has

$$\omega_{\mathcal{L}_2} \subseteq \omega_{\mathcal{L}_1}.$$

Hence,

$$\mathcal{L}_{\omega_{\mathcal{L}_1}} \subseteq \mathcal{L}_{\omega_{\mathcal{L}_2}},$$

that is, the operator is monotonic. To show that it is idempotent one verifies that

$$\omega_{\mathcal{L}_1} = \omega_{\mathcal{L}_{\omega_{\mathcal{L}_1}}}$$

using the definition of $\omega_{\mathcal{L}_1}$. Finally,

$$\mathcal{L}_1 \subseteq \mathcal{L}_{\omega_{\mathcal{L}_1}}$$

is an immediate consequence of the definition of $\Omega(\mathcal{L}_1)$ and the fact that $\omega_{\mathcal{L}_1} \in \Omega(\mathcal{L}_1)$. □

Theorem 2.18 *Let $\mathcal{L}$ and $\mathcal{L}_1$ be strictly non-trivial families of languages over X such that*

$$\mathcal{L} \subseteq \mathcal{L}_1 \subsetneq \mathcal{L}_{\omega_{\mathcal{L}}}.$$

Then there is no binary relation ω such that $\mathcal{L}_1 = \mathcal{L}_\omega$.

Proof: By Lemma 2.17,

$$\mathcal{L}_{\omega_{\mathcal{L}}} = \mathcal{L}_{\omega_{\mathcal{L}_1}}.$$

Assume that $\mathcal{L}_1 = \mathcal{L}_\omega$ for some binary binary relation ω. Without loss of generality we may assume that ω is reflexive and symmetric. Then

$$\omega = \max \Omega(\mathcal{L}_1) = \omega_{\mathcal{L}_1} = \omega_{\mathcal{L}},$$

hence

$$\mathcal{L}_1 = \mathcal{L}_{\omega_{\mathcal{L}}},$$

a contradiction! □

Theorem 2.18 provides a powerful tool for proving that certain classes of languages cannot be characterized as families of independent sets with respect to binary relations.

Corollary 2.19 *If $\mathcal{L}$ is a family of languages satisfying*

$$\mathcal{C} \subseteq \mathcal{L} \subsetneq \mathcal{C}_2 \quad or \quad \mathcal{C}_p \cup \mathcal{C}_s \subseteq \mathcal{L} \subsetneq \mathcal{PS}_2$$

then there is no binary relation ω on X^* with $\mathcal{L} = \mathcal{L}_\omega$.

Proof: The families $\mathcal{C}$ and $\mathcal{C}_p \cup \mathcal{C}_s$ cannot be characterized by binary relations. Moreover

$$\mathcal{L}_{\omega_\mathcal{C}} = \mathcal{C}_2 \quad \text{and} \quad \mathcal{L}_{\omega_{\mathcal{C}_p \cup \mathcal{C}_s}} = \mathcal{PS}_2.$$

□

Corollary 2.20 ([Ito2], [Ito3]) *For $n = 3, 4, \ldots$ there is no binary relation ω on X^* with $\mathcal{L}_\omega = \mathcal{C}_n$ or $\mathcal{L}_\omega = \mathcal{PS}_n$.*

2.2 p-Infix Codes

This section deals with properties of p-infix codes. The partial order $\leq_{pi}$ is defined to characterize the class of p-infix codes. The partial order $\leq_{si}$ and s-infix codes can be defined and characterized analogously.

Definition 2.21 *The binary relation $\leq_{pi}$ is defined as follows: For $u, v \in X^*$,*

$$u \leq_{pi} v \iff u = v \text{ or } v = xuy \text{ for some } x \in X^* \text{ and } y \in X^+.$$

Lemma 2.22 *The relation $\leq_{pi}$ is a partial order on X^*.*

Proof: It is clear that $u \leq_{pi} u$ for all $u \in X^*$. If $u, v \in X^*$ such that $u \leq_{pi} v$ and $v \leq_{pi} u$, then clearly $u = v$. Now let $u, v, w \in X^*$ such that $u \leq_{pi} v$ and $v \leq_{pi} w$. If $u = v$ or $v = w$, then the result $u \leq_{pi} w$ is clear. If u, v, w are distinct words, then $v = xuy$ and $w = x'vy'$ for some $x, x' \in X^*$, $y, y' \in X^+$. Hence $w = x'xuyy'$ and then $u \leq_{pi} w$. □

Theorem 2.23 $\mathcal{C}_{pi} = \mathcal{L}_{\leq_{pi}}$.

Proof: First we show that $\mathcal{C}_{pi} \subseteq \mathcal{L}_{\leq_{pi}}$. Let $L \in \mathcal{C}_{pi}$. If there exist $u, v \in L$, $u \neq v$, such that $u \leq_{pi} v$, then $v = xuy$ for some $x \in X^*$, $y \in X^+$. This contradicts the definition of p-infix codes. Hence, $L \in \mathcal{L}_{\leq_{pi}}$.

Now we show that $\mathcal{L}_{\leq_{pi}} \subseteq \mathcal{C}_{pi}$. Let $L \in \mathcal{L}_{\leq_{pi}}$. If $L \notin \mathcal{C}_{pi}$, then there exist two distinct words $u, v \in L$ such that $v = xuy$ for some $x \in X^*$, $y \in X^+$. This means that there exist two distinct words $u, v \in L$ such that $u \leq_{pi} v$. Thus $L \notin \mathcal{L}_{\leq_{pi}}$, a contradiction. □

The following result concerning p-infix codes is taken from [Ito4]. For more details, see that paper.

Corollary 2.24 ([Ito4]) *A language $L \subseteq X^+$ is a p-infix code if and only if L is a subset of a right semaphore code.*

A language $L \subseteq X^+$ is called *r-shifting* if $X^*L \subseteq LX^*$. We have the following result.

Lemma 2.25 ([Jür1]) *If L is a r-shifting language and a code, then L is a maximal prefix code, a right semaphore code, and a maximal code.*

Corollary 2.24 and Lemma 2.25 imply the following theorem:

Theorem 2.26 *Let $L \subseteq X^+$. Then the following three statements are equivalent:*

(1) L is a right semaphore code;

(2) L is a p-infix code and L is a maximal prefix code;

(3) L is a p-infix code and L is r-shifting.

Proof: Let L be a right semaphore code. Corollary 2.24 implies that L is a p-infix code. The definition of right semaphore codes implies that L is a r-shifting language. By Lemma 2.25, L is also a maximal prefix code. This shows that (1) implies (2).

Now let L be a p-infix code and a maximal prefix code. Corollary 2.24 implies that L is a subset of a right semaphore code L'. Lemma 2.25 implies that L' is a maximal prefix code. Since L is also a maximal prefix code, the case can only be that $L = L'$. This shows that (2) implies (1).

Now the result that (1) implies (3) is derived directly from (2) and the definition of right semaphore codes. By Lemma 2.25, (3) implies (1). □

Let X_1 and X_2 be two alphabets. A homomorphism $h : X_1^* \to X_2^*$ is said to be *non-erasing* if $h(x) \neq 1$ for all $x \in X_1$. Whether a homomorphism preserves the p-infix code property or not can be tested by just testing whether or not this homomorphism preserves p-infix codes each of whose words has a length not exceeding 2.

Theorem 2.27 *Let X be an alphabet with $|X| \geq 2$ and let h be a homomorphism of X^*. The following two statements are equivalent:*

(1) $h(L)$ is a p-infix code for every p-infix code $L \subseteq X^+$;

(2) $h(L)$ is a p-infix code for every p-infix code $L \subseteq X \cup X^2$.

Proof: It is obvious that condition (1) implies condition (2). Now we show that condition (2) implies condition (1). If h is erasing then there is a letter $a \in X$ such that $h(a) = 1$. Since $|X| \geq 2$, there is a letter $b \in X$ such that $b \neq a$. The set $L = \{b^2, ba\} \subseteq X^2$ is a p-infix code while $h(L) = \{h(b)h(b), h(b)\}$ is not. This implies that h must be non-erasing. If h is not injective on X, that is, there are distinct letters $a, b \in X$ such that $h(a) = h(b)$, then h does not preserve the p-infix property for some $L \subseteq X \cup X^2$. Indeed, the set $L = \{a, ba\}$ is a p-infix code while

$h(L)$ is not. Therefore, h has to be non-erasing and injective on X and we may assume that $|h(X)| = |X|$.

Let $L \subseteq X^+$ be a p-infix code such that $h(L)$ is not a p-infix code. Then there are distinct words $u, v \in L$ such that $\{u, v\}$ is a p-infix code whereas $h(\{u, v\})$ is not. Without loss in generality, we may assume that $L = \{u, v\}$. Of course, $L \not\subseteq X \cup X^2$. Indeed, the case that $L \subseteq X \cup X^2$ contradicts condition (2). Let L be chosen in such a way that $|u| + |v|$ is minimal with these properties. Let

$$u = u_1 u_2 \cdots u_r \quad \text{and} \quad v = v_1 v_2 \cdots v_s$$

with $u_1, u_2, \ldots, u_r, v_1, v_2, \ldots, v_s \in X$. By the choice of L, $h(L)$ is not a p-infix code. Without loss in generality, we assume that $h(u) \leq_{pi} h(v)$, that is, there exist words x, y and an index j such that

$$1 \leq j \leq s, \;\; h(v_1 \cdots v_j) = xh(u)y \;\; \text{and} \;\; |y| < |h(v_j)|.$$

Thus, if $y = 1$, the condition that $h(L)$ is not a p-infix code implies that $j < s$ and $u \leq_{pi} v$, contradicting the fact that L is a p-infix code.

Therefore, $y \neq 1$. Moreover, for $k = 1, 2, \ldots, r$ the inequalities

$$(2.1) \qquad |h(v_1) \cdots h(v_k)| < |xh(u_1) \cdots h(u_k)| < |h(v_1) \cdots h(v_{k+1})|$$

hold true. Otherwise, as u is not a p-infix factor of v, some $h(u_n)$ is a p-infix factor of some $h(v_m)$ contradicting the minimality of L. This implies that $h(u_1)$ is a p-infix factor of $h(v_1 v_2)$. By condition (2), u_1 is a p-infix factor of $v_1 v_2$. This case can happen only when $u_1 = v_1$. For an index k such that $u_i = v_i$ for all $i \leq k$, consider the case $k + 1$. The inequalities (2.1) and $u_i = v_i$ for all $i \leq k$ imply that $|h(v_{k+1})| < |x_{k+1} h(u_{k+1})| < |h(v_{k+1} v_{k+2})|$ for some $x_{k+1} \in X^*$. Again, the case can happen only when $u_{k+1} = v_{k+1}$. By induction on k, u is a p-infix factor of v, contradicting the fact that L is a p-infix code. □

If a homomorphism h is p-infix preserving then it is injective. But an injective homomorphism is not necessary p-infix preserving. For example, let $h : X \to Y$ where $X = \{a, b, c\}$ and $Y = \{a, b, c, d\}$ such that $h(a) = ab$, $h(b) = bc$, $h(c) = cd$. Then $h(\{ac, b\}) = \{abcd, bc\}$ which maps a p-infix code onto a set which is not a p-infix code.

Morphisms whose inverses preserve the p-infix property are characterized in the following theorem.

Theorem 2.28 *Let X and Y be alphabets and let h be a homomorphism of Y^* into X^*. The following conditions are equivalent:*

(1) $h^{-1}(L)$ is a p-infix code for every p-infix code $L \subseteq X^*$;

(2) $h^{-1}(1) = \{1\}$.

Proof: First we show that (2) implies (1). Suppose there exists a p-infix code L over X such that $h^{-1}(L)$ is not a p-infix code. Then there are words $x, u, y \in Y^*$ such that $u, xuy \in h^{-1}(L)$ and $y \neq 1$. Since $h(u) \in L$ and $h(xuy) = h(x)h(u)h(y) \in L$, $h(y) = 1$ where $y \neq 1$. This contradicts condition (2). This proves that (2) implies (1). The converse is obvious. □

2.3 Shuffle Relations and Codes

The binary relations, $\leq_h$, $\leq_i$, $\leq_p$ and $\leq_s$, are called the *embedding, infix, prefix*, and *suffix orders*, respectively.

In this section, we define an infinite hierarchy of reflective and antisymmetric binary relations on X^*. These relations, called *n-shuffle relations*, contain the embedding order and are contained in the infix order. They are not transitive except for $n = 1$ and their transitive closure is the embedding order. The corresponding independent sets form the classes of n-shuffle codes. The purpose of this section is to study these hierarchies of n-shuffle relations and n-shuffle codes. Several types of operations on these classes of codes are considered. Furthermore, some closure operations in relation with the notions of convexity associated with the embedding order are investigated.

2.3.1 n-Shuffle Relations

For every positive integer n, the binary relation ω_{sh_n} is defined on X^* by:

$$u\omega_{sh_n}v \Leftrightarrow u = u_1u_2\cdots u_n,\ v = v_0u_1v_1\cdots u_nv_n,\ \text{where } u_i, v_j \in X^*.$$

The relation ω_{sh_n} is ***reflexive*** and ***antisymmetric*** and it is called the *n-shuffle relation*. Clearly

$$\omega_{sh_1} \subseteq \omega_{sh_2} \subseteq \ldots \subseteq \omega_{sh_n} \subseteq \ldots$$

and this hierarchy is *strict*. For example, let $X = \{a, b\}$ and let $u = b^n$, $v = a(ba)^n$. Then we have $u\omega_{sh_n}v$ but not $u\omega_{sh_{n-1}}v$.

Lemma 2.29 *The following two statements hold:*

(1) For every $n \geq 1$, ω_{sh_n} is not compatible.

(2) The relation ω_{sh_1} is transitive, but for $n \geq 2$, ω_{sh_n} is not transitive.

Proof: (1) We have $a^n\omega_{sh_n}(ab)^n$, but not $a^na\omega_{sh_n}(ab)^na$.

(2) It is immediate that ω_{sh_1} is transitive. Let $n \geq 2$. We have $a^{n+1}b\omega_{sh_n}(ab)^{n-1}a^2b$ and $(ab)^{n-1}a^2b\omega_{sh_n}(ab)^{n-1}abab$, but not $a^{n+1}b\omega_{sh_n}(ab)^{n-1}abab$. □

Since ω_{sh_1} is transitive, it is a partial order and coincide with the infix order, that is, $\omega_{sh_1} = \leq_i$.

Theorem 2.30 *Let $u, \bar{u}, v, \bar{v}, w \in X^+$. Then:*

(1) $u\bar{u}\omega_{sh_n}v\bar{v}$ implies $u\omega_{sh_n}v$ or $\bar{u}\omega_{sh_n}\bar{v}$.

(2) $uw\omega_{sh_n}vw$ implies $u\omega_{sh_n}v$ and similarly $wu\omega_{sh_n}wv$ implies $u\omega_{sh_n}v$.

Proof: (1) Since $u\bar{u}\omega_{sh_n}v\bar{v}$, $u\bar{u} = u_1 \cdots u_n$, $v\bar{v} = v_0u_1v_1 \cdots u_nv_n$ for $u_i, v_j \in X^*$. This implies that $v = v_0u_1v_1 \cdots u_jv_jx$ where $x = u'$ or $x = u_{j+1}v'$, u' is a prefix of u_{j+1} and v' is a prefix of v_{j+1}, $j < n$. In the first case, $v = v_0u_1v_1 \cdots u_jv_ju'$. If $|u| \leq |u_1 \cdots u_ju'|$, then $u\omega_{sh_n}v$. Otherwise, $\bar{u}\omega_{sh_n}\bar{v}$. In the second case, $v = v_0u_1v_1 \cdots u_jv_ju_{j+1}v'$. If $|u| \leq |u_1 \cdots u_{j+1}|$, then $u\omega_{sh_n}v$. Otherwise, $\bar{u}\omega_{sh_n}\bar{v}$. Both cases imply that $u\omega_{sh_n}v$ or $\bar{u}\omega_{sh_n}\bar{v}$.

(2) Since $uw\omega_{sh_n}vw$, $uw = x_1x_2 \cdots x_n$ and $vw = y_0x_1y_1 \cdots x_ny_n$ where $x_i, y_j \in X^*$. There exist $i \leq n$ and $u', w' \in X^*$ such that $u'w' = x_i$, $u = x_1 \cdots x_{i-1}u'$ and $w = w'x_{i+1} \cdots x_n$. This implies that $w\omega_{sh_n}w'y_i \cdots x_ny_n$ and $u\omega_{sh_n}y_0x_1y_1 \cdots x_{i-1}y_{i-1}u'$. Now it is clear that $|w'y_i \cdots x_ny_n| \geq |w|$. Since

$$vw = y_0x_1y_1 \cdots x_{i-1}y_{i-1}u'w'y_ix_{i+1}y_{i+1} \cdots x_ny_n,$$

$|v| \geq |y_0x_1y_1 \cdots x_{i-1}y_{i-1}u'|$. That is, there exists $v' \in X^*$ such that

$$v = y_0x_1y_1 \cdots x_{i-1}y_{i-1}u'v'.$$

Since $i \leq n$, $u\omega_{sh_n}v$. □

Since ω_{sh_n} is reflective and anti-symmetric, trans ω_{sh_n} is a partial order relation. It is not difficult to show that $u \leq_h v$ if and only if $u\omega_{sh_n}v$ for all n. Moreover, we have the following theorem:

Theorem 2.31 *For every $n \geq 2$ one has* trans $\omega_{sh_n} = \leq_h$.

Proof: The fact that $\omega_{sh_n} \subseteq \leq_h$ implies that trans $\omega_{sh_n} \subseteq \leq_h$ as $\leq_h$ is transitive. On the converse, we want to show that $u \leq_h v$ implies that u trans $\omega_{sh_n}v$. Since $u \leq_h v$, $u = u_1u_2 \cdots u_k$, $v = v_0u_1v_1 \cdots u_kv_k$ for some $k \geq 1$. Let $m = k + 1$. Then $m \geq 2$ and $u\omega_{sh_m}v$.

If $m \leq n$, then the fact that $\omega_{sh_m} \subseteq \omega_{sh_n} \subseteq$ trans ω_{sh_n} implies that u trans $\omega_{sh_n}v$.

For the case $m > n$, we first show that $\omega_{sh_m} \subseteq \text{trans}\,\omega_{sh_{m-1}}$. Since $u\omega_{sh_m}v$,

$$u = u_1u_2\cdots u_m,\ v = v_0u_1v_1u_2\cdots u_{m-1}v_{m-1}u_mv_m.$$

Let $v' = v_0u_1v_1\cdots u_{m-1}u_mv_m$. Since $m > n \geq 2$, $m-1 \geq 2$. Hence $u\omega_{sh_{m-1}}v'$ and $v'\omega_{sh_{m-1}}v$. Thus $u\,\text{trans}\,\omega_{sh_{m-1}}v$. The fact that $u\,\text{trans}\,\omega_{sh_n}v$ is then obtained by decreasing the index m.

Therefore $\omega_n = \leq_h$. □

By the similarity with the definition of ω_{sh_n}, we define the following two binary relations π_n and φ_n on X^* by: for $u, v \in X^*$,

$$u\pi_nv \iff u = u_1u_2\cdots u_n,\ v = u_1v_1u_2v_2\cdots u_nv_n \quad \text{where} \quad u_i, v_j \in X^*;$$

$$u\varphi_nv \iff u = u_0u_1\cdots u_n,\ v = u_0v_1u_1\cdots u_{n-1}v_{n-1}u_n \quad \text{where} \quad u_i, v_j \in X^*.$$

By the definitions of ω_{sh_n}, π_n, φ_n, $\leq_h$, $\leq_p$, $\leq_o$ and $\leq_i$, it is not difficult to show the following hierarchy:

$$\pi_1 = \leq_p \subsetneq \omega_{sh_1} = \leq_i \subsetneq \pi_2 \subsetneq \ldots \subsetneq \pi_n \subsetneq \omega_{sh_n} \subsetneq \pi_{n+1} \subsetneq \ldots \subsetneq \leq_h;$$

$$\pi_1 = \leq_p \subsetneq \varphi_1 = \leq_o \subsetneq \pi_2 \subsetneq \ldots \subsetneq \pi_n \subsetneq \varphi_n \subsetneq \pi_{n+1} \subsetneq \ldots \subsetneq \leq_h.$$

Using those three relations, ω_{sh_n}, π_n and φ_n, and the concept of the independent sets of binary relations, we exhibit some subclasses of prefix codes, infix codes and outfix codes in the next section.

2.3.2 n-Shuffle Codes

A non-empty set $L \subseteq X^+$ of ω_{sh_n}-independent words is a code, called an *n-shuffle code*. Clearly every n-shuffle code is an m-shuffle code for every $m \leq n$. Because of the strict hierarchy of the relations ω_{sh_n}, an m-shuffle code is not in general an n-shuffle code for $m < n$. If L is an n-shuffle code for all n, then L is called a *hypercode* (see [Shy5]). Hence a hypercode is a $\leq_h$-independent set of words where $\leq_h$ is the transitive closure of ω_{sh_n} for any $n \geq 2$. Let $\mathcal{C}_{sh_n}$ be the class of all n-shuffle codes. Then we have the following strict hierarchy:

$$\mathcal{C}_h \subsetneq \ldots \subsetneq \mathcal{C}_{sh_n} \subsetneq \mathcal{C}_{sh_{n+1}} \subsetneq \ldots \subsetneq \mathcal{C}_p$$

It is well known that $\mathcal{C}_p$ is a free monoid with respect to the operation of catenation of languages and that $\mathcal{C}_h$ is a free submonoid of $\mathcal{C}_p$ (see [Ito4], [Shy1] and [Shy3])[1]. The next two theorems show that, for every positive integer n, $\mathcal{C}_{sh_n}$ is a submonoid of $\mathcal{C}_p$, but not a free submonoid.

Theorem 2.32 *$\mathcal{C}_{sh_n}$ is a submonoid of $\mathcal{C}_p$.*

Proof: Let $C_1, C_2 \in \mathcal{C}_{sh_n}$. If $C_1 = \{1\}$ or $C_2 = \{1\}$, then clearly $C_1C_2 \in \mathcal{C}_{sh_n}$. Now let $C_1 \neq \{1\}$ and $C_2 \neq \{1\}$. Suppose $C_1C_2 \notin \mathcal{C}_{sh_n}$. Then there exist $\{u\bar{u}, v\bar{v}\} \subseteq C_1C_2$, with $u, v \in C_1$ and $\bar{u}, \bar{v} \in C_2$ such that $u\bar{u}\omega_{sh_n}v\bar{v}$. Then by Theorem 2.30, $u\omega_{sh_n}v$ or $\bar{u}\omega_{sh_n}\bar{v}$. This implies that $C_1 \notin \mathcal{C}_{sh_n}$ or $C_2 \notin \mathcal{C}_{sh_n}$, a contradiction. Therefore, $C_1C_2 \in \mathcal{C}_{sh_n}$. □

We use Theorem 1.1, to show that $\mathcal{C}_{sh_n}$ is not free.

Theorem 2.33 *For every positive integer n, $\mathcal{C}_{sh_n}$ is not free.*

Proof: Consider an alphabet X such that $\{a, b\} \subseteq X$. Let $C_1 = \{(ab)^n a, b^n\}$ and let $C_2 = \{b\}$. Then $C_1 \notin \mathcal{C}_{sh_n}$ and $C_2 \in \mathcal{C}_{sh_n}$. But $C_1C_2, C_2C_1 \in \mathcal{C}_{sh_n}$. Thus $C_1 \in \mathcal{C}_{sh_n}{}^{-1}\mathcal{C}_{sh_n} \cap \mathcal{C}_{sh_n}\mathcal{C}_{sh_n}{}^{-1}$. This implies that $\mathcal{C}_{sh_n}$ is not free. □

It is well known that every hypercode over a finite alphabet is finite. However this is not the case for n-shuffle codes. For example let $X = \{a, b\}$ and, for every $n \geq 1$, let $L = \{(a^k b)^{n+1} a^k \mid k \geq 1\}$. Then L is an infinite n-shuffle code. In some special cases, finiteness conditions can be obtained for n-shuffle codes.

---

[1] Here we let $\mathcal{C}_h, \mathcal{C}_{sh_n}$ and $\mathcal{C}_p$ contain the set $\{1\}$. Then they form monoids.

Theorem 2.34 *Let X be an alphabet. Then:*

(1) Regular n-shuffle codes over X are finite for $n > 1$;

(2) Context-free n-shuffle codes over X are finite for $n > 2$.

Proof: This is a direct consequence of the pumping lemmata for regular and context-free languages and the definition of ω_{sh_n}. □

Theorem 2.35 *Let L be an infinite language over an alphabet X. Then there exists an integer m such that L is not a n-shuffle code for every $n \geq m$.*

Proof: Since every hypercode is finite and L is infinite, there exists an integer m such that L is not a m-shuffle code. The proof then follows from the fact that every s-shuffle code is a r-shuffle code for $r \leq s$. □

It follows from the Zorn's lemma that every n-shuffle code over the alphabet X is contained in a maximal n-shuffle code over X.

Theorem 2.36 *Let L be a code over the alphabet X. Then L is an n-shuffle code and maximal as a code if and only if L is a full uniform code.*

Proof: It is clear that every full uniform code is a maximal code and an n-shuffle code. Conversely, let L be a maximal code and an n-shuffle code. Then, clearly, L is a maximal prefix code because every n-shuffle code is a prefix code. Let $A = \{w \in L \mid |w| \leq |u| \text{ for all } u \in L\}$ and $up(A) = \cup_{x \in A} up(x)$ where $up(x) = \{y \in X^* \mid x\omega_{sh_n}y\}$. The elements in A have the same length, that is, $\emptyset \neq A \subseteq X^m$ for some m. Since L is a n-shuffle code, $(L \setminus A) \cap up(A) = \emptyset$, that is, $L \cap up(A) = A$. Suppose that L is not a full uniform code. Since L is a maximal prefix code and $L \neq X^k$ for all k, $L \setminus A \neq \emptyset$. As $A \subseteq X^m$ and $A \neq X^m$, one has $XA \not\subseteq AX$. For, if $XA \subseteq AX$, then $X^m A \subseteq AX^m$ by induction on m. Thus $A = X^m$, a contradiction! Let $w \in XA \setminus AX$. Then $wX^* \subseteq up(A) \setminus A$ and $wX^* \cap L = \emptyset$. Hence there exists no $v \in L$ such that $w = vx$ or $v = wx$ for some $x \in X^+$. Thus $L \cup \{w\}$ is a prefix code, a contradiction! Therefore, L is a full uniform code. □

In [Ito4], it has been shown that many subclasses of maximal codes are, under certain conditions, just subclasses of uniform codes. The previous theorem shows that an n-shuffle code is maximal as a code if and only if it is a uniform code. The connection between a specific code which is maximal as a code and a maximal specific code is a very interesting topic for further research.

For any binary relation ϱ defined on X^*, a language L over X is said to be *right (left) ϱ-convex* if $u\varrho v$ ($v\varrho u$) with $u \in L$ implies $v \in L$. The language L is said to be *ϱ-convex* if $u\varrho w$ and $w\varrho v$ with $u, v \in L$ implies $w \in L$.

Theorem 2.37 *If a language is right or left ω_{sh_n}-convex with $n \geq 2$, then it is $\leq_h$-convex.*

Proof: Suppose that L is right ω_{sh_n}-convex and let $u \leq_h v$ with $u \in L$. Then $u = u_1u_2\cdots u_k$ and $v = x_0u_1x_1u_2x_2\cdots u_kx_k$. Let $v_0 = x_0u_1u_2\cdots u_k$, $v_1 = x_0u_1x_1u_2u_3\cdots u_k$, $v_2 = x_0u_1x_1u_2x_2u_3u_4\cdots u_k, \ldots, v_k = v$. Since L is right ω_{sh_n}-convex, L is right ω_{sh_2}-convex. From $u\omega_{sh_2}v_0, v_0\omega_{sh_2}v_1, \ldots, v_{k-1}\omega_{sh_2}v_k$ and $v_k = v$ it follows that $v_0 \in L, v_1 \in L, \ldots, v_k = v \in L$. Hence L is $\leq_h$-convex. The proof is similar if L is left ω_{sh_n}-convex. □

The previous theorem is not true for $n = 1$. For example $L = \{a^2, aba^2\}$ over $X = \{a, b\}$ is right ω_{sh_1}-convex but not $\leq_h$-convex. A ω_{sh_n}-convex language is not $\leq_h$-convex in general. For example, the language $L = \{(ab)^n aba, (a^2b)^n a^2ba^2\}$ is ω_{sh_n}-convex, but not $\leq_h$-convex. However, a $\leq_h$-convex language is ω_{sh_n}-convex for all n.

Theorem 2.38 *Let $L \subseteq X^+$ be a nonempty language and let $n \geq 2$. If L is ω_{sh_n}-convex and an outfix code, then L is an n-shuffle code.*

Proof: Suppose there exist $u, v \in L$ such that $u\omega_{sh_n}v$, that is, $u = u_1u_2\cdots u_n$ and $v = v_0u_1v_1u_2v_2\cdots u_nv_n$, $u_i, v_j \in X^*$. Then for every $i, i = 0, 1, \ldots, n$, we have:

$$u_1u_2\cdots u_n\omega_{sh_n}u_1u_2\cdots u_iv_iu_{i+1}\cdots u_n,\ u_1u_2\cdots u_iv_iu_{i+1}\cdots u_n\omega_{sh_n}v.$$

Since L is ω_{sh_n}-convex, $u_1 \cdots u_i v_i u_{i+1} \cdots u_n \in L$ and hence $v_i = 1$. Therefore, $u = v$ and L is n-shuffle. □

The above theorem is not true for $n = 1$. For example, $L = \{a^2, aba\}$ over $X = \{a, b\}$ is an 1-shuffle code but not an outfix code.

Let π_n and φ_n be defined as in preceding section. It is not difficult to show that, for every $n \geq 1$, $\mathcal{L}_{\pi_n}$ and $\mathcal{L}_{\varphi_n}$ are two subclasses of prefix codes. Moreover, we have the following hierarchy:

$$\mathcal{C}_h \subsetneq \ldots \subsetneq \mathcal{L}_{\pi_{n+1}} \subsetneq \mathcal{C}_{sh_n} \subsetneq \mathcal{L}_{\pi_n} \subsetneq \ldots \subsetneq \mathcal{L}_{\pi_2} \subsetneq \mathcal{C}_{sh_1} = \mathcal{C}_i \subsetneq \mathcal{L}_{\pi_1} = \mathcal{C}_p;$$

$$\mathcal{C}_h \subsetneq \ldots \subsetneq \mathcal{L}_{\pi_{n+1}} \subsetneq \mathcal{L}_{\varphi_n} \subsetneq \mathcal{L}_{\pi_n} \subsetneq \ldots \subsetneq \mathcal{L}_{\pi_2} \subsetneq \mathcal{L}_{\varphi_1} = \mathcal{C}_o \subsetneq \mathcal{L}_{\pi_1} = \mathcal{C}_p.$$

These hierarchies clarify connections between hypercodes, prefix codes, infix codes and outfix codes.

Since regular outfix codes are finite, every regular language in $\mathcal{L}_{\pi_n}$, $\mathcal{C}_{sh_n}$, $\mathcal{L}_{\varphi_n}$ where $n > 1$ is finite. By Theorem 2.34, every context-free n-shuffle code over X is finite for $n > 2$. Similarly, every context-free language in $\mathcal{L}_{\pi_n}$, $\mathcal{C}_{sh_n}$, $\mathcal{L}_{\varphi_n}$ where $n > 2$ is finite.

2.3.3 The Syntactic Monoid of n-Shuffle Codes

In the sequel, a monoid M is called ***non-trivial*** if it contains at least three elements. An element $u \in M$ is called ***n-strict*** with n a positive integer if the equality

$$x_0 u_1 x_1 \cdots u_n x_n = u_1 \cdots u_n = u$$

with $x_i, u_j \in M$ implies $x_0 = x_1 = \ldots = x_n = e$ where e is the identity of the monoid M. The identity e is said to be ***isolated*** if $M \setminus \{e\}$ is a subsemigroup of M.

Given a language $L \subseteq X^*$, the ***principal congruence*** P_L determined by L is defined by the following condition:

$$\forall x, y \in X^* : \ u \equiv v(P_L) \text{ if and only if } (xuy \in L \iff xvy \in L).$$

The monoid X^*/P_L is called the ***syntactic monoid*** of L and denoted by $\mathrm{Syn}(L)$. Recall that a monoid M is ***subdirectly irreducible*** if it has a unique minimal congruence (see [Sch1] and [Thi1]).

The following theorem is a generalization of the characterization of the syntactic monoid of infix codes and hypercodes (see [Pet1], [Jür4] and [Shy1]).

Theorem 2.39 *A monoid M is isomorphic with the syntactic monoid* $\mathrm{Syn}(L)$ *of an n-shuffle code L if and only if M is a non-trivial subdirectly irreducible monoid with an isolated identity and a non-zero n-strict disjunctive element.*

Proof: Let P_L be the syntactic congruence of L and let $M = \mathrm{Syn}(L)$. Since L is n-shuffle, L is an infix code and, by [Jür4], $\mathrm{Syn}(L)$ is a subdirectly irreducible monoid. Furthermore L is a class of P_L and the residue W of L is non-empty and, hence, also a class of P_L. Since W is an ideal of X^*, $W \neq L$. It is also immediate that the class e of the empty word 1 is distinct from W and L. Hence M is non-trivial. Let x be an element of the class of 1. Then $x \equiv 1(P_L)$ and $xc \equiv c(P_L)$ for every $c \in L$. Hence $c, xc \in L$ and, since L is n-shuffle, $x = 1$. Therefore $e = \{1\}$ and the identity e of M is isolated.

Since L is an n-shuffle code and also a class of the syntactic congruence P_L, it follows then that $\mathrm{Syn}(L)$ contains a non-zero n-strict element, the class of L, and that L is a disjunctive element of $\mathrm{Syn}(L)$.

For the converse, as M is subdirectly irreducible it has a unique minimal congruence and by [Sch1] it contains at least two different disjunctive elements. Let 0 and e be respectively the zero and the identity elements of M and let c be a non-zero n-strict disjunctive element of M. Since M contains more than two elements and e is isolated, e cannot be disjunctive. Hence, c is different from e.

Let X be an alphabet such that $|X| = |M \setminus \{e\}|$ and let φ be a bijection of X onto $M \setminus \{e\}$. As usual, we extend φ (in a unique fashion) to a homomorphism of X^* onto M. The relation $\vartheta = \varphi\varphi^{-1}$ is a congruence of X^* such that the quotient monoid X^*/ϑ is isomorphic with M. Let $L = \{u \mid u \in X^*,\ u\varphi = c\}$. As c is disjunctive one has $\vartheta = P_L$ (see [Thi2] for example).

We need to show that L is n-shuffle. Clearly $L \subseteq X^+$. Let $u\omega_{sh_n}v$, $u, v \in L$. Then

$$u = u_1u_2\cdots u_n,\ v = x_0u_1x_1u_2x_2\cdots u_nx_n \text{ with } x_i, u_j \in X^*.$$

Since $u, v \in L$, $u\varphi = v\varphi = c$ and $u_1\varphi\cdots u_n\varphi = x_0\varphi u_1\varphi\cdots u_n\varphi x_n\varphi = c$. Since $c \neq 0$ and c is n-strict, we must have $x_0\varphi = x_1\varphi = \ldots = x_n\varphi = e$. Therefore $x_0 = x_1 = \ldots = x_n = 1$ and $u = v$.

Since M is isomorphic with X^*/ϑ and since $\vartheta = P_L$, M is isomorphic with the syntactic monoid $\mathrm{Syn}(L)$ of the n-shuffle code L. □

Theorem 2.40 *If L is a $n+2$-shuffle code with $n \geq 1$, then every non-zero element of the syntactic monoid* $\mathrm{Syn}(L)$ *is n-strict.*

Proof: Let $\bar{u} = \bar{x}_0\bar{u}_1\bar{x}_1\bar{u}_2\cdots\bar{u}_n\bar{x}_n = \bar{u}_1\bar{u}_2\cdots\bar{u}_n \neq 0$, where $\bar{x}$ denotes the P_L-class of x, that is

$$u \equiv x_0u_1x_1u_2x_2\cdots u_nx_n \equiv u_1u_2\cdots u_n(P_L).$$

Since $\bar{u} \neq 0$, $u \notin W$, where W is the residue of L. Therefore there exist $x, y \in X^*$ such that $v = xx_0u_1x_1u_2x_2\cdots u_nx_ny \in L$, hence $w = xu_1u_2\cdots u_ny \in L$ because L is a P_L-class. It follows then that $w\omega_{sh_{n+2}}v$ with $w, v \in L$. Since L is $n+2$-shuffle, we have $w = v$, $x_0 = x_1 = x_2 = \ldots = x_n = 1$. Therefore, $\bar{x}_0 = \bar{x}_1 = \ldots = \bar{x}_n = e$ and $\bar{u}$ is n-strict. Therefore, every non zero element of $\mathrm{Syn}(L)$ is n-strict. □

CHAPTER 3

Solid Codes

3.1 Solid Codes

Solid codes can be used for information transmission over a noisy channel to allcw one to decode the correct parts of a disturbed message correctly. For this reason we call them solid codes. In this chapter, a characterization of solid codes is given and some closure and non-closure properties of the class of all solid codes are studied. Moreover, some conditions for a solid code to be maximal are also investigated in this chapter.

Shyr and the author introduced solid codes in [Shy6]. Some properties of solid codes with cardinality less than or equal to 2 were already studied in [Shy6]. First, we give some basic notions and definitions as follows:

For $x, y \in X^*$, if $x = uyv$ for some $u, v \in X^*$, then we call y a *factor* of x. For any $x \in X^+$, let $u, v \in X^*$ such that $x = uv$. Then u is called a *left factor* of x, and v is called a *right factor* of x. We define the proper left factor set $P(x)$, the proper right factor set $S(x)$ and the factor set $E(x)$ of x as

$$P(x) = \{y \in X^+ \mid x \in yX^+\},$$

$$S(x) = \{y \in X^+ \mid x \in X^+y\},$$

and

$$E(x) = \{y \in X^+ \mid x \in X^*yX^*\}.$$

Given a set $L \subseteq X^+$, any word $w \in X^+$ can be represented as follows:

$$w = x_1y_1x_2y_2 \ldots x_ny_nx_{n+1},$$

where $y_j \in L$, $j = 1, 2, \ldots, n$, $E(x_i) \cap L = \emptyset$, $i = 1, 2, \ldots, n+1$. If $E(w) \cap L = \emptyset$ or $L = \emptyset$, then we let $w = x_1$. Any such representation of w is called an *L-representation of w*.

Definition 3.1 *A non-empty language $L \subseteq X^+$ is said to be solid if for every $w \in X^+$ there is a unique L-representation of w.*

The following theorem justifies the terminology.

Theorem 3.2 *Let L be a solid language over X. Then L is a code.*

Proof: Suppose that L is a solid language, but not a code. Then there are words $x_1, \ldots, x_n, y_1, \ldots, y_m \in L$ such that

$$x_1 \cdots x_n = y_1 \cdots y_m$$

with $n \neq m$ or $x_i \neq y_i$ for some i, $1 \leq i \leq n$. But then $x_1 \cdots x_n$ and $y_1 \cdots y_m$ are two different L-representations of the same word, a contradiction! □

In view of Theorem 3.2, we use the term "solid code" to denote a solid language. By the definition of codes, the unique factorization property implies that the factorization of a string composed from code words is unique. For example, the language $L = \{ab, ba\}$ is a uniform code. The factorization is unique for any word over L. But for an arbitrary word there may be more than one parsing. For example, $aba = (ab)a = a(ba)$, and $a \notin L$. Thus aba has two distinct L-representations. Such a situation may arise, for instance, when a noisy channel inserts symbols into the message stream. In that case, a partial but unique factorization would be an important goal. With solid codes, the factorization is unique for arbitrary words.

The following result provides a necessary and sufficient condition for a language L to be solid. It generalizes a characterization obtained in [Shy6] for the case of $|L| \leq 2$.

Theorem 3.3 *A language L over X is a solid code if and only if every two words $u, v \in L$ satisfy the following conditions:*

(a) $P(u) \cap S(v) = \emptyset$.

(b) If $u \neq v$ then $u \notin E(v)$ and $v \notin E(u)$.

Proof: For $|L| \leq 2$, this is the result of [Shy6].

First assume that L is solid. Every subset L' of L is solid. In particular, this holds true for every subset L' with $|L'| \leq 2$, that is, by [Shy6], the conditions (a) and (b) are satisfied for L'. This proves the necessity of the conditions.

For the converse, let L be a language which satisfies the conditions (a) and (b) and which is not a solid code. By the above, $|L| > 2$. As L is not solid there is a word $w \in X^*$ which has two different L-representations

$$w = x_1 y_1 \cdots x_n y_n x_{n+1} = x'_1 y'_1 \cdots x'_m y'_m x'_{m+1}$$

where

$$y_1, \ldots, y_n, y'_1, \ldots, y'_m \in L$$

and

$$E(x_i) \cap L = \emptyset = E(x'_j) \cap L$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. We may assume that $x_1 y_1 \neq x'_1 y'_1$. Otherwise we could cancel these factors and obtain a proper suffix of w which also has two different L-representations.

From the definition of L-representations, one concludes that

$$y_1 \notin E(x'_1) \text{ and } y'_1 \notin E(x_1).$$

The language $\{y_1\} \cup \{y'_1\}$ satisfies the conditions (a) and (b).

If $y_1 \neq y'_1$, then $y_1 \notin E(y'_1)$ and $y'_1 \notin E(y_1)$ by condition (b). This is impossible if $|x_1| = |x'_1|$. Therefore, $S(y'_1) \cap P(y_1) \neq \emptyset$ or $S(y_1) \cap P(y'_1) \neq \emptyset$, contradicting condition (a).

Now assume that $y_1 = y'_1$. If $|x_1| = |x'_1|$ then one has $x_1y_1 = x'_1y'_1$ contrary to the assumptions. Otherwise, again condition (a) is violated. □

Corollary 3.4 *The class of solid codes is a proper subclass of the class of infix codes.*

Proof: Infix codes can be characterized as $\leq_i$-independent sets where

$$u \leq_i v \iff v \in X^*uX^*$$

for $u, v \in X^*$. This is just the above condition (b). On the other hand, the set $\{ab, ba\}$ is an infix code which is not solid. □

Recall that a word $u \in X^+$ is called ***unbordered*** if no proper nonempty left factor of u is a right factor of u. In other words, u is unbordered if and only if $P(u) \cap S(u) = \emptyset$, that is, $\{u\}$ is a solid code. In the following examples we indicate how sets which do not satisfy the conditions mentioned in Theorem 3.3 fail to be solid codes.

Examples 3.5

(1) $\{a^2\}$ *is not a solid code by Theorem 3.3. Indeed, for example* $a^3 = a(a^2) = (a^2)a$.

(2) $L = \{a^2b^3, ab\}$ *is not a solid code according to Theorem 3.3; for instance, the word* a^3b^3 *has two L-representations:* $a^3b^3 = a(a^2b^3) = a^2(ab)b^2$.

(3) By Theorem 3.3, the languages $\{ab\}$ *and* $\{a^2b^3, abab^3\}$ *are solid codes.*

By a remark of [Shy6], a solid code can contain only primitive words. If a set $L \subseteq X^*$ is an infix code and $P(x) \cap S(y) = \emptyset$ for all $x, y \in L$, then L is called a strong infix code (see [Ito1]). Using Theorem 3.3, this proves the following result:

Corollary 3.6 *A language* $L \subseteq X^+$ *is a solid code if and only if it is a strong infix code.*

Like prefix codes, infix codes, outfix codes, etc., solid codes are another example of a class C of codes whose definition can be expressed in terms of a property of the subsets of size at most 2. This is particularly important with respect to deciding the respective code property. Clearly, if L is a finite language, then for any such class C it is decidable whether $L \in C$. The statement can be extended to regular languages.

Theorem 3.7 *Let L be a regular language over X. It is decidable whether or not L is a solid code.*

Proof: By [Ito4], one can decide, whether L is an infix code. If L is not an infix then it is not solid.

Otherwise, let

$$\mathrm{Pref}(L) = \{u \mid u \in X^+,\ uX^* \cap L \neq \emptyset\}$$

and

$$\mathrm{Suff}(L) = \{u \mid u \in X^+,\ X^*u \cap L \neq \emptyset\}.$$

These sets are known to be regular for regular L. Moreover,

$$P(L) = \bigcup_{w \in L} P(w) = \mathrm{Pref}(L) \setminus L$$

and

$$S(L) = \bigcup_{w \in L} S(w) = \mathrm{Suff}(L) \setminus L$$

as L is an infix code.

Indeed, if $u \in P(L)$ then $u \in X^+$ and $w = uv$ for some $w \in L$ and $v \in X^+$. Hence $u \in \mathrm{Pref}(L)$. As L is an infix code and $u \leq_i w$ it follows that $u \notin L$. The converse inclusion is obvious. The equality concerning $S(L)$ is obtained dually.

The equalities show that the sets $P(L)$ and $S(L)$ are regular and that automata can be constructed for them from an automaton accepting L.

Now let $w \in P(L) \cap S(L)$. Then there are words $u, v \in L$ such that $w \in P(u)$ and $w \in S(v)$. Hence, L is not solid. On the other hand, if $P(L) \cap S(L) = \emptyset$ then every two words $u, v \in L$ have $P(u) \cap S(v) = \emptyset$.

Thus, in order to decide whether L is solid one decides whether the intersection $P(L) \cap S(L)$ is empty. The decidability of this latter question for regular sets is known from automata theory (see [Woo1], for example). □

3.2 Closure Properties

In this section we discuss closure properties of the class of solid codes. It terns out to be closed under reversal, arbitrary intersections, inverse non-erasing homomorphisms, and a restricted kind of products, while it is not closed under union, complement, product, catenation closure, and (non-erasing) homomorphisms. The first two properties are immediate consequences of the definitions.

Theorem 3.8 *The class of solid codes is closed under reversal and intersection with arbitrary sets, and not closed under union, complement, product, catenation closure, homomorphisms, and non-erasing homomorphisms.*

Proof: By the definition of solid codes, it is clear that the class of solid codes is closed under reversal and intersection with arbitrary sets. Now, consider the solid codes $\{a\}$ and $\{ab\}$. The sets $\{a, ab\} = \{a\} \cup \{ab\}$, $\{aba\} = \{ab\}\{a\}$, and $\{a\}^+$ are not solid. This proves the statement for union, product, and catenation closure. Moreover, as the class of solid codes is closed under intersections, but not closed under unions, therefore it is also not closed under complement.

Consider the alphabets $X = \{a\}$ and $Y = \{a, b\}$ and let h be the homomorphism of X^* into Y^* which is defined by $h(a) = aba$. h is non-erasing and maps the solid code $\{a\}$ onto $\{aba\}$, the latter not a solid code. Thus, the class of solid codes is not closed under non-erasing homomorphisms nor — a fortiori — under arbitrary homomorphisms. □

While the class of solid codes is not closed under arbitrary non-erasing homomorphisms in general, the class of homomorphisms which preserve solid codes can be characterized effectively.

Theorem 3.9 *Let X and Y be alphabets. A non-erasing homomorphism $h : X^* \to Y^*$ preserves solid codes if and only if its restriction to X is injective and for any $a, b \in X$, $a \neq b$, the sets $\{h(a)\}$ and $\{h(a), h(b)\}$ are solid. Therefore, for a given non-erasing homomorphism h it is decidable whether it preserves solid codes.*

Proof: Assume that h preserves solid codes. Consider $a, b \in X$. As $\{a\}$ is solid, it follows that $\{h(a)\}$ is solid. Similarly, if $a \neq b$, then $\{h(a), h(b)\}$ is solid as $\{a, b\}$ is solid. Finally, if $a \neq b$ then $\{ab\}$ is solid and, therefore, also $\{h(a)h(b)\}$ is solid. This implies $h(a) \neq h(b)$.

For the converse, assume that h satisfies the condition, but does not preserve solid codes. Let L be a solid code over X with $h(L)$ not solid. Hence there are words $u, v \in L$ such that one of the following statements holds true:

(1) $h(u) \neq h(v)$ and $h(u) \in E(h(v))$;

(2) $P(h(u)) \cap S(h(v)) \neq \emptyset$.

Let $u = u_1 \cdots u_n$ and $v = v_1 \cdots v_m$ with $u_1, \ldots, u_n, v_1, \ldots, v_m \in X$.

Consider case 1: Then $u \neq v$. The fact that L is solid implies that there are no $x, y \in X^*$ with $v = xuy$ or $u = xvy$. Let k be such that

$$h(u) \notin E(h(v_1 \cdots v_{k-1})) \quad \text{and} \quad h(u) \in E(h(v_1 \cdots v_k)).$$

We show that, for $j = n-1, n-2, \ldots, 0$, one has $u_{j+1} \cdots u_n = v_{k-(n-j)+1} \cdots v_k$ and $h(u_1 \cdots u_j)$ is a suffix of $h(v_1 \cdots v_{k-(n-j)})$.

Let $j = n-1$. If $h(u_1 \cdots u_{n-1})$ is not a suffix of $h(v_1 \cdots v_{k-1})$ then $P(h(v_k)) \cap S(h(u_n)) \neq \emptyset$ or $h(u_n) \in E(h(v_k))$ contradicting the assumptions about h. This shows that $h(u_1 \cdots u_{n-1})$ is a suffix of $h(v_1 \cdots v_{k-1})$. Consequently, if $h(u_n) \neq h(v_k)$

then $h(u_n) \in P(h(v_k))$, which is impossible by the assumptions about h. Therefore, $h(u_n) = h(v_k)$, and this implies $u_n = v_k$.

Now assume that $u_{j+1} \cdots u_n = v_{k-(n-j)+1} \cdots v_k$ and that $h(u_1 \cdots u_j)$ is a suffix of $h(v_1 \cdots v_{k-(n-j)})$ where $0 < j < n$. If $h(u_j) \neq h(v_{k-(n-j)})$ then $h(u_j) \in S(h(v_{k-(n-j)}))$ or $h(v_{k-(n-j)}) \in S(h(u_j))$, contradicting the assumptions about h. Therefore, $h(u_j) = h(v_{k-(n-j)})$, and this implies $u_j = v_{k-(n-j)}$. Thus,

$$u_j \cdots u_n = v_{k-(n-j)} \cdots v_k,$$

and $h(u_1 \cdots u_{j-1})$ is a suffix of $h(v_1 \cdots v_{k-(n-j)-1})$.

By induction on j, this proves that

$$u_1 \cdots u_n = v_{k-n+1} \cdots v_k,$$

that is, $u \in E(v)$ contradicting the solidity of L. This shows that case 1 is impossible.

Now consider case 2: As L is solid, one has $P(u) \cap S(v) = \emptyset$. Let w be the longest word in $P(h(u)) \cap S(h(v))$. Let r be maximal such that $h(u_1 \cdots u_r)$ is a prefix of w, that is, $w = h(u_1 \cdots u_r)w'$ for some $w' \in Y^*$ where $|w'| < |u_{r+1}|$. The assumptions about h imply that $h(v_m) \notin E(h(u_{r+1}))$. Therefore, if $|w'| > 0$ then $w' \in P(h(u_{r+1})) \cap S(h(v_m))$ contradicting the assumptions about h. This implies $|w'| = 0$, that is, $w = h(u_1 \cdots u_r)$ and $h(u_1 \cdots u_r)$ is a suffix of $h(v)$.

As in case 1, one shows by induction that $u_1 \cdots u_r = v_{m-r+1} \cdots v_m$, that is, $P(u) \cap S(v) \neq \emptyset$ contradicting the solidity of L. Hence, also case 2 is impossible.
□

Theorem 3.10 *The class of solid codes is closed under inverse non-erasing homomorphisms.*

Proof: Let X and Y be alphabets and let $h : X^* \to Y^*$ be a non-erasing homomorphism. It suffices to show that for any $u, v \in Y^*$, if $\{u\} \cup \{v\}$ is solid, then $\{\overline{u}\} \cup \{\overline{v}\}$ is solid for every $\overline{u} \in h^{-1}(u)$ and $\overline{v} \in h^{-1}(v)$.

Suppose that for some such $\overline{u}$ and $\overline{v}$ this is not true. If $\overline{u} \neq \overline{v}$ and $\overline{u} \in E(\overline{v})$ then $u \neq v$ and $u \in E(v)$, a contradiction! The case of $\overline{u} \neq \overline{v}$ and $\overline{v} \in E(\overline{u})$ is analogous. Otherwise, if $P(\overline{u}) \cap S(\overline{v}) \neq \emptyset$ or $P(\overline{v}) \cap S(\overline{u}) \neq \emptyset$ then $P(u) \cap S(v) \neq \emptyset$ or $P(v) \cap S(u) \neq \emptyset$, respectively, again a contradiction! □

We conclude this section with a theorem stating that the class of solid codes is closed under a restricted kind of product.

Theorem 3.11 *Let A, B be disjoint languages over X. If $A \cup B$ is solid then AB is solid.*

Proof: The statement is obviously true if one of A and B is empty. Therefore, suppose that $A \neq \emptyset \neq B$, and assume that AB is not solid. Then there exist $u_A, v_A \in A$ and $u_B, v_B \in B$ such that the set $\{u_A u_B, v_A v_B\}$ is not solid. Hence for some $x, y \in \{u_A u_B, v_A v_B\}$ one has $P(x) \cap S(y) \neq \emptyset$ or, assuming $x \neq y$, $x \in E(y)$. This implies the existence of $x', y' \in \{u_A, v_A\} \cup \{u_B, v_B\}$ such that $P(x') \cap S(y') \neq \emptyset$ or, with $x' \neq y'$, $x' \in E(y')$. But this contradicticts the fact that $\{u_A, v_A\} \cup \{u_B, v_B\}$ is a subset of the solid code $A \cup B$ and, hence, itself solid. □

3.3 Maximal Solid Codes

In this section we study properties of maximal solid codes. The questions addressed include the following: For which cardinalities do there exist solid codes? Is every finite solid code contained in a finite maximal solid code? How does one construct maximal solid codes? Note that by Zorn's lemma every solid code L is contained in a maximal solid code L'. However, if L is finite, can L' be chosen finite? Is there a situation when L' is uniquely determined by L? The results of this section give partial answers to some of these questions.

First we consider the case of an alphabet of size 2. Assume that $X = \{a, b\}$. The following is proved in [Shy6].

Lemma 3.12 ([Shy6]) *Let $X = \{a, b\}$ and $u, v \in X^+$ such that the set $\{u\} \cup \{v\}$ is solid. The following properties obtain:*

If $|u| > 1$ and $|v| > 1$, then either $u, v \in aX^*b$ or $u, v \in bX^*a$. If, in addition, $u \neq v$, then $|u| \geq 4$ and $|v| \geq 4$. If $u, v \in aX^*b$ then $u, v \notin a^+b \cup ab^+$.

We use Lemma 3.12 to prove the maximality of certain solid codes.

Theorem 3.13 *Let $X = \{a, b\}$ and let k be a positive integer. The language*

$$L_k = \{aba^ib^2 \mid i = 1, 2, \ldots, k\} \cup \{a^{k+1}b^2a^ib^2 \mid i = 1, 2, \ldots, k\} \cup \{a^{k+1}b^3\}$$

is a maximal solid code over X.

Proof: One verifies, that for every $u, v \in L_k$ the set $\{u\} \cup \{v\}$ is solid. Hence, L_k is solid.

Suppose, L_k is not maximal. Then there is a word $w \in X^+ \setminus L_k$ such that $L_k \cup \{w\}$ is solid. Lemma 3.12 implies that w has the form

$$w = a^{r_1}b^{s_1}a^{r_2}b^{s_2}\cdots a^{r_n}b^{s_n}$$

for some n, $n > 0$, and some r_i, s_i with $r_i > 0$ and $s_i > 0$ for $i = 1, 2, \ldots, n$. The proof consists of three 'steps.'

As a first step, we observe that $s_i \leq 2$ or $r_i \leq k$ for $i = 1, \ldots, n$. Otherwise $a^{k+1}b^3 \leq_i w$ and, therefore, $L_k \cup \{w\}$ is not solid, a contradiction!

In the second step, we show that $s_j \neq 2$ or $r_i \leq k$ for $i = 1, \ldots, n$. Suppose this is not true, and let j be maximal with $s_j = 2$ and $r_j \geq k + 1$.

If $j = n$ then $w = ua^{k+1}b^2$. Thus $a^{k+1}b^2 \in P(a^{k+1}b^2ab^2) \cap S(w)$ where $a^{k+1}b^2ab^2 \in L_k$. Hence, $L_k \cup \{w\}$ is not solid, a contradiction! This proves that $j < n$.

Let $j < m \leq n$ and assume that $r_m \geq k+1$. Then, by the result of the first step above, $s_m \leq 2$. The choice of j implies that $s_m = 1$. This shows that $s_m = 1$ or $r_m \leq k$ for $m = j+1, \ldots, n$.

Suppose that $r_{j+1} \leq k$ and $s_{j+1} \geq 2$. Then $a^{k+1}b^2a^{r_{j+1}}b^2 \in L_k \cap E(w)$, a contradiction!. Thus $r_{j+1} \geq k+1$ or $s_{j+1} = 1$. On the other hand, as shown above, $r_{j+1} \leq k$ or $s_{j+1} = 1$ must hold true, too. This implies that $s_{j+1} = 1$.

Let m be maximal with $j < m \leq n$ and $s_i = 1$ for $i = j+1, \ldots, m$. If $m = n$ then $ab \in S(w) \cap P(abab^2)$ where $abab^2 \in L_k$, a contradiction! Therefore, $m < n$ and $s_{m+1} \geq 2$, $r_{m+1} \leq k$. Then $aba^{r_{m+1}}b^2 \in E(w) \cap L_k$, a contradiction!

We have shown that the assumption about j leads to a contradiction in every case. This proves that $s_i \neq 2$ or $r_i \leq k$ for $i = 1, \ldots, n$.

As a third step, we prove that $s_i \neq 1$ for $i = 1, \ldots, n$. Suppose, $s_i = 1$ for some i, and let m be maximal with $s_m = 1$. If $m = n$ then $ab \in S(w) \cap P(abab^2)$ where $abab^2 \in L_k$, a contradiction. Therefore, $m < n$, $s_{m+1} \geq 2$. If $r_{m+1} \leq k$ then $aba^{r_{m+1}}b^2 \in E(w) \cap L_k$, a contradiction. Therefore, $r_{m+1} \geq k+1$ and $s_{m+1} \leq 2$ by the result of the first step of this proof. Hence, $s_{m+1} = 2$, contradicting the result of the second step. This shows that $s_i \neq 1$ for all i.

When combined, the results of the three steps imply that

$$s_i \geq 2 \quad \text{and} \quad r_i \leq k$$

for $i = 1, \ldots, n$. Thus, $a^{r_1}b^2 \in P(w) \cap S(aba^{r_1}b^2)$ where $aba^{r_1}b^2 \in L_k$, a contradiction! This completes the proof. □

Corollary 3.14 *For every odd number $n \geq 3$ there is a maximal solid code L with $|L| = n$ over $X = \{a, b\}$.*

As there are infinite solid codes, Zorn's lemma implies the existence of infinite maximal solid codes. The following theorem provides an example.

Theorem 3.15 *The language $L = \{aba^ib^2 \mid i = 1, 2, \ldots\}$ is a maximal solid code over $X = \{a, b\}$.*

Proof: For any $u, v \in L$ one verifies that $\{u\} \cup \{v\}$ is solid; hence, L is solid. Now suppose that $L \cup \{w\}$ is solid for some word $w \notin L$. Then w has the form

$$w = a^{r_1}b^{s_1}a^{r_2}b^{s_2}\cdots a^{r_n}b^{s_n}$$

for some n, $n > 0$, and some r_i, S_i with $r_i, s_i \geq 1$ for $i = 1, \ldots, n$.

If $s_1 \geq 2$ then $a^{r_1}b^2 \in P(w) \cap S(aba^{r_1}b^2)$ and $aba^{r_1}b^2 \in L$, a contradiction! Therefore, $s_1 = 1$.

Suppose there is an $s_i \geq 2$, and let m be minimal such that $s_m \geq 2$. Then $aba^{r_m}b^2 \in E(w) \cap L$, a contradiction! Therefore, $s_i = 1$ for all i. But then $ab \in P(abab^2) \cap S(w)$ and $abab^2 \in L$, again a contradiction! Hence $L \cup \{w\}$ is not solid. □

As is quite common in the theory of codes, the assumption that $|X| = 2$ renders constructions more complicated than they would be for $|X| > 2$. For example, let $X = \{a, b, c\}$. For every $k \geq 2$ the language

$$M_k = \{ac^ib \mid i = 0, \ldots, k\} \cup \{c^{k+1}b\}$$

is a maximal solid code, and $|M_k| = k+2$. Similarly, when $|X| \geq 3$, infinite maximal solid codes can be obtained quite easily. One general method for constructing certain maximal solid codes is provided by the following theorem.

Theorem 3.16 *Let $|X| \geq 2$ and let A, B, C be mutually disjoint subsets of X such that $A \neq \emptyset \neq C$ and $A \cup B \cup C = X$. Then the language AB^*C is a maximal solid code.*

Proof: Clearly AB^*C is solid. Consider a word $w \notin AB^*C$ where $w = x_1 \cdots x_n$ with $x_i \in X$ for $i = 1, \ldots, n$. We show that $AB^*C \cup \{w\}$ is not solid.

Assume there is an index i with $x_i \in A$ and, in fact, let i be maximal with this property. If $i = n$ then $x_i \in P(v) \cap S(w)$ for $v \in AB^*C$, hence $AB^*C \cup \{w\}$ is not solid.

If $i < n$ then $x_{i+1}, \ldots, x_n \in B \cup C$. If $x_{i+1} \in C$ then $x_i x_{i+1} \in AB^*C \cap E(w)$ and $AB^*C \cup \{w\}$ is not solid.

Therefore, assume that $x_{i+1} \in B$. If $x_{i+1} \cdots x_n \in B^+$ then $x_i x_{i+1} \cdots x_n \in P(v) \cap S(w)$ for some $v \in AB^*C$, and $AB^*C \cup \{w\}$ is not solid.

Thus, there is an index j with $i + 1 < j \leq n$ and $x_j \in C$. Choose j minimal with these properties. Then $x_i x_{i+1} \cdots x_j \in AB^*C \cap E(w)$, hence $AB^*C \cup \{w\}$ is not solid.

So far we have proved that w cannot contain a symbol from A if $AB^*C \cup \{w\}$ is to be solid. The dual proof shows that it must not contain a symbol from B either. Hence $w \in B^*$, that is, $w \in E(v)$ for some $v \in AB^*C$. But then $AB^*C \cup \{w\}$ is not solid either. □

This result has several interesting consequences. For $X = \{a, b\}$ it implies that the set $\{ab\}$ is a maximal solid code. On the other hand, for $X = \{a, b, c\}$ it shows that the set $\{ab, ac\}$, $\{ac, bc\}$, and $\{ac^i b \mid i = 0, 1, \ldots\}$ are maximal solid codes. In view of the above example M_k, this establishes the existence of a solid code, that is, $\{ab\}$ in this case, for which there is no unique maximal solid code into which it can be embedded. Moreover, this solid code can be embedded in finite and in infinite maximal solid codes. As a consequence of the preceding results and proofs one obtains the following observations.

Remark 3.17 *Let X be an alphabet.*

(1) If $X_1, \ldots, X_n$ is a partition of X then every language $X_i X_j$ with $i \neq j$ is solid.

(2) If X_1, X_2 is a partition of X then X_1X_2 is a maximal solid code.

(3) Let $Y \subseteq X$ and $L \subseteq Y^+$. Then L is a maximal solid code over Y if and only if $L \cup (X \setminus Y)$ is a maximal solid code over X.

(4) If L is a maximal solid code over X then for every $a \in X$ one has $a \in E(w)$ for some $w \in L$.

Our next results concern finite maximal solid codes. If L is a language over the alphabet X, let

$$P_X(L) = \{a \mid a \in X, aX^+ \cap L \neq \emptyset\}$$

and

$$S_X(L) = \{a \mid a \in X, X^+a \cap L \neq \emptyset\}.$$

The following auxiliary result holds true without any finiteness or maximality assumptions.

Lemma 3.18 *Let L be a solid code over the alphabet X. For every $a \in P_X(L)$ and every $b \in S_X(L)$ one has $|L \cap a^+b^+| \leq 1$.*

Proof: Consider $a \in P_X(L)$ and $b \in S_X(L)$. Assume that $w_1 = a^{r_1}b^{s_1} \in L$ and $w_2 = a^{r_2}b^{s_2} \in L$ and that these two words are distinct.

If $r_1 \leq r_2$ and $s_1 \leq s_2$ then $w_1 \in E(w_2)$ which is impossible. If $r_1 \leq r_2$ and $s_1 \geq s_2$ then $a^{r_1}b^{s_2} \in P(w_1) \cap S(w_2)$, again a contradiction! The remaining two cases are analogous. This shows that $|L \cap a^+b^+| \leq 1$. □

Theorem 3.19 *Let L be a finite maximal solid code over the alphabet X. Then the sets $P_X(L)$, $S_X(L)$, and $L \cap X$ are mutually disjoint with*

$$P_X(L) \cup S_X(L) \cup (L \cap X) = X.$$

Moreover, for every $a \in P_X(L)$ and every $b \in S_X(L)$ one has $|L \cap a^+b^+| = 1$.

Proof: As L is solid, the sets $L \cap X$, $P_X(L)$, and $S_X(L)$ are mutually disjoint. Suppose, there is a symbol $a \in X$ such that $a \notin P_X(L) \cup S_X(L) \cup (L \cup X)$. By the above remark, there is a word $w \in L$ such that $w = uav$ for some words $u, v \in X^+$. These may be chosen in such a way that $a \notin E(v)$.

Let $k = \max_{x \in L}|x|$. Then the set

$$L \cup \{ua^{k+1}v\}$$

is a solid code, contradicting the maximality of L. Therefore,

$$P_X(L) \cup S_X(L) \cup (L \cup X) = X.$$

Now consider $a \in P_X(L)$ and $b \in S_X(L)$. If $L \cap a^+b^+ = \emptyset$ then the set

$$L \cup \{a^{2k}b^k a^k b^{2k}\}$$

is a solid code, a contradiction! Therefore, using the lemma, one concludes that $|L \cap a^+b^+| = 1$. □

We have already given a few examples of finite maximal solid codes L such that $|L| = |X|$ where X is the underlying alphabet. This rather special situation is analysed in the following theorem.

Theorem 3.20 *Let X be an alphabet. The following statements hold true:*

(1) If $|X| \leq 2$ then L is a maximal solid code over X with $|L| = |X|$ if and only if $L = X$.

(2) If $|X| > 4$ and L is a maximal solid code over X with $|L| = |X|$ and $L \neq X$. Then $|X| - 3 \geq |L \cap X| \geq |X| - 4$.

Proof: Let X be an alphabet. Clearly, X is a maximal solid code. Let L be a maximal solid code over X such that $|L| = |X|$. If $|X| = 1$ then $L = X$ as this is the only possibility for a solid code over a singleton alphabet.

Suppose that $|X| = 2$ and $L = \{u, v\} \neq X$. By Lemma 3.12 one has

$$L \cap (a^+b \cup ab^+ \cup b^+a \cup ba^+) = \emptyset.$$

Theorem 3.19 implies that

$$|L \cap a^+b^+| = 1 \quad \text{and} \quad |L \cap b^+a^+| = 0$$

or vice versa. Hence, without loss of generality, we may assume that $u = a^ib^j \in L$ with $i, j > 1$. Therefore, the other word $v \in L$ has the form

$$v = a^{r_1}b^{s_1}a^{r_2}b^{s_2}\cdots a^{r_m}b^{s_m}$$

where $m \geq 2$ and $r_k, s_k \geq 1$ for $k = 1, \ldots, m$. Moreover, $v \neq abab\cdots ab$ as $P(abab\cdots ab) \cap S(abab\cdots ab) \neq \emptyset$. Choose $p > \max_k r_k$ and $q > \max_k s_k$ and let

$$w = a^{r_1}b^{s_1}a^pbab^qa^{r_2}b^{s_2}\cdots a^{r_m}b^{s_m}.$$

Then $w \notin L$ and $L \cup \{w\}$ is solid, contradicting the maximality of L.

Now let $|X| > 4$ and $L \neq X$, but $|L| = |X|$. If $a \in L \cap X$ then $a \notin E(w)$ for any $w \in L$ with $w \neq a$. Let $Y = X \setminus (L \cap X)$ and $M = L \cap Y^+$. Then M is a maximal solid code over Y by Remark 3.17 (3). The fact that $M \neq Y$ and $|M| = |Y|$ then implies that $|M| \geq 3$, that is, $|L \cap X| \leq |X| - 3$.

Assume that $|L \cap X| = |X| - n$ where $n > 4$, that is, $|Y| = n = |M|$. The sets $P_Y(M)$ and $S_Y(M)$ form a partition of Y such that $|M \cap a^+b^+| = 1$ for every $a \in P_Y(M)$ and $b \in S_Y(M)$. Therefore,

$$n = |M| \geq |P_Y(M)| \cdot |S_Y(M)| \quad \text{and} \quad |P_Y(M)| + |S_Y(M)| = n.$$

If both $P_Y(M)$ and $S_Y(M)$ contain more than one element then $n > 4$ implies

$$|P_Y(M)| \cdot |S_Y(M)| > n,$$

a contradiction! Therefore, without loss of generality, we may assume that $|P_Y(M)| = 1$ and $|S_Y(M)| = n - 1 \geq 3$. Let $P_Y(M) = \{a\}$. Thus,

$$|P_Y(M)| \cdot |S_Y(M)| = n - 1 < |M|,$$

that is, there is a (unique) word $w \in M$ which is not of the form a^ib^j with $b \in S_Y(M)$. Hence, w has the form $w = avb$ for some $v \in Y^+ \setminus a^*b^*$.

Let $c \in Y \cap E(w)$ such that $c \neq b$. Such a letter c always exists; it may or may not be equal to a. Then w has the form $w = axcyb$.

First, assume that $c \neq a$. As $|S_Y(M)| \geq 3$, there exists another letter $d \in S_Y(M)$ such that $b \neq d \neq c$. Then the set $M \cup \{axcdyb\}$ is solid, contradicting the maximality of M.

Finally, assume that $c = a$. Then there is a letter d, $d \neq a$, such that $w = axdyb$ where $yb \in a^+b^+$. As $|S_Y(M)| \geq 3$, there exists another letter $e \in S_Y(M)$ such that $e \neq d$. Then the set $M \cup \{axdeyb\}$ is solid, again a contradiction!

This proves that $n > 4$ is impossible. □

The first statement of Theorem 3.20 completely solves the case of maximal solid codes of cardinality $|X|$ over the alphabet X for the case when $|X| \leq 2$. The second statement deals with the case of $|X| > 4$: If $L \neq X$ then L will contain 3 or 4 words whose lengths are greater than 1. The case of $|X| = 3$ and $|X| = 4$ are open. We know that maximal solid codes L with $L \neq X$ and $|L| = |X|$ also exist in these cases. For instance, if $X = \{a, b, c\}$ then $L = \{ab, c^2b, acb^2\}$ is such a maximal solid code; for $X = \{a, b, c, d\}$ the set $L = \{ac, ad, bc, bd\}$ is an example.

All examples of maximal solid codes discussed so far are regular or even finite. However, as is shown next, there are non-regular maximal solid codes.

Theorem 3.21 *Let $X = \{a, b\}$ and let $T = \{2^i \mid i \in \mathbb{N}\}$ where $\mathbb{N}$ denotes the set of all natural numbers. The language*

$$L = \{aba^ib^2 \mid i \in T\} \cup \{aba^iba^jb^2 \mid i \in \mathbb{N}, j \in \mathbb{N} \setminus T\}$$

is a maximal solid code which is not regular where $\mathbb{N} = \{1, 2, 3, \ldots\}$.

Proof: One verifies that $\{u\} \cup \{v\}$ is solid for any $u, v \in L$; hence, L is solid.

Now suppose that there is a word $w \notin L$ such that $L \cup \{w\}$ is solid. Then w has the form

$$w = a^{r_1} b^{s_1} a^{r_2} b^{s_2} \cdots a^{r_n} b^{s_n}.$$

If $s_1 > 1$ then $a^{r_1} b^2 \in P(w) \cap S(v)$ for some $v \in L$. Hence, $s_1 = 1$. If $s_n = 1$ then $ab \in P(v) \cap S(w)$ for all $v \in L$. Hence, $s_n > 1$. Let i be minimal with $s_i > 1$. If $r_i \in T$ then $aba^{r_i} b^2 \in E(w) \cap L$, a contradiction! On the other hand, if $r_i \in \mathbb{N} \setminus T$ then one of the following two cases applies: If $i = 2$ then $aba^{r_i} b^2 \in P(w) \cap S(ababa^{r_i} b^2)$ where $ababa^{r_i} b^2 \in L$; if $i > 2$ then $ababa^{r_i} b^2 \in E(w) \cap L$. Both cases are impossible. This proves that L is maximal.

It follows from the Pumping Lemma for regular languages that L is not regular. □

3.4 Some Further Investigations

The class of solid codes is interesting from various points of view. Its combinatorial properties can be exploited to derive properties of disjunctive domains [Shy6] and the f-disjunctivity of certain congruences [Rei1]. For more details one can consult [Shy6] and [Rei1].

Solid codes are also interesting from an information theoretic point of view. We plan to investigate this issue further in a separate study.

CHAPTER 4

Intercodes

4.1 Intercodes

There is an infinite hierarchy of subclasses of the class of bifix codes called intercodes. In this section, we show that every intercode is uniformly synchronous, limited, and circular. In particularly, for an intercode of index n but not of index $n-1$, we have that the synchronous decoding delay $\iota(L)$ is n. A sufficient and necessary condition for an infix code to be solid is also given in this section.

Intercodes are defined below as languages $L \subseteq X^+$ satisfying

$$L^{m+1} \cap X^+ L^m X^+ = \emptyset \text{ for some } m.$$

First, however, we show that a language $L \subseteq X^+$ which satisfies this condition is a bifix code.

Theorem 4.1 *A non-empty language $L \subseteq X^+$ which satisfies the condition $L^{m+1} \cap X^+ L^m X^+ = \emptyset$ for some $m \geq 1$ is a bifix code.*

Proof: Suppose that L is a language satisfying the condition $L^{m+1} \cap X^+ L^m X^+ = \emptyset$ of some index $m \geq 1$, but not a bifix code. Then L is not a prefix code or not a suffix code. Assume L is not a prefix code. Then there exist $u, v \in L$ such that $u \neq v$ and $u = vy$ for some $y \in X^+$. This implies that $u^{m+1} = u(u^{m-1}v)y \in L^{m+1} \cap X^+ L^m X^+$, a contradiction! The case of L not being a suffix code is similar. Therefore, L is a bifix code. □

It is now clear that a language L such that $L^{m+1} \cap X^+ L^m X^+ = \emptyset$ is a code. This permits the following definition:

Definition 4.2 *A non-empty language $L \subseteq X^+$ is said to be an intercode of index m, $m \geq 1$, if $L^{m+1} \cap X^+ L^m X^+ = \emptyset$.*

A language L is called *dense* if for every $u \in X^*$, $L \cap X^*uX^* \neq \emptyset$. We now show that the converse of Theorem 4.1 does not hold. We prove the following result which is of interest in its rights.

Lemma 4.3 *No intercode is dense.*

Proof: For any $x \in L$ and $m > 1$, $X^*x^mX^* \cap L = \emptyset$. □

Corollary 4.4 *There are bifix codes which are not intercodes.*

Proof: To show this corollary, we simply give an example. Let $X = \{a, b\}$ and let $L = \{a^n bxba^n \mid x \in X^+ \text{ and } n = \#(x)\}$. Then L is a bifix code and L is dense. This implies that there exists a bifix code which is not an intercode of any index $m \geq 1$. □

For a regular language L and an integer $m \geq 1$, the problem whether or not L is an intercode of an integer index m is decidable.

Theorem 4.5 *Let L be a regular language. Then for a given $m \geq 1$, it is decidable whether or not L is an intercode of index m.*

Proof: By the definition of intercodes, L is an intercode of index m if and only if $L^{m+1} \cap X^+L^mX^+ = \emptyset$. Since the family of regular languages is closed under catenation and intersection, L^{m+1}, $X^+L^mX^+$ and $L^{m+1} \cap X^+L^mX^+$ are regular languages. Since it is decidable whether or not a regular language is empty, whether $L^{m+1} \cap X^+L^mX^+$ is empty or not is decidable. □

Given a regular language L, to decide whether L is an intercode of index m for some m is still an open problem.

Theorem 4.6 *Let $|X| \geq 2$ and let $L \subseteq X^+$ be an intercode of index n with $n \geq 1$. Then L has the following properties:*

(1) For every m, $m \geq n$, L is an intercode of index m.

(2) $L^p \cap X^+L^nX^+ = \emptyset$ for all $p \leq n+1$.

Proof: Let $L^{n+1} \cap X^+L^nX^+ = \emptyset$.

First we show that (1) holds. Suppose that there exists an integer $m \geq n$ such that

$$L^{m+1} \cap X^+L^mX^+ = \emptyset \text{ and } L^{m+2} \cap X^+L^{m+1}X^+ \neq \emptyset.$$

Then there exist $x_1, x_2, \ldots, x_{m+2}, y_1, y_2, \ldots, y_{m+1} \in L$ and $u, v \in X^+$ such that $x_1x_2\cdots x_{m+2} = uy_1y_2\cdots y_{m+1}v$. We have the following three cases:

Case 1: $|u| \geq |x_1|$. For this case, it is clear that

$x_2\cdots x_{m+2} \in X^+y_2\cdots y_{m+1}X^+$ and $L^{m+1} \cap X^+L^mX^+ \neq \emptyset$, a contradiction.

Case 2: $|v| \geq |x_{m+2}|$. Similar to the above case.

Case 3: $|u| < |x_1|$ and $|v| < |x_{m+2}|$. In this case, we have

$y_1y_2\cdots y_{m+1} \in X^+x_2x_3\cdots x_{m+1}X^+$ and $L^{m+1} \cap X^+L^mX^+ \neq \emptyset$, a contradiction.

Hence $L^{m+1} \cap X^+L^mX^+ = \emptyset$ for any $m \geq n$ implies that $L^{m+2} \cap X^+L^{m+1}X^+ = \emptyset$. By induction on m, $L^{m+1} \cap X^+L^mX^+ = \emptyset$ for all $m \geq n$.

Now we show that (2) holds. Assume on the contrary that $L^p \cap X^+L^nX^+ \neq \emptyset$ for some $p \leq n+1$. Then $\emptyset \neq L^{n+1} \cap L^{n+1-p}X^+L^nX^+ \subseteq L^{n+1} \cap X^+L^nX^+$, a contradiction! □

From Theorem 4.6, we have that every intercode of index n is an intercode of index m for all $m \geq n$. For any $m \geq 1$, we let $\mathcal{I}_m$ be the family of all intercodes of index m. Then from Theorem 4.6 we have $\mathcal{I}_m \subseteq \mathcal{I}_{m+1}$, $m \geq 1$. Now let $a \neq b \in X$ and let $u_i = a^ib^ia^i$ for $i \geq 1$. The language $L = \{u_1u_2\cdots u_{m+1}u_{m+2}, u_2, u_3, \ldots, u_m, u_{m+1}\}$ satisfies the condition $L^{m+2} \cap X^+L^{m+1}X^+ = \emptyset$, that is, $L \in \mathcal{I}_{m+1}$. Clearly,

$$u_1u_2\cdots u_{m+1}u_{m+2} \in X^+u_2u_3\cdots u_{m+1}X^+ \quad \text{and} \quad L^{m+1} \cap X^+L^mX^+ \neq \emptyset.$$

This proves:

Theorem 4.7 *The hierarchy of intercodes is strict, that is $\mathcal{I}_m \subsetneq \mathcal{I}_{m+1}$ for all m, $m \geq 1$.*

Theorem 4.8 *Let $L \subseteq X^+$. If L is an intercode, then $L \subseteq Q$.*

Proof: Let L be an intercode of index m for some m. Assume that $L \not\subseteq Q$, that is, there is a word $u = f^i \in L$ for some $i > 1$ and $f \in X^+$. Then $u^{m+1} = f^{i(m+1)} = f(f^{im})f^{i-1} \in L^{m+1} \cap X^+L^mX^+$, that is $L^{m+1} \cap X^+L^mX^+ \neq \emptyset$. This implies that $L \notin \mathcal{I}_m$, a contradiction. Thus $L \subseteq Q$. □

Of course, $\{ab, ba\} \in \mathcal{Q}$ but $\{ab, ba\} \notin \mathcal{I}_m$ for any $m \geq 1$. Moreover, $\{ab, ba\} \in \mathcal{C}_b$. This implies that $\mathcal{I}_m \subsetneq \mathcal{Q} \cap \mathcal{C}_b$. From Theorems 4.1, 4.6 and 4.8, we have:

Corollary 4.9 $\mathcal{I}_1 \subsetneq \mathcal{I}_2 \subsetneq \ldots \subsetneq \mathcal{I}_m \subsetneq \ldots \subsetneq \mathcal{Q} \cap \mathcal{C}_b$.

For a word $x \in X^+$, if $x = a_1a_2\cdots a_n$ for $a_i \in X$, $i = 1, 2, \ldots, n$, then the word $a_n \cdots a_2a_1$ is called *the reverse of x* and denoted by $\mathrm{rev}(x)$. For a language $L \subseteq X^+$, we define the reverse $\mathrm{rev}(L)$ of L as $\mathrm{rev}(L) = \{\mathrm{rev}(x) \mid x \in L\}$. The following characterization of intercodes of index m is an immediate result from the definition of intercodes.

Theorem 4.10 *Let $L \subseteq X^+$. Then the following statements are equivalent:*

(1) L is an intercode of index m;

(2) $\mathrm{rev}(L)$ is an intercode of index m;

(3) For any $u \in L^m$, $x, y \in X^*$, $xuy \in L^{m+1}$ implies that $x = 1$ or $y = 1$.

Now we give the definition of synchronously decipherable codes as follows:

Definition 4.11 *A code L is synchronously decipherable if there is a non-negative integer n such that*

$$\forall u, v \in X^* \wedge x \in L^n \wedge uxv \in L^* \Rightarrow u, v \in L^*.$$

If a code L is synchronously decipherable, the smallest n such that this condition holds is the synchronous decoding delay $\iota(L)$ of L.

Theorem 4.12 *Let $L \subseteq X^+$ be a code. Then L is an intercode of index n if and only if L is synchronously decipherable with delay $\iota(L) \leq n$.*

Proof: Let $L^{n+1} \cap X^+L^nX^+ = \emptyset$. We first show that L is synchronously decipherable. Let $x \in L^n$, let $u, v \in X^*$ and let $uxv \in L^*$. Since $x \in L^n$ and $n \geq 1$, $uxv \in L^+$. If $u = 1$ or $v = 1$, then $v \in L^*$ or $u \in L^*$ as L is a bifix code. Hence $u, v \in L^*$.

Hence assume that there exist $u, v \in X^+$ such that $uxv \in L^m$ with m minimal. We then have $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m \in L$ such that

$$y_1y_2\cdots y_m = uxv \quad \text{and} \quad x_1x_2\cdots x_n = x.$$

If $m \leq n+1$, then $L^m \cap X^+L^nX^+ \neq \emptyset$, contradicting Theorem 4.6. Hence $m > n+1$. Since m is minimal, $y_2\cdots y_m \notin X^+xv$ and $y_1y_2\cdots y_{m-1} \notin uxX^+$. Thus, $x_1x_2\cdots x_n \in X^*y_2\cdots y_{m-1}X^*$. That is, $x_1x_2\cdots x_n = u_1y_2\cdots y_{m-1}v_1$ for some $u_1, v_1 \in X^*$.

The case $u_1, v_1 \in X^+$ implies that $L^n \cap X^+L^{m-2}X^+ \neq \emptyset$. Since $m-2 \geq n$, by Theorem 4.6, $L^{m-2+1} \cap X^+L^{m-2}X^+ \neq \emptyset$ and $L^{n+1} \cap X^+L^nX^+ \neq \emptyset$, a contradiction. Thus $v_1 = 1$ or $u_1 = 1$. Assume $v_1 = 1$. Since L is a bifix code, the equality $x_1x_2\cdots x_n = u_1y_2\cdots y_{m-1}$ can hold true only when $m = n+2$ and $u_1 = 1$. That is, $u = y_1$, $v = y_m$, and $u, v \in L$. Similarly $u_1 = 1$ implies that $u, v \in L$.

For the converse, let L be a synchronously decipherable code with $m = \iota(L)$ and suppose that $L^{m+1} \cap X^+L^mX^+ \neq \emptyset$. Then there exist $x_1, x_2, \ldots, x_m$, $y_1, y_2, \ldots, y_{m+1} \in L$ and $u, v \in X^+$ such that

$$y_1y_2\cdots y_{m+1} = ux_1x_2\cdots x_mv.$$

Since $m = \iota(L)$ and $u, v \in X^+$, $u, v \in L^+$. There exist $i, j \geq 1$ and $u_1, u_2, \ldots, u_i$, $v_1, v_2, \ldots, v_j \in L$ such that $u = u_1u_2\cdots u_i$ and $v = v_1v_2\cdots v_j$. Thus

$$y_1y_2\cdots y_{m+1} = u_1u_2\cdots u_ix_1x_2\cdots x_mv_1v_2\cdots v_j.$$

Since L is a bifix code, $m+1 = i+m+j \geq m+2$, a contradiction.

Thus, $L^{m+1} \cap X^+L^mX^+ = \emptyset$. By Theorem 4.6 (1), $L^{n+1} \cap X^+L^nX^+ = \emptyset$ for all $n \geq m$. □

A code L is called ***(p,q)-limited*** if for all $u_0, u_1, u_2, \ldots, u_{p+q} \in X^*$,

$$u_{i-1}u_i \in L^* \text{ for all } 1 \leq i \leq p+q \text{ implies that } u_p \in L^*.$$

The following theorem shows that an intercode of index m is (p,q)-limited for all p,q with $p+q = 2m+1$. Since every intercode of index m is an intercode of index n where $n \geq m$. Thus every intercode of index m is (p,q)-limited for all p,q with $p+q = 2n+1$ for all $n \geq m$.

Theorem 4.13 *Let $L \subseteq X^+$ be an intercode of index m. Then L is (p,q)-limited for all p,q with $p+q = 2m+1$.*

Proof: Let $u_0, u_1, u_2, \ldots, u_{2m+1} \in X^*$ such that $u_{i-1}u_i \in L^*$ for all $1 \leq i \leq 2m+1$. Suppose that there exixts k with $0 \leq k \leq 2m+1$ such that $u_k = 1$. Then $u_k \in L^*$. Moreover, $u_{k+1}, u_{k+1}u_{k+2} \in L^*$. Since L is a bifix code, $u_{k+2} \in L^*$. It follows that $u_{k+3}, \ldots, u_{2m+1} \in L^*$, and in a similar way $u_{k-1}, u_{k-2}, \ldots, u_0 \in L^*$.

If there exists no k with $0 \leq k \leq 2m+1$ such that $u_k = 1$, then $u_{2i-1}u_{2i} \in L^+$ for all $i = 1, 2, \ldots, m$. Hence $u_1u_2 \ldots u_{2m} \in L^j$ for some $j \geq m$. By Theorem 4.6 (1), one has $L^{j+1} \cap X^+L^jX^+ = \emptyset$. By Theorem 4.12, $u_0u_1u_2 \cdots u_{2m}u_{2m+1} \in u_0L^ju_{2m+1}$ implies that $u_0, u_{2m+1} \in L^*$. Since L is a bifix code, $u_1, u_2, \ldots, u_{2m} \in L^*$. Hence L is (p,q)-limited for all p,q with $p+q = 2m+1$. □

A code L is called a ***circular code*** if for all $n, m \geq 1$, $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m \in L$, and $u \in X^+$, $v \in X^*$, the following two equalities $x_1 = vu$ and $ux_2 \cdots x_nv = y_1y_2 \cdots y_m$ imply that $v = 1$, $n = m$ and $x_i = y_i$ for all $1 \leq i \leq n$. From ([Ber2]), we have the following theorem:

Theorem 4.14 ([Ber2]) ***Every limited code is circular.***

As a consequence, we have:

Corollary 4.15 *Every intercode is circular.*

We now turn to considering non-empty infix codes which are (p,q)-limited. This results in special classes of intercodes.

Theorem 4.16 *If an infix code L is (p,q)-limited for some p,q with $p+q=2m+1$ for some integer $m \geq 1$, then the following two statements hold true:*

(1) L is (p,q)-limited for all p,q with $p+q=2m+1$;

(2) L is an intercode of index m.

Proof: Let L be an infix code which is (p,q)-limited for some p,q with $p+q=2m+1$ for some integer $m \geq 1$.

First we show that statement (1) holds. Let $u_0, u_1, \ldots, u_{2m+1} \in X^*$ such that $u_{i-1}u_i \in L^*$ for all $1 \leq i \leq 2m+1$. Since L is (p,q)-limited, $u_p \in L^*$. Suppose that $u_n \in L^*$, $n \geq p$. Then $u_n u_{n+1} \in L^*$ and the fact that L is infix imply that $u_{n+1} \in L^*$. Hence by induction on n, $u_n \in L^*$ for all $n \geq p$. Similarly $u_n \in L^*$ for all $n \leq p$. Thus statement (1) holds.

Now we show that L is an intercode of index m. Suppose on the contrary that $L^{m+1} \cap X^+L^mX^+ \neq \emptyset$. Then there exist $x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_{m+1} \in L$ and $u, v \in X^+$ such that

$$y_1 y_2 \ldots y_{m+1} = u x_1 x_2 \ldots x_m v.$$

Since L is infix, there exist $u_1, u_2, \ldots, u_{2m} \in X^+$ such that

$$x_i = u_{2i-1}u_{2i} \text{ for all } 1 \leq i \leq m,$$

$$y_1 = uu_1,\ y_{m+1} = u_{2m}v \text{ and } y_j = u_{2j-2}u_{2j-1} \text{ for all } 2 \leq j \leq m.$$

Let $u_0 = u$ and let $u_{2m+1} = v$. Then by statement (1), $u_i \in L^*$ for all $0 \leq i \leq 2m+1$ and actually $u_i \in L^+$ because of $u_i \in X^+$. This contradicts the fact that L is an infix code. Thus $L^{m+1} \cap X^+L^mX^+ = \emptyset$. □

Thus, when we add the infix condition we obtain another strict hierarchy:

Corollary 4.17 *$\mathcal{I}_{i_1} \subsetneq \mathcal{I}_{i_2} \subsetneq \ldots \subsetneq \mathcal{I}_{i_m} \subsetneq \ldots \subsetneq Q \cap C_b$ where $\mathcal{I}_{i_m}$ is the family of infix intercodes of index m.*

Proof: We need only to show the strictness of the inclusions. It can be done by giving the following example: Let $a \neq b \in X$ and let $u_i = a^i b^i a^i$ for $i \geq 1$ and let $L = \{u_i u_{i+1} \mid 1 \leq i \leq 2m+2\} \cup \{u_{2m+3}\}$. Then $L \in \mathcal{I}_{i_{m+1}} \cap Q \cap C_b$ but $L \notin \mathcal{I}_{i_m}$. □

4.2. n-Intercodes

In this section, we investigate the properties of n-interocdes. First we give the definition of n-intercodes as follows:

Definition 4.18 *A language $L \subseteq X^+$ is called an n-intercode of index m if every subset of L with at most n elements is an intercode of index m.*

It is clear that every n-intercode of index m is a k-intercode of index m for every $k \leq n$. Thus the class of all n-intercodes is a subclass of all k-intercodes for $k \leq n$. From the definition of intercodes, it is clear that every word in a 1-intercode is primitive. Moreover, we have the following property of 1-intercodes.

Theorem 4.19 *A language L is a 1-intercode if and only if $L \in Q$.*

Proof: Let L be a 1-intercode. For every $u \in L$, $\{u\}$ is an intercode. By Theorem 4.8, $u \in Q$. Thus, $L \in Q$.

For the converse, let L be a non-empty subset of Q. If a word $u \in X^+$ has the property that $u^2 \in X^+uX^+$, then $u \notin Q$ (see [Lal1] and [Lyn1], for example). Hence, for every $u \in L$, $u^2 \notin X^+uX^+$. Theorem 4.6 implies that L is also a 1-intercode of any index $m \geq 1$. □

It is also clear that every intercode of index m is an n-intercode of index m for all $n \geq 1$. But for $n \leq 2m$, an n-intercode of index m is not necessarily an intercode of index m. For example: Let $a \neq b \in X$ and let $u_i = a^i b^i a^i$ for $i \geq 1$. Then clearly the language $L = \{u_i u_{i+1} \mid i = 1, 2, \ldots, 2m+1\}$ is a $2m$-intercode of index m, but not an intercode of index m. The property that a language is an intercode of index m if and only if it is a $(2m+1)$-interocde of index m, shown in the following theorem, is a very important result in this thesis. It is one of the basic examples that help us to generalize the mechanism investigated in Section 2.1 to finitary relations. More details about the hierarchy of intercodes and n-intercodes will be investigated in Section 5.4 by using the construction of finitary relations. We show this property of intercodes and n-intercodes as follows:

Theorem 4.20 *Let $L \subseteq X^+$. Then L is an intercode of index m if and only if L is a (2m+1)-intercode of index m.*

Proof: The sufficient condition is derived directly from the definitions.
For the converse, assume that $L^{m+1} \cap X^+ L^m X^+ \neq \emptyset$, that is, there exist $x_1, x_2, \ldots, x_m, x_{m+1} \in L$, $y_1, y_2, \ldots, y_m \in L$ and $u, v \in X^+$ such that

$$x_1 x_2 \cdots x_m x_{m+1} = u y_1 y_2 \cdots y_m v.$$

This shows that L is not a $(2m+1)$-intercode of index m, a contradiction! □

A diagram to illustrate the relationships between classes of intercodes and n-intercodes of different indices is shown in Figure 3 in Section 1.1.

4.3 Comma-Free Codes

In this section, we study a subclass of the intercodes called comma-free codes. Comma-free codes have been investigated by many researchers. For more details, see [Ber2], [Cri1], [Hsi1], [Gol2], [Jig1] and [Sch3].

A comma-free code L has the property that for a message x in L^+, if one decodes this message and finds a factor y of x in L, then y is a term of the unique L-factorization of x. Thus a comma-free code has a very good performance in message decoding. Moreover, the synchronous decoding delay of a comma-free code is just one. A word in a message consisting of words taken from a comma-free code can be identified when the last symbol of this word is received.

Now we give the definition of synchronization delay (see [Gol1]) as follows:

Definition 4.21 *A code L is uniformly synchronous if there is a non-negative integer n such that*

$$\forall x \in L^n \wedge u, v \in X^* \wedge uxv \in L^* \Rightarrow ux, xv \in L^*.$$

If a code L is uniformly synchronous, the smallest n such that this property holds is the uniform synchronization delay of L, denoted by $\sigma(L)$.

Definition 4.22 [Gol1] *A code is comma-free if L is bifix and $\sigma(L) = 1$.*

From [Ber1] we have the following characterization of comma-free codes which is related to the limited and infix properties.

Theorem 4.23 ([Ber1]) *A code L is comma-free if and only if L satisfies the following two statements:*

(1) L is (p, q)-limited for all p, q with $p + q = 3$;

(2) $X^+LX^+ \cap L = \emptyset$.

A simpler characterization of comma-free codes is given as follows by using the set-intersection form:

Theorem 4.24 *A language $L \subseteq X^+$ is a comma-free code if and only if L is an intercode of index one.*

Proof: Let L be a comma-free code, that is, L is a bifix code and $\sigma(L) = 1$. For $x \in L$ and $u, v \in X^*$, $uxv \in L^+$ implies that $ux, xv \in L^+$. Since L is a bifix code, $u, v \in L^*$. Theorem 4.12 implies that L is an intercode of index one. This proves the sufficient condition.

For the converse, let L be an intercode of index one. Theorem 4.1 implies that L is a bifix code. By Theorem 4.12, for every $x \in L$ and $u, v \in X^*$, $uxv \in L^+$ implies that $u, v \in L^*$. Clearly, $ux, xv \in L^+$. This complete this proof. □

Using the properties of intercodes, we show the following characterization of comma-free codes:

Theorem 4.25 *Let $L \subseteq X^+$ be a code. Then L is infix and (p,q)-limited for some p, q with $p + q = 3$ if and only if L is comma-free.*

Proof: By Theorem 4.16, the sufficient condition holds true. We need only to show the necessary condition. By Theorem 4.1, L is bifix. By Theorem 4.23, $L \cap X^+LX^+ = \emptyset$. Hence L is infix. From Theorem 4.13, we have that L is (p,q)-limited for all p, q with $p + q = 3$. Thus the necessary part holds true. □

In [Hsi1], it is shown that a comma-free code is an infix code. But this is not the case for intercodes of index $m \geq 2$. Consider the following example: Let $X = \{a, b\}$ and let $L = \{b^2abab^2, a\}$. The language L is an intercode of index 2 but not an infix code. It follows that the class of comma-free codes is a proper subclass of the class of intercodes.

4.4 n-Comma-Free Codes

A language L is called an ***n-comma-free code*** if every subset of L with at most n elements is a comma-free code. We now consider the n-comma-free codes. We have:

Theorem 4.26 *A language is an n-comma-free code for some $n \geq 3$ if and only if it is a comma-free code.*

Proof: From Theorem 4.24, we need only to show that a language is an n-intercode of index 1 for some $n \geq 3$ if and only if it is an intercode of index 1. From Theorem 4.20, we derive the result that a language is an n-intercode of index 1 for some $n \geq 3$ if and only if it is an intercode of index 1. □

Moreover, Theorem 4.19 implies that a language is a 1-intercode if and only if it is a non-empty subset of Q. Hence, in the sequel, we have only to consider 2-comma-free codes.

A language L is ***nonoverlapping*** if $u, v \in L$, $u = xy$, $v = yz$ for some $x, y, z \in X^*$ imply that $y = 1$. The following results are obtained in [Hsi1].

Theorem 4.27 ([Hsi1]) *The following two statements hold:*

(1) Every 2-comma-free code is an infix code.

(2) Let $L \subseteq Q$ be a nonoverlapping language. If L is an infix code then L is a 2-comma-free code.

For $L \subseteq X^+$, if every word in L is unbordered and if L is without the condition of being nonoverlapping, then the conclusion of Theorem 4.27 need not be true. For example: Let $X = \{a, b, c\}$ and let $L = \{cab, bc\}$. Then $L \subseteq Q$ and every word in L is unbordered. But L is not 2-comma-free. We give a characterization of 2-comma-free codes as follows:

Theroem 4.28 *Let $L \subseteq X^+$ and let every word in L be unbordered. Then the following two statements are equivalent:*

(1) L is a 2-comma-free code;

(2) L is an infix code and, for $u, v \in X^+$, $uv \in L$ implies that $L \cap vX^*u = \emptyset$.

Proof: Let L be a 2-comma-free code. By Theorem 4.27 (1), L is an infix code. If $L' = \{uv, vxu\} \subseteq L$ for some $x \in X^*$, then $(vxu)^2 \in L'^2 \cap X^+L'X^+$. This implies that L is not a 2-comma-free code, a contradiction!

For the converse, let L satisfy condition (2). The condition $L \cap vX^*u = \emptyset$ implies that $L \subseteq Q$ and L is a 2-code. Assume that L is not a 2-comma-free code. Then there exist $r, s \in X^+$ and $x, y \in L$ such that $rxs = y^2$, $rxs = x^2$, $rxs = xy$, or $rxs = yx$. Since L is an infix code and every word in L is unbordered, cases $rxs = x^2$, $rxs = xy$ and $rxs = yx$ can not be true. We need only to consider the case $rxs = y^2$. Since L is an infix code, we have that $x = uv$ for some $u, v \in X^+$ such that $y = y'u = vy''$, $y', y'' \in X^+$. If $|y'| \geq |v|$, then $L \cap vX^*u \neq \emptyset$, a contradiction. If $|y'| < |v|$, then $u = z_1z_2$, $v = z_3z_1$ for some $z_1 \in X^+$. This implies that $x = z_1z_2z_3z_1$ is not unbordered, a contradiction! □

We now discuss closure properties of the class of 2-comma-free codes.

Theorem 4.29 *The class of 2-comma-free codes is closed under reversal and intersection with arbitrary sets, and is not closed under union, product, complement, catenation closure, and non-earsing homomorphisms.*

Proof: By the definition of intercodes of index 1, the class of 2-comma-free codes is closed under reversal and intersection with arbitrary sets. The proof of the non-closure properties will be done by constructing some examples. The languages $\{ab\}$ and $\{ba\}$ are two 2-comma-free codes. It is clear that the union $\{ab, ba\} = \{ab\} \cup \{ba\}$, the product $\{(ab)^2\} = \{ab\}\{ab\}$, and the catanation closure $\{ab\}^+$ are not 2-comma-free codes. This proves the statements for union, product, and catenation closure. Moreover, as the class of 2-comma-free codes is closed under intersections, but not closed under unions, it is also not closed under complement.

Consider alphabets $X = \{a, b\}$, $Y = \{a\}$ and let h be the homomorphism of X^* into Y^* which is defined by $h(a) = h(b) = a$. h is non-eracing and maps the

2-comma-free code $\{ab\}$ onto $\{a^2\}$, the latter is not a 2-comma-free code. Thus, the class of 2-comma-free codes is not closed under non-erasing homomorphisms. □

Theorem 4.30 *The class of 2-comma-free codes is closed under inverse non-erasing homomorphisms.*

Proof: Let X and Y be alphabets and let $h : X^* \rightarrow Y^*$ be a non-erasing homomorphism. It suffices to show that for any $u, v \in Y^+$, if $\{u, v\}$ is a comma-free code, then $\{\overline{u}, \overline{v}\}$ is a comma-free code for every $\overline{u} \in h^{-1}(u)$ and $\overline{v} \in h^{-1}(v)$.

Suppose that for some $\overline{u}$ and $\overline{v}$ this is not true. If $\overline{u} \neq \overline{v}$ and $\{\overline{u}, \overline{v}\}^2 \cap X^+\overline{v}X^+ \neq \emptyset$ then $\{u, v\}^2 \cap Y^+vY^+ \neq \emptyset$, a contradiction! The case of $\overline{u} \neq \overline{v}$ and $\{\overline{u}, \overline{v}\}^2 \cap X^+\overline{u}X^+ \neq \emptyset$ is also impossible. □

It has been shown in Section 3.2 that the class of solid codes has the same closure and non-closure properties as the class of 2-comma-free codes does. Moreover the class of all codes also has the same closure and non-closure properties.

CHAPTER 5

n-ary Relations on Free Monoids and Their Independent Sets

5.1 n-ary Relations and Their Independent Sets

As shown in the preceding chapters, even though there are many classes of languages which can be expressed as the independent sets with respect to some binary relations, there are still many other classes of languages which cannot be characterized in such a way. As examples, we have the classes of codes, intercodes, and n-codes.

In this chapter, we generalize the definitions and some results of Section 2.1 to n-ary relations. It turns out that some of the classes of codes and n-codes mentioned before can be characterized as the independent sets with respect to n-ary relations. The generalization is twofold: We consider relations of arbitrary finite arity and their independent sets. Moreover, we develop the theory over arbitrary sets first before considering the case of relations on free monoids. This latter point results from the observation that most of the general results of Section 2.1 are actually true for binary relations on arbitrary sets.

In the sequel, let S be an arbitrary non-empty set. For an n-tuple $x = (x_1, x_2, \ldots, x_n) \in [S]^n$, let $\operatorname{cont} x$ denote the set of components of x, that is,

$$\operatorname{cont} x = \{y \mid y \in S, \exists i : x_i = y\}.$$

$\operatorname{cont} x$ is called the ***contents*** of x.

Definition 5.1 Let ω be an n-ary relation on S. A set $L \subseteq S$ is said to be *ω-independent* if $x \in \omega$ implies $\operatorname{cont} x \not\subseteq L$.

Note that for $n = 2$ this new definition of ω-independence does not specialize to the one used in the preceding sections and in the literature. As shown in Section 2.1 only symmetric and reflexive binary relations need be considered given the older definition. This is no longer true with this new definition. However, we believe that the new definition is more natural than the one used before. Details of the connection between these notions are discussed in Section 5.2 below. Whenever it is necessary to refer to the notions of Section 2.1 along with their re-defined versions, we attach the subscript "$_{\text{old}}$" to the former, that is, for instance ω-independence$_{\text{old}}$ would mean ω-independence in the sense of Section 2.1.

For an n-ary relation ω on S, $\mathcal{L}_\omega$ denotes the family of all non-empty, ω-independent subsets of S. Given our new definition of ω-independence, it is no longer true that $\mathcal{L}_\omega$ contains all the singleton sets. Thus we drop this condition from the definition of strictly non-trivial families of sets. Hence, a family $\mathcal{L}$ of subsets of S is said to be ***non-trivial*** if $\mathcal{L} \subseteq 2^S$ and $\emptyset \notin \mathcal{L}$.

Remark 5.2 *Let ω be an n-ary relation on S. Then $\mathcal{L}_\omega$ is a non-trivial family of languages which is closed under taking non-empty subsets. If $(x, x, \ldots, x) \in \omega$ for all $x \in S$ then $\mathcal{L}_\omega = \emptyset$.*

We first investigate the connection between the lattice structure of the set $\mathfrak{R}^{(n)}$ of all n-ary relations on S and the set

$$\mathfrak{L}^{(n)} = \{\mathcal{L} \mid \exists\omega \subseteq [S]^n : \mathcal{L} = \mathcal{L}_\omega\}$$

of non-trivial families of subsets of S which can be characterized by n-ary relations.

Lemma 5.3 *Let $\omega_1, \omega_2 \in \mathfrak{R}^{(n)}$. Then $\omega_1 \subseteq \omega_2$ implies that $\mathcal{L}_{\omega_2} \subseteq \mathcal{L}_{\omega_1}$.*

Proof: Consider $L \in \mathcal{L}_{\omega_2}$ and $a_1, a_2, \ldots, a_n \in L$. Then $(a_1, a_2, \ldots, a_n) \notin \omega_2$. Thus $(a_1, a_2, \ldots, a_n) \notin \omega_1$. This implies $L \in \mathcal{L}_{\omega_1}$. $\square$

The converse of the Lemma 5.3 is not true in general as is shown in the following example.

Example 5.4 For $X = \{a, b, c\}$, $n = 2$, and $S = X^+$, consider

$$\omega_1 = [X^+]^2 \setminus \{(a,a),(b,b),(a,b),(b,a),(a,c)\}$$

and

$$\omega_2 = [X^+]^2 \setminus \{(a,a),(b,b),(a,b),(b,a),(b,c)\}.$$

Then

$$\mathcal{L}_{\omega_1} = \mathcal{L}_{\omega_2} = \{\{a,b\},\{a\},\{b\}\}.$$

However, ω_1 and ω_2 are not comparable.

For any n-ary relation ω on S let

$$\Phi(\omega) = \{\varphi \mid \varphi \in \mathfrak{R}^{(n)}, \mathcal{L}_\varphi = \mathcal{L}_\omega\},$$

and let

$$\overline{\omega} = \bigcup_{\varphi \in \Phi(\omega)} \varphi \quad \text{and} \quad \underline{\omega} = \bigcap_{\varphi \in \Phi(\omega)} \varphi.$$

From $\omega \in \Phi(\omega)$ it is clear that $\underline{\omega} \subseteq \omega \subseteq \overline{\omega}$, that is, $\mathcal{L}_{\overline{\omega}} \subseteq \mathcal{L}_\omega \subseteq \mathcal{L}_{\underline{\omega}}$.

Theorem 5.5 *For $\omega \in \mathfrak{R}^{(n)}$ the set $\Phi(\omega)$ is a complete $\cup$-semilattice with $\overline{\omega}$ as its maximum.*

Proof: Let Φ' be a non-empty subset of $\Phi(\omega)$ and let

$$\varphi' = \bigcup_{\varphi \in \Phi'} \varphi.$$

We have to show that $\mathcal{L}_{\varphi'} = \mathcal{L}_\omega$.

Let L be a non-empty subset of S. If $L \notin \mathcal{L}_\omega$ then $L \notin \mathcal{L}_\varphi$ for every $\varphi \in \Phi' \subseteq \Phi(\omega)$, that is, for every $\varphi \in \Phi'$, there exists $x \in \varphi$ such that $\operatorname{cont} x \subseteq L$.

This implies $x \in \varphi'$ for every such x and, therefore, $L \notin \mathcal{L}_{\varphi'}$. This proves that $\mathcal{L}_{\varphi'} \subseteq \mathcal{L}_{\omega}$.

For the converse, assume that there exists a set $L \in \mathcal{L}_{\omega}$ such that $L \notin \mathcal{L}_{\varphi'}$. Then there exists $x \in \varphi'$ with $\operatorname{cont} x \subseteq L$. This implies that there exists $\varphi \in \Phi'$ such that $x \in \varphi$, hence $L \notin \mathcal{L}_{\varphi} = \mathcal{L}_{\omega}$, a contradiction! □

Corollary 5.6 *For $\omega \in \mathfrak{R}^{(n)}$ one has $\mathcal{L}_{\omega} = \mathcal{L}_{\overline{\omega}}$.*

Our next result is provides a characterization of $\overline{\omega}$. For this we need the following notions of symmetric and upward symmetric n-ary relation. For $n = 2$, the former specializes to the usual notion.

Definition 5.7 Let ω be an n-ary relation on S.

(1) ω is said to be ***symmetric*** if it has the following property: For $x, y \in [S]^n$, if $x \in \omega$ and $\operatorname{cont} x = \operatorname{cont} y$ then $y \in \omega$.

(2) ω is said to be ***upward symmetric*** if it has the following property: For $x, y \in [S]^n$, if $x \in \omega$ and $\operatorname{cont} x \subseteq \operatorname{cont} y$ then $y \in \omega$.

Clearly, every upward symmetric relation is also symmetric. Let $\mathfrak{S}^{(n)}$ and $\mathfrak{US}^{(n)}$ denote the sets of symmetric and upward symmetric n-ary relations on S, respectively. For an n-ary relation ω, let $\operatorname{symm} \omega$ and $\operatorname{upsymm} \omega$ denote the symmetric and the upward symmetric closures of ω, respectively.

Theorem 5.8 *Let φ and ω be n-ary relations on S. Then $\varphi = \overline{\omega}$ if and only if φ satisfies the following two conditions:*

(1) $\varphi \in \Phi(\omega)$.

(2) For every $x, y \in [S]^n$, if $x \in \omega$ and $\operatorname{cont} x \subseteq \operatorname{cont} y$ then $y \in \varphi$.

Proof: Assume that $\varphi = \overline{\omega}$. Then condition (1) is satisfied by Corollary 5.6. Suppose that there are $x, y \in [S]^n$ such that $x \in \omega$, $\operatorname{cont} x \subseteq \operatorname{cont} y$ and $y \notin \varphi = \overline{\omega}$. Then $\operatorname{cont} y \notin \mathcal{L}_{\omega} = \mathcal{L}_{\varphi}$. Consider $\varphi' = \varphi \cup \{y\}$. Then $\mathcal{L}_{\varphi'} = \mathcal{L}_{\varphi} = \mathcal{L}_{\omega}$. Thus $\varphi' \in \Phi(\omega)$ and $\varphi \subsetneq \varphi' \subseteq \overline{\omega}$, a contradiction!

For the converse, assume that φ satisfies the two conditions. Condition (1) implies that $\varphi \subseteq \overline{\omega}$. Consider $y \in \overline{\omega}$. It follows that $\operatorname{cont} y \notin \mathcal{L}_{\overline{\omega}} = \mathcal{L}_\omega$. By the definition of $\mathcal{L}_\omega$, there exists $x \in [S]^n$ with $\operatorname{cont} x \subseteq \operatorname{cont} y$ and $y \in \omega$. By condition (2), $y \in \varphi$, hence also $\overline{\omega} \subseteq \varphi$. □

Corollary 5.9 *For any n-ary relation ω one has* $\overline{\omega} = \operatorname{upsymm} \omega$.

The fact that "upsymm" is a closure operator together with Corollary 5.9 implies that the mapping $\omega \mapsto \overline{\omega}$ is monotonic.

Corollary 5.10 *Let ω be an n-ary relation on S. One has*

$$x \notin \overline{\omega} \iff \operatorname{cont} x \in \mathcal{L}_{\overline{\omega}}$$

for all $x \in [S]^n$.

Theorem 5.11 (Inclusion Theorem) *Let ω_1, ω_2 be n-ary relations on S. One has*

$$\overline{\omega}_1 \subseteq \overline{\omega}_2 \iff \mathcal{L}_{\omega_2} \subseteq \mathcal{L}_{\omega_1}.$$

Proof: If $\overline{\omega}_1 \subseteq \overline{\omega}_2$ then $\mathcal{L}_{\omega_2} = \mathcal{L}_{\overline{\omega}_2} \subseteq \mathcal{L}_{\overline{\omega}_1} = \mathcal{L}_{\omega_1}$.

For the converse, assume that $\mathcal{L}_{\omega_2} \subseteq \mathcal{L}_{\omega_1}$ and $\overline{\omega}_1 \not\subseteq \overline{\omega}_2$. Hence there is an $x \in [S]^n$ with $x \in \overline{\omega}_1$ and $x \notin \overline{\omega}_2$. By Corollary 5.10 one has $\operatorname{cont} x \in \mathcal{L}_{\overline{\omega}_2} = \mathcal{L}_{\omega_2}$ and $\operatorname{cont} x \notin \mathcal{L}_{\overline{\omega}_1} = \mathcal{L}_{\omega_1}$, a contradiction! □

Corollary 5.12 *For every $n \in \mathbb{N}$ the mapping $\omega \mapsto \mathcal{L}_\omega$ is an antitonic bijection of $\mathfrak{US}^{(n)}$ onto $\mathfrak{L}^{(n)}$.*

The mapping $\omega \mapsto \mathcal{L}_\omega$ of $\mathfrak{US}^{(n)}$ onto $\mathfrak{L}^{(n)}$ can be shown to be a lattice isomorphism with respect to naturally defined lattice structures on $\mathfrak{US}^{(n)}$ and $\mathfrak{L}^{(n)}$.

Lemma 5.13 $(\mathfrak{US}^{(n)}, \cup, \cap)$ *is a complete lattice. Moreover, for* $\omega_1, \omega_2 \in \mathfrak{US}^{(n)}$ *one has*

$$\mathcal{L}_{\omega_1} \cap \mathcal{L}_{\omega_2} = \mathcal{L}_{\omega_1 \cup \omega_2} \quad \textit{and} \quad \mathcal{L}_{\omega_1} \cup \mathcal{L}_{\omega_2} \subseteq \mathcal{L}_{\omega_1 \cap \omega_2}.$$

Proof: Consider a non-empty family $\Psi = \{\omega_i \mid i \in I\}$ of n-ary relations in $\mathfrak{US}^{(n)}$ where I is an arbitrary index set. Clearly, with $\psi_\cup$ and $\psi_\cap$ the union and the intersection over Ψ one has

$$\psi_\cup \subseteq \overline{\psi}_\cup \quad \text{and} \quad \psi_\cap \subseteq \overline{\psi}_\cap.$$

If $x \in \overline{\psi}_\cup$ then $y \in \psi_\cup$ for some $y \in [S]^n$ with $\operatorname{cont} y \subseteq \operatorname{cont} x$. Thus $y \in \omega_i$ for some $i \in I$ and, therefore, $x \in \omega_i$. This implies $x \in \psi_\cup$.

If $x \in \overline{\psi}_\cap$ then $y \in \psi_\cap$ for some $y \in [S]^n$ with $\operatorname{cont} y \subseteq \operatorname{cont} x$. Thus $y \in \omega_i$ for all $i \in I$ and, therefore, $x \in \omega_i$ for all i. Hence $x \in \psi_\cap$.

This proves that $(\mathfrak{US}^{(n)}, \cup, \cap)$ is a complete lattice. Now consider $\mathcal{L}_{\omega_1} \cap \mathcal{L}_{\omega_2}$. Using the fact that the mapping $\omega \mapsto \mathcal{L}_\omega$ is antitonic, one has $\mathcal{L}_{\omega_1 \cup \omega_2} \subseteq \mathcal{L}_{\omega_1} \cap \mathcal{L}_{\omega_2}$. Let $L \in \mathcal{L}_{\omega_1} \cap \mathcal{L}_{\omega_2}$ and $x \in [S]^n$ with $\operatorname{cont} x \subseteq L$. Then $x \notin \omega_1$ and $x \notin \omega_2$, that is $x \notin \omega_1 \cup \omega_2$. Therefore, $L \in \mathcal{L}_{\omega_1 \cup \omega_2}$. This proves that $\mathcal{L}_{\omega_1} \cap \mathcal{L}_{\omega_2} = \mathcal{L}_{\omega_1 \cup \omega_2}$.

The last statement is again an immediate consequence of the antitonicity of the mapping $\omega \mapsto \mathcal{L}_\omega$. □

By Lemma 5.13 the mapping $\omega \mapsto \mathcal{L}_\omega$ of $\mathfrak{US}^{(n)}$ onto $\mathfrak{L}^{(n)}$ induces a lattice structure on $\mathfrak{L}^{(n)}$ via $\mathcal{L}_{\omega_1} \vee \mathcal{L}_{\omega_2} = \mathcal{L}_{\omega_1 \cap \omega_2}$ and $\mathcal{L}_{\omega_1} \wedge \mathcal{L}_{\omega_2} = \mathcal{L}_{\omega_1 \cup \omega_2}$. This implies the following statement.

Theorem 5.14 *The mapping* $\omega \mapsto \mathcal{L}_\omega$ *is a lattice isomorphism of* $(\mathfrak{US}^{(n)}, \cup, \cap)$ *onto* $(\mathfrak{L}^{(n)}, \wedge, \vee)$.

The last statement of Lemma 5.13 concerning the union of language families can be strengthened as follows.

Lemma 5.15 *Let $\omega_1, \omega_2 \in \mathfrak{US}^{(n)}$. Then there is an n-ary relation ω with $\mathcal{L}_\omega = \mathcal{L}_{\omega_1} \cup \mathcal{L}_{\omega_2}$ if and only if $\mathcal{L}_{\omega_1} \cup \mathcal{L}_{\omega_2} = \mathcal{L}_{\omega_1 \cap \omega_2}$.*

Proof: Suppose that $\mathcal{L}_{\omega_1} \cup \mathcal{L}_{\omega_2} = \mathcal{L}_\omega$ for some n-ary relation ω. We may assume that $\omega \in \mathfrak{US}^{(n)}$. Consider $L \in \mathcal{L}_\omega$. Lemma 5.13 implies that $\omega_1 \cap \omega_2 \subseteq \omega$. Consider $x \in \omega$. Then $\operatorname{cont} x \notin \mathcal{L}_\omega$, hence $\operatorname{cont} x \notin \mathcal{L}_{\omega_1}$ and $\operatorname{cont} x \notin \mathcal{L}_{\omega_2}$. Therefore $x \in \omega_1$ and $x \in \omega_2$. This proves $\omega_1 \cap \omega_2 = \omega$. The converse is obvious. □

We now turn to considering $\underline{\omega}$. While $\Phi(\omega)$ is a complete $\cup$-semilattice it turns out not to be a $\cap$-semilattice in general.

Theorem 5.16 *For $\omega \in \mathfrak{R}^{(n)}$ one has*

$$\mathcal{L}_{\underline{\omega}} = \mathcal{L}_\omega \cup \{L \mid L \subseteq X^+, L \neq \emptyset, \forall u \in L : (u, u, \ldots, u) \notin \omega\}.$$

Proof: Let

$$\mathcal{L} = \{L \mid L \subseteq X^+, L \neq \emptyset, \forall u \in L : (u, u, \ldots, u) \notin \omega\}.$$

We show first that $\mathcal{L}_\omega \cup \mathcal{L} \subseteq \mathcal{L}_{\underline{\omega}}$.

From $\underline{\omega} \subseteq \omega$ it follows that $\mathcal{L}_\omega \subseteq \mathcal{L}_{\underline{\omega}}$ by Lemma 5.3. Consider $L \in \mathcal{L}$ and $x \in [S]^n$ with $\operatorname{cont} x \subseteq L$. We have to prove that $x \notin \underline{\omega}$.

Because of $\underline{\omega} \subseteq \omega$ this is obvious for $x \notin \omega$. Hence we assume that $x \in \omega$. By the definition of $\mathcal{L}$ this implies that $|\operatorname{cont} x| > 1$. Let $x = (x_1, x_2, \ldots, x_n)$ and let i, j be such that $1 \leq i < j \leq n$ and $a_i \neq a_j$. Let

$$x' = (x_1, x_2, \ldots, x_{i-1}, x_j, x_{i+1}, \ldots, x_{j-1}, x_i, x_{j+1}, \ldots, x_n)$$

and consider the relation $\omega' = (\omega \setminus \{x\}) \cup \{x'\}$. In order to prove that $x \notin \underline{\omega}$ it suffices to show that $\omega' \in \Phi(\omega)$ because, in this case, $\underline{\omega} \subseteq \omega \cap \omega'$ with $x \notin \omega'$.

Now suppose that $M \in \mathcal{L}_\omega$ and $M \notin \mathcal{L}_{\omega'}$. Then there is a $y \in \omega'$ such that $\operatorname{cont} y \subseteq M$. It is impossible that $y = x'$ because $\operatorname{cont} x' = \operatorname{cont} x \not\subseteq M$ by $M \in \mathcal{L}_\omega$. However, also $y \neq x'$ is impossible because $y \in \omega$ in this case. This shows $\mathcal{L}_\omega \subseteq \mathcal{L}_{\omega'}$. The proof of the converse inclusion is analogous.

We now turn to proving that $\mathcal{L}_{\underline{\omega}} \subseteq \mathcal{L}_\omega \cup \mathcal{L}$. Suppose there exists a set $L \in \mathcal{L}_{\underline{\omega}}$ such that $L \notin \mathcal{L}_\omega \cup \mathcal{L}$. Hence there is an element $u \in L$ with $(u, u, \ldots, u) \in \omega$ and $(u, u, \ldots, u) \notin \underline{\omega}$. This implies the existence of a relation $\omega' \in \Phi(\omega)$ with $(u, u, \ldots, u) \notin \omega'$. Thus $\{u\} \in \mathcal{L}_{\omega'}$ while $\{u\} \notin \mathcal{L}_\omega$, a contradiction! Hence $\mathcal{L}_{\underline{\omega}} \subseteq \mathcal{L}_\omega \cup \mathcal{L}$. □

Corollary 5.17 *For every S with $|S| > 1$ and every $n \in \mathbb{N}$, $n > 1$, there is an n-ary relation ω on S such that $\Phi(\omega)$ is not a $\cap$-semilattice.*

Proof: Let $a, b \in S$, $a \neq b$. For $n > 1$ choose $\omega = \{(a, a, \ldots, a, b)\}$ and $\omega' = \{(b, b, \ldots, b, a)\}$. Then $\mathcal{L}_{\omega'} = \mathcal{L}_\omega$, that is, $\omega' \in \Phi(\omega)$ while $\omega \cap \omega' = \emptyset \notin \Phi(\omega)$. □

Remark 5.18 *If ω is a unary relation on S then $\Phi(\omega) = \{\omega\}$. Moreover, one has $L \in \mathcal{L}_\omega$ if and only if $L \cap \omega = \emptyset$.*

By Theorem 5.16 one always has the inclusion $\mathcal{L}_\omega \subseteq \mathcal{L}_{\underline{\omega}}$. The following example illustrates the general situation when the inclusion is proper.

Example 5.19 Consider $X = \{a, b\}$, $S = X^+$, and $n = 3$. Let $\omega = \{(a, b, b)\}$. Then $\varphi = \{(a, a, b)\} \in \Phi(\omega)$ and $\underline{\omega} = \emptyset$. We have $\{a, b\} \in \mathcal{L}_{\underline{\omega}}$ but $\{a, b\} \notin \mathcal{L}_\omega$.

For any non-trival family $\mathcal{L}$ of subsets of S and any $n \in \mathbb{N}$ we consider the sets

$$\Omega_n(\mathcal{L}) = \{\omega \mid \omega \in \mathfrak{US}^{(n)}, \mathcal{L} \subseteq \mathcal{L}_\omega\}$$

and

$$\omega_n(\mathcal{L}) = \{\omega \mid \omega \in \mathfrak{US}^{(n)}, \mathcal{L}_\omega \subseteq \mathcal{L}\}$$

of n-ary relations. These are the obvious counterparts of the sets$_{\text{old}}$ $\Omega(\mathcal{L})$ and $\omega(\mathcal{L})$ of Section 2.1. However, note that due to the change in definition of independence these sets are actually not quite the same in the case of $n = 2$.

Both $\Omega_n(\mathcal{L})$ and $\omega_n(\mathcal{L})$ are always non-empty. The former contains at least the empty relation. The latter contains at least the *universal* relation $\omega = [S]^n$ for which $\mathcal{L} = \emptyset$.

Clearly, if $\mathcal{L} \in \mathfrak{L}^{(n)}$, $\mathcal{L} = \mathcal{L}_\omega$ with $\omega \in \mathfrak{US}^{(n)}$ say, then $\omega \in \Omega_n(\mathcal{L}) \cap \omega_n(\mathcal{L})$. In the sequel we show that, in this case, ω is the maximum of $\Omega_n(\mathcal{L})$ and the minimum of $\omega_n(\mathcal{L})$. As a preparation we generalize the corresponding results of Section 2.1 concerning the structure of these sets.

Remark 5.20 *For every non-trivial family $\mathcal{L}$ of subsets of S and every $n \in \mathbb{N}$ one has*

$$\Omega_n(\mathcal{L}) \cap \omega_n(\mathcal{L}) = \begin{cases} \emptyset, & \text{if } \mathcal{L} \notin \mathfrak{L}^{(n)}, \\ \{\omega\}, & \text{otherwise, where } \omega \in \mathfrak{US}^{(n)} \text{ and } \mathcal{L} = \mathcal{L}_\omega. \end{cases}$$

Theorem 5.21 *Let $\mathcal{L}$ be a non-trivial family of subsets of S, let $n \in \mathbb{N}$, and let $\omega_1, \omega_2 \in \mathfrak{US}^{(n)}$.*

(1) Let $\omega_0 \in \Omega_n(\mathcal{L})$. If $\omega_1 \subseteq \omega_0$ then $\omega_1 \in \Omega_n(\mathcal{L})$. Moreover, $\omega_0 \cap \omega_2 \in \Omega_n(\mathcal{L})$ and $\Omega_n(\mathcal{L})$ is a $\cap$-complete $\cap$-semilattice with the empty relation as its minimum.

(2) $\Omega_n(\mathcal{L})$ is a $\cup$-complete $\cup$-semilattice. Hence $\Omega_n(\mathcal{L})$ has a maximum.

(3) Let $\omega_0 \in \omega_n(\mathcal{L})$. If $\omega_0 \subseteq \omega_1$ then $\omega_1 \in \omega_n(\mathcal{L})$. Moreover, $\omega_0 \cup \omega_2 \in \omega_n(\mathcal{L})$, and $\omega_n(\mathcal{L})$ is a $\cup$-complete $\cup$-semilattice with the universal relation as its maximum.

Proof: For (1), let $\omega_1 \subseteq \omega_0 \in \Omega_n(\mathcal{L})$. Then $\mathcal{L} \subseteq \mathcal{L}_{\omega_0} \subseteq \mathcal{L}_{\omega_1}$, thus $\omega_1 \in \Omega_n(\mathcal{L})$. As $\omega_0 \cap \omega_2 \subseteq \omega_0$ one has $\omega_0 \cap \omega_2 \in \Omega_n(\mathcal{L})$. Let $\{\omega_i \mid i \in I\}$ be any family of relations with $i\omega_i \in \Omega_n(\mathcal{L})$ for $i \in I$. Then

$$\omega = \bigcap_{i \in I} \omega_i \subseteq \omega_j$$

for every $j \in I$. Hence $\omega \in \Omega_n(\mathcal{L})$. As $\Omega_n(\mathcal{L})$ contains the empty relation, it follows that $\cap_{\omega \in \Omega_n(\mathcal{L})} \omega = \emptyset$.

For (2), consider a family $\{\omega_i \mid i \in I\}$ of relations $\omega_i \in \Omega_n(\mathcal{L})$ for $i \in I$. Let

$$\omega = \bigcup_{i \in I} \omega_i.$$

We show that $\omega \in \Omega_n(\mathcal{L})$, that is, that $\mathcal{L} \subseteq \mathcal{L}_\omega$. Assume the contrary. Then there is $L \in \mathcal{L}$ such that L is not ω-independent. Therefore, for some $x \in [S]^n$ we have $x \in \omega$ and $\operatorname{cont} x \subseteq L$. Hence there exists an $i \in I$ such that $x \in \omega_i$. But then $L \notin \mathcal{L}_{\omega_i}$, that is, $\omega_i \notin \Omega_n(\mathcal{L})$, a contradiction! Clearly, the relation

$$\bigcup_{\omega \in \Omega_n(\mathcal{L})} \omega$$

is the maximum of $\Omega_n(\mathcal{L})$.

For (3), let $\omega_0 \in \omega_n(\mathcal{L})$ and $\omega_0 \subseteq \omega_1$. Then $\mathcal{L}_{\omega_1} \subseteq \mathcal{L}_{\omega_0} \subseteq \mathcal{L}$. Thus $\omega_1 \in \omega_n(\mathcal{L})$. Since $\omega_0 \subseteq \omega_0 \cup \omega_2$, $\omega_0 \cup \omega_2 \in \omega_n(\mathcal{L})$. Let $\{\omega_i \mid i \in I\}$ be a family with $\omega_i \in \omega_n(\mathcal{L})$ for $i \in I$. Then

$$\omega_j \subseteq \omega = \bigcup_{i \in I} \omega_i$$

for every $j \in I$. Hence $\omega \in \omega_n(\mathcal{L})$. As $\omega_n(\mathcal{L})$ contains the universal relation, it follows that

$$\bigcup_{\omega \in \omega_n(\mathcal{L})} \omega$$

is the universal relation. □

For a set L, let $\operatorname{Fin}(L)$ be the set of non-empty, finite subsets of L. A family $\mathcal{L}$ of subsets of S is said to be Fin-*determined* if it is non-trivial and

$$L \in \mathcal{L} \iff \operatorname{Fin}(L) \subseteq \mathcal{L}$$

holds true.

Theorem 5.22 *Let $\mathcal{L}$ be a non-trivial, Fin-determined family of subsets of S and let $n \in \mathbb{N}$. Then the following properties obtain:*

(1) The relation

$$\omega_{\mathcal{L}}^{(n)} = \{x \mid x \in [S]^n, \text{cont}\, x \notin \mathcal{L}\}$$

is the maximum of $\Omega_n(\mathcal{L})$.

(2) For every $\omega_0 \in \omega_n(\mathcal{L})$ there is a minimal element $\omega_\infty \in \omega_n(\mathcal{L})$ such that $\omega_\infty \subseteq \omega_0$.

Proof: Let ω be the maximum of $\Omega_n(\mathcal{L})$. Then $\omega, \omega_{\mathcal{L}}^{(n)} \in \mathfrak{US}^{(n)}$. Hence in order to prove that $\omega = \omega_{\mathcal{L}}$ it is sufficient to show that $\mathcal{L}_\omega = \mathcal{L}_{\omega_{\mathcal{L}}^{(n)}}$. Consider $L \in \mathcal{L}_\omega$ and suppose that $L \notin \mathcal{L}_{\omega_{\mathcal{L}}^{(n)}}$. Then there exists $x \in \omega_{\mathcal{L}}^{(n)}$ such that $\text{cont}\, x \subseteq L$. By the definition of $\omega_{\mathcal{L}}^{(n)}$ this implies that $\text{cont}\, x \notin \mathcal{L}$. Now consider the relation

$$\tilde{\omega} = \text{upsymm}(\omega \cup \{x\}).$$

Then $L' \in \mathcal{L}_\omega \setminus \mathcal{L}_{\tilde{\omega}}$ if and only if $\text{cont}\, x \subseteq L'$. But for such L' one has $L' \notin \mathcal{L}$ as $\mathcal{L}$ is Fin-determined. Therefore, $\tilde{\omega} \in \Omega_n(\mathcal{L})$. However, ω is a proper subset of $\tilde{\omega}$ contradicting the maximality of ω. This shows $L \in \mathcal{L}_{\omega_{\mathcal{L}}^{(n)}}$, that is $\mathcal{L}_\omega \subseteq \mathcal{L}_{\omega_{\mathcal{L}}^{(n)}}$.

Conversely, let $L \in \mathcal{L}_{\omega_{\mathcal{L}}^{(n)}}$. Then $\text{cont}\, x \in \mathcal{L}$ for all $x \in [S]^n$ with $\text{cont}\, x \subseteq L$. Therefore, $x \notin \tilde{\omega}$ for every $\tilde{\omega} \in \Omega_n(\mathcal{L})$. In particular, $x \notin \omega$, that is $L \in \mathcal{L}_\omega$. This proves $\mathcal{L}_{\omega_{\mathcal{L}}^{(n)}} \subseteq \mathcal{L}_\omega \subseteq \mathcal{L}_{\omega_{\mathcal{L}}^{(n)}}$, that is, $\omega_{\mathcal{L}}^{(n)}$ is the maximum of $\Omega_n(\mathcal{L})$.

To prove the second statement, assume that ω_0 is not minimal—otherwise nothing needs to be proved. Consider a decreasing chain

$$\omega_0 \supseteq \omega_1 \supseteq \omega_2 \supseteq \dots$$

in $\omega_n(\mathcal{L})$ and let $\omega = \bigcap_{i \geq 0} \omega_i$. Let $L \in \mathcal{L}_\omega$. We show that $L \in \mathcal{L}$. As $\mathcal{L}$ is Fin-determined, it is sufficient to show that $L' \in \mathcal{L}$ for every $L' \in \text{Fin}(L)$.

Let $L' \in \text{Fin}(L)$. For every $x \in [S]^n$ with $\text{cont}\, x \subseteq L'$ there is an index $i(x)$ such that $x \notin \omega_{i(x)}$. Let

$$j = \max\{i(x) \mid x \in [X^+]^n, \text{cont}\, x \subseteq L'\}.$$

As L' is finite, j exists. Then $L' \in \mathcal{L}_{\omega_j} \subseteq \mathcal{L}$. This implies $L \in \mathcal{L}$ and therefore $\omega \in \omega_n(\mathcal{L})$. Now the existence of ω_∞ follows by Zorn's lemma. □

Theorem 5.23 *Let $\mathcal{L}$ be a non-trivial family of subsets of S.*

(1) If $\mathcal{L} = \mathcal{L}_{\tilde{\omega}}$ for some n-ary relation $\tilde{\omega}$ then $\omega_n(\mathcal{L})$ has a minimum ω and

$$\omega = \operatorname{upsymm} \tilde{\omega} = \max \Omega_n(\mathcal{L}).$$

(2) If $\mathcal{L}$ is Fin-determined and $\omega_n(\mathcal{L})$ has a minimum ω then $\mathcal{L} = \mathcal{L}_\omega$ and $\omega = \max \Omega_n(\mathcal{L})$.

Proof: Suppose that $\mathcal{L} = \mathcal{L}_{\tilde{\omega}}$. Without loss of generality we assume that $\tilde{\omega}$ is upward symmetric. Then $\tilde{\omega} \in \omega_n(\mathcal{L})$. Let $\omega_1 \in \omega_n(\mathcal{L})$. Then $\mathcal{L}_{\omega_1} \subseteq \mathcal{L} = \mathcal{L}_{\tilde{\omega}}$ by the definition of $\omega_n(\mathcal{L})$, hence $\tilde{\omega} \subseteq \omega_1$ using the fact that ω_1 is upward symmetric. This proves that $\omega_n(\mathcal{L})$ has a minimum and that $\tilde{\omega}$ is this minimum. Clearly, $\tilde{\omega} \in \Omega_n(\mathcal{L})$, hence $\tilde{\omega} = \max \Omega_n(\mathcal{L})$. This proves statement (1).

For part (2) we only need to show that $\mathcal{L} = \mathcal{L}_\omega$. By the definition of $\omega_n(\mathcal{L})$ one has $\mathcal{L}_\omega \subseteq \mathcal{L}$. Suppose there is a set $L \in \mathcal{L} \setminus \mathcal{L}_\omega$. Then there is $x \in [S]^n$ with $x \in \omega$ and $\operatorname{cont} x \in \mathcal{L}$ as $\mathcal{L}$ is Fin-determined. Let

$$\tilde{\omega} = \{y \mid y \in [S]^n, \operatorname{cont} y \not\subseteq \operatorname{cont} x\}.$$

Clearly, $\tilde{\omega} \in \mathfrak{US}^{(n)}$. Moreover,

$$\mathcal{L}_{\tilde{\omega}} = \{L \mid \emptyset \neq L \subseteq \operatorname{cont} x\}.$$

As $\mathcal{L}$ is Fin-determined this implies $\mathcal{L}_{\tilde{\omega}} \subseteq \mathcal{L}$, hence $\tilde{\omega} \in \omega_n(\mathcal{L})$. As $\mathcal{L}$ is Fin-determined there is a minimal element $\hat{\omega}$ of $\omega_n(\mathcal{L})$ with $\hat{\omega} \subseteq \tilde{\omega}$. From $x \in \omega$ and $x \notin \hat{\omega}$ it follows that $\hat{\omega} \neq \omega$, contradicting the fact that ω is unique. Thus $\mathcal{L} = \mathcal{L}_\omega$.

□

Let $\mathfrak{F}$ denote the set of non-trivial families of subsets of S. The following lemmata generalize the corresponding statements of Section 2.1. For arbitrary $\mathcal{L} \in \mathfrak{F}$ let $\omega_{\mathcal{L}}^{(n)} = \max \Omega_n(\mathcal{L})$, generalizing the notation introduced above.

Lemma 5.24 *$\mathfrak{F}$ is a complete lattice with $\emptyset$ and $2^S \setminus \{\emptyset\}$ as its minimum and maximum, respectively.*

Lemma 5.25 *Let $\mathcal{L}_1, \mathcal{L}_2 \in \mathfrak{F}$ with $\mathcal{L}_1 \subseteq \mathcal{L}_2$. Then $\omega_{\mathcal{L}_2}^{(n)} \subseteq \omega_{\mathcal{L}_1}^{(n)}$.*

Lemma 5.26 *The mapping*

$$\mathcal{L} \mapsto \mathcal{L}_{\omega_{\mathcal{L}}^{(n)}}$$

is a closure operator on $\mathfrak{F}$.

These results enable us to generalize the "gap theorem," that is, Theorem 2.21, to n-ary relations. It provides a powerful criterion by which to prove of a given family of sets whether it can be characterized as the independent sets with respect an n-ary relation.

Theorem 5.27 (Gap Theorem) *Let $n \in \mathbb{N}$ and $\mathcal{L}, \mathcal{L}_1 \in \mathfrak{F}$ such that*

$$\mathcal{L} \subseteq \mathcal{L}_1 \subsetneq \mathcal{L}_{\omega_{\mathcal{L}}^{(n)}}.$$

Then there is no n-ary relation ω such that $\mathcal{L}_1 = \mathcal{L}_\omega$.

Proof: By Lemma 5.26 one has

$$\mathcal{L}_{\omega_{\mathcal{L}}^{(n)}} = \mathcal{L}_{\omega_{\mathcal{L}_1}^{(n)}}.$$

Assume that $\mathcal{L}_1 = \mathcal{L}_\omega$ for some n-ary relation ω. Without loss of generality we may assume that $\omega \in \mathfrak{U}\mathfrak{S}^{(n)}$. Then

$$\omega = \max \Omega_n(\mathcal{L}_1) = \omega_{\mathcal{L}_1}^{(n)} = \omega_{\mathcal{L}}^{(n)},$$

hence

$$\mathcal{L}_1 = \mathcal{L}_{\omega_{\mathcal{L}}^{(n)}},$$

a contradiction! $\square$

5.2 Binary Relations Re-Visited

As mentioned in section 5.1, the new definitions of that section are inconsistent with the earlier one of Section 2.1. In the present section we give the necessary hints for a "translation." Moreover we point out a few new results that can be achieved for binary relations given our modified definitions. In the sequel we assume that $S = X^+$ where X is an alphabet with $|X| \geq 2$.

In Section 2.1, the definition of ω-independence$_{\text{old}}$ for $\omega \in [S]^2$ made no distinction between ω and its reflexive closure. To obtain the same results, we have to re-define the binary relations considered there by eliminating the diagonal. These re-defined relations are listed in Table 5.1 below.

Given the new definition of ω-independence, if ω is a reflexive binary relation then $\mathcal{L}_\omega = \emptyset$. In detail, if ω is a binary relation and $(u, u) \in \omega$ then $u \notin L$ for every $L \in \mathcal{L}_\omega$, that is, diagonal elements in ω exclude the corresponding elements from every ω-independent set. In this way, the diagonal elements of ω "behave like a unary relation." This idea is to be explored in general terms in Section 5.3 below. Here we just provide another example.

Recall that a non-empty language $L \subseteq X^+$ is called a ***solid code*** if every word $w \in X^+$ has a unique decomposition of the form $w = x_1 y_1 x_2 y_2 \cdots x_n y_n x_{n+1}$ with $y_1, \ldots, y_n \in L$ and $\{u \mid u \leq_i x_i\} \cap L = \emptyset$ for $i = 1, \ldots, n+1$. We now give a characterization of the class of solid codes in terms of a binary relation. Observe that such a characterization was impossible with the earlier notion of independence.

A word $w \in X^+$ is said to be an ***overlap*** of words $u, v \in X^+$ if $w <_p u$ and $w <_s v$ or if $w <_s u$ and $w <_p v$. The ***overlap relation*** ω_{ov} is defined by

$$(u, v) \in \omega_{ov} \iff \text{there is an overlap } w \in X^+ \text{ of } u \text{ and } v$$

for $u, v \in X^+$. Obviously, ω_{ov} is symmetric, but not upward symmetric. The languages in $\mathcal{L}_{\omega_{ov}}$ are the ***overlap-free*** languages. Now define the binary relation

languages	property	relation
uniform codes	u	$w <_u v \iff \lvert w\rvert < \lvert v\rvert$
hypercodes	h	$w <_h v \iff w \neq v \wedge \exists n, \exists x_1, \ldots, x_n \in X^* : w = x_1 \cdots x_n \wedge v \in X^* x_1 X^* \cdots x_n X^*$
n-shuffle codes	sh_n	$w \omega_{sh_n} v \iff w \neq v \wedge \exists x_1, \ldots, x_n : w = x_1 \cdots x_n \wedge v \in X^* x_1 X^* \cdots x_n X^*$
prefix codes	p	$w <_p v \iff v \in wX^+$
suffix codes	s	$w <_s v \iff v \in X^+ w$
bifix codes	b	$\omega_b = <_p \cup <_s$
2-codes	c	$w <_c v \iff \exists x \in X^+ : v = wx = xw$
infix codes	i	$w <_i v \iff w \neq v \wedge v \in X^* w X^*$
p-infix codes	pi	$w <_{pi} v \iff v \in X^* w X^+$
s-infix codes	si	$w <_{si} v \iff v \in X^+ w X^*$
outfix codes	o	$w \omega_o v \iff \exists w_1, w_2 : w = w_1 w_2 \wedge v \in w_1 X^+ w_2$
solid codes	σ	$\omega_\sigma = \omega_{ov} \cup \mathrm{symm} <_i$ where $w \omega_{ov} v \iff$ there is an overlap $u \in X^+$ of w and v
2-ps-codes	d	$<_d = <_p \cap <_s$

Table 5.1. Some classes of languages and their defining properties, given the new definition of independence. In all cases one would have to take the (upward) symmetric closure of the defining relation.

ω_σ on X^+ as

$$\omega_\sigma = \omega_{ov} \cup \text{symm} \underset{i}{<} .$$

The relation ω_σ is also symmetric and again not upward symmetric.

Theorem 5.28 *$L \in \mathcal{L}_{\omega_\sigma}$ if and only if L is a solid code.*

Proof: In [Jür3] the following conditions are shown to be necessary and sufficient for a non-empty language $L \subseteq X^+$ to be a solid code:

(1) For any $u, v \in L$, not necessarily distinct, u and v do not have an overlap.

(2) For any $u, v \in L$, if $u \neq v$ then $u \notin X^*vX^*$.

The first condition is captured by ω_{ov}. The second one is captured by symm $<_i$. By Lemma 5.13 $\mathcal{L}_{\omega_{ov} \cup \text{symm} <_i} = \mathcal{L}_{\omega_{ov}} \cap \mathcal{L}_{<_i}$. □

5.3 Countable Sets of Finitary Relations, Relations of Different Arities, and Their Independent Sets

The construction of the binary relation ω_σ which characterizes the solid codes could have been carried out differently. Given the conditions derived in [Jür3], the solid codes can be considered as languages which are independent with respect to two relations, the binary relation $\omega_\sigma \setminus \{(u,u) \mid u \in X^+\}$ and the unary relation $\{u \mid (u,u) \in \omega_{ov}\}$. This idea can be captured formally and generalized as follows:

Again let S be an arbitrary, non-empty set. Let $\Psi = \{\psi_j \mid j \in J\}$ be a family of finitary relations on S where J is an index set and where ν_j is the arity of ψ_j. A set $L \subseteq S$ is said to be *Ψ-independent* if it is ψ_j-independent for every $j \in J$. Let $\mathcal{L}_\Psi$ denote the family of ψ-independent subsets of S, that is, $\mathcal{L}_\Psi = \bigcap_{j \in J} \mathcal{L}_{\psi_j}$. In the sequel we restrict the attention to countable index sets J, that is, we assume that $J \subseteq \mathbb{N}$. Some generalizations to uncountable index sets are possible, but will not be considered.

If J is infinite then, without loss of generality, we may assume that $J = \mathbb{N}$. If J is finite we may assume that $J = \{1, \ldots, n\}$ for some $n \in \mathbb{N}$. In this case, instead of Ψ, we consider the family

$$\Psi' = \left\{ \psi'_j \,\middle|\, j \in \mathbb{N}, \psi'_j = \begin{cases} \psi_j, & \text{if } j \le n, \\ \psi_n, & \text{if } j > n \end{cases} \right\}.$$

Clearly, L is Ψ-independent if and only if it is Ψ'-independent. Therefore, to simplify notation we can always assume that $J = \mathbb{N}$. For $j \in \mathbb{N}$ let $\overline{\nu}_j = \max_{h \le j} \nu_h$.

For $n \in \mathbb{N}$ let ω be an arbitrary n-ary relation on S and let $m \in \mathbb{N}$ with $m \ge n$. We construct an m-ary relation denoted by $\text{upsymm}_m \omega$ from ω such that $\mathcal{L}_\omega = \mathcal{L}_{\text{upsymm}_m \omega}$. This relation is referred to as the *m-ary completion* of ω. It is defined as follows:

$$\text{upsymm}_m \omega = \text{upsymm}\{(x_1, \ldots, x_m) \mid (x_1, \ldots, x_n) \in \omega, x_{n+1} = \ldots = x_m = x_n\}.$$

Lemma 5.29 *For $n, m \in \mathbb{N}$ with $n \le m$ and $\omega \subseteq [S]^n$ one has*

$$\mathcal{L}_\omega = \mathcal{L}_{\text{upsymm}_m \omega}.$$

Proof: Consider L with $\emptyset \ne L \subseteq S$. If $L \notin \mathcal{L}_\omega$ then there is $x \in \omega$ such that $\text{cont}\, x \subseteq L$. Letting $x = (x_1, \ldots, x_n)$ one has $(x_1, \ldots, x_n, x_n, \ldots, x_n) \in \text{upsymm}_m \omega$. Thus $L \notin \mathcal{L}_{\text{upsymm}_m \omega}$.

Conversely, assume that $L \notin \mathcal{L}_{\text{upsymm}_m \omega}$. Then there is an $x \in \text{upsymm}_m \omega$ such that $\text{cont}\, x \subseteq L$. By the definition of $\text{upsymm}_m \omega$ there is an $x' \in \omega$ such that $\text{cont}\, x' \subseteq \text{cont}\, x$. Therefore, $L \notin \mathcal{L}_\omega$. □

Lemma 5.30 *Let $\Psi = \{\psi_j \mid j \in \mathbb{N}\}$ be a family of finitary relations on S with ν_j the arity of ψ_j. Then*

$$\mathcal{L}_\Psi = \bigcap_{j \in \mathbb{N}} \mathcal{L}_{\bigcup_{h \le j} \text{upsymm}_{\overline{\nu}_j} \omega_h}.$$

Proof: By Lemma 5.13 one has

$$\bigcap_{h\leq j} \mathcal{L}_{\mathrm{upsymm}_{\overline{\nu}_j}\,\omega_h} = \mathcal{L}_{\bigcup_{h\leq j}\mathrm{upsymm}_{\overline{\nu}_j}\,\omega_h}.$$

Thus

$$\bigcap_{j\in\mathbf{N}} \mathcal{L}_{\bigcup_{h\leq j}\mathrm{upsymm}_{\overline{\nu}_j}\,\omega_h} = \bigcap_{j\in\mathbf{N}}\bigcap_{h\leq j} \mathcal{L}_{\mathrm{upsymm}_{\overline{\nu}_j}\,\omega_h} = \mathcal{L}_{\Psi}.$$

□

If, in particular, Ψ contains only finitely many different finitary relations then $\max_{j\in\mathbf{N}}\nu_j$ exists and, therefore, the sequence of relations $\bigcup_{h\leq j}\mathrm{upsymm}_{\overline{\nu}_j}\,\omega_h$ converges to a finitary relation on S. This proves the following property.

Lemma 5.31 *Let Ψ be a finite family of finitary relations on S and let n be the maximum of the arities of the relations in Ψ. Then there is an n-ary relation $\omega(\Psi) \in \mathfrak{U}\mathfrak{S}^{(n)}$ such that $\mathcal{L}_{\omega(\Psi)} = \mathcal{L}_{\Psi}$. Moreover, $\omega(\Psi)$ is the union of the n-ary completions of the relations in Ψ.*

We can use these results for instance to prove the following "hierarchy theorem."

Theorem 5.32 *Let $\mathcal{L}$ be an nontrivial family of subsets of S. The following properties obtain:*

(1) For every $n \in \mathbf{N}$ one has

$$\mathcal{L} \subseteq \mathcal{L}_{\omega_{\mathcal{L}}^{(n+1)}} \subseteq \mathcal{L}_{\omega_{\mathcal{L}}^{(n)}}.$$

(2) For every $n \in \mathbf{N}$, if

$$\mathcal{L}_{\omega_{\mathcal{L}}^{(n+1)}} \neq \mathcal{L}_{\omega_{\mathcal{L}}^{(n)}}$$

then there is no n-ary relation ω such that

$$\mathcal{L}_{\omega} = \mathcal{L}_{\omega_{\mathcal{L}}^{(n+1)}}.$$

Proof: For (1), one has

$$\mathcal{L}_{\text{upsymm}_{n+1}\,\omega_{\mathcal{L}}^{(n)}} = \mathcal{L}_{\omega_{\mathcal{L}}^{(n)}}$$

and

$$\text{upsymm}_{n+1}\,\omega_{\mathcal{L}}^{(n)} \subseteq \max \Omega_{n+1}(\mathcal{L}) = \omega_{\mathcal{L}}^{(n+1)},$$

hence the inclusion. Statement (2) follows from Theorem 5.27. □

It seems natural to conjecture that the sequence of families $\mathcal{L}_{\omega_{\mathcal{L}}^{(n)}}$ would have the following stabilization property: If $\mathcal{L}_{\omega_{\mathcal{L}}^{(n+1)}} = \mathcal{L}_{\omega_{\mathcal{L}}^{(n)}}$ for some $n \in \mathbb{N}$ then $\mathcal{L}_{\omega_{\mathcal{L}}^{(n+m)}} = \mathcal{L}_{\omega_{\mathcal{L}}^{(n)}}$ for all $m \in \mathbb{N}$. It turns out that this is not true as shown in the following example.

Example 5.33 Let $S = \{a_1, a_2, \ldots, a_n\}$ with $n \geq 6$ and let $S_k = \{a_1, a_2, \ldots, a_k\}$ for $k \leq n$. Define

$$\mathcal{L}_{m,k} = \{X \mid X \subseteq S_k, |X| = m\}$$

and

$$\mathcal{L} = \mathcal{L}_{1,n} \cup \mathcal{L}_{2,5} \cup \mathcal{L}_{3,5} \cup \mathcal{L}_{4,4}.$$

Then

$$\mathcal{L}_{\omega_{\mathcal{L}}^{(4)}} \subsetneq \mathcal{L}_{\omega_{\mathcal{L}}^{(3)}} = \mathcal{L}_{\omega_{\mathcal{L}}^{(2)}} \subsetneq \mathcal{L}_{\omega_{\mathcal{L}}^{(1)}}.$$

5.4 n-ary Relations and Codes

In this section, we apply our theory to n-codes, n-intercodes, and n-ps-languages. We strengthen the hierarchy results obtained in [Ito2] and add to those obtained in [Shy7].

Recall that a language L is an n-code if every non-empty subset of L of cardinality at most n is a code. As an immediate consequence one has the following results for the hierarchy of n-codes.

Theorem 5.34 *For $n \in \mathbb{N}$ let γ_n be the n-ary relation on X^+ given by*

$$x \in \gamma_n \iff \text{cont}\, x \notin C$$

for $x \in [X^+]^n$. The following properties obtain:

(1) $C_n = \mathcal{L}_{\gamma_n}$ for all n.

(2) $\gamma_n = \omega_C^{(n)} = \omega_{C_m}^{(n)}$ for all n and all m with $m \geq n$.

(3) There is no finitary relation ω such that $C = \mathcal{L}_\omega$.

(4) For all $m \in \mathbb{N}$, there is no n-ary relation ω with $\mathcal{L}_\omega = C_m$ and $n < m$.

Proof: Statements (1) and (2) follow directly from the definition of γ_n. Statements (3) and (4) are consequences of Theorem 5.27. □

A hierarchy similar to that of n-codes can be built within the class of intercodes. Recall that a language L is called an *intercode of index* m if $\emptyset \neq L \subseteq X^+$ and $L^{m+1} \cap X^+ L^m X^+ = \emptyset$. Let $\mathcal{I}_m$ denote the family of all intercodes of index m over X and let $\mathcal{I} = \bigcup_{m \in \mathbb{N}} \mathcal{I}_m$ denote the family of all intercodes. An *n-intercode of index m* is a non-empty language L, $L \subseteq X^+$, such that every non-empty subset of L of cardinality at most n is an intercode of index m. Let $\mathcal{I}_{n,m}$ denote the family of n-intercodes of index m over X and let $\mathcal{I}_{n,\infty} = \bigcup_{m \in \mathbb{N}}$ denote the family of n-intercodes. Note that

$$\mathcal{I}_m = \bigcap_{n \in \mathbb{N}} \mathcal{I}_{n,m} \quad \text{and} \quad \mathcal{I} = \bigcap_{n \in \mathbb{N}} \mathcal{I}_{n,\infty}.$$

In [Shy7] it is shown that a non-empty language L, $L \subseteq X^+$ is an intercode of index m if and only if it is a $(2m+1)$-intercode of index m, that is, $\mathcal{I}_m = \mathcal{I}_{2m+1,m}$.

To describe n-intercodes of index m, we consider the n-ary relation $\varrho_{n,m}$ on X^+ given by the following condition:

$$x \in \varrho_{n,m} \iff (\text{cont}\, x)^{m+1} \cap X^+ (\text{cont}\, x)^m X^+ \neq \emptyset$$

for all $x \in [X^+]^n$. Clearly, $\mathcal{L}_{\varrho_{n,m}} = \mathcal{I}_{n,m}$. The relation between intercodes and codes is described by the following result. The complete situation is illustrated in Figure 3 in Section 1.1.

Theorem 5.35 *For every $n, m \in \mathbb{N}$ the following statements hold true:*

(1) If $n \geq 2$ then $\mathcal{I}_m \subseteq \mathcal{I}_{n,m} \subsetneq \mathcal{I}_{n,\infty} \subsetneq \mathcal{C}_b$ and $\mathcal{I}_m \subsetneq \mathcal{I} \subsetneq \mathcal{I}_{n,\infty}$. Hence every n-intercode with $n \geq 2$ is a bifix code.

(2) If $n \geq 2m+1$ then $\mathcal{I}_{n,m} = \mathcal{I}_m$.

(3) If $1 \leq n < 2m$ then $\mathcal{I}_{n+1,m} \subsetneq \mathcal{I}_{n,m}$.

(4) If $n \geq 2$ and $n \leq 2m+1$ then $\mathcal{I}_{n,m} \subsetneq \mathcal{I}_{n,m+1}$.

(5) $\mathcal{I}_{n+1,\infty} \subsetneq \mathcal{I}_{n,\infty}$

(6) $\mathcal{I}_{1,\infty} = \mathcal{I}_{1,m} = \mathcal{Q} \subseteq \mathcal{C}_2$ and $\mathcal{I}_{1,\infty} \not\subseteq \mathcal{C}_n$ for $n > 2$.

(7) $\mathcal{I}_m \subsetneq \mathcal{I}_{m+1}$.

(8) $\mathcal{I}_{2,\infty} \subsetneq \mathcal{Q} \cap \mathcal{C}_b$.

Proof: For (1), consider a non-empty language $L \subseteq X^+$ which is not a bifix code. Then there are distinct words $u, v \in L$ such that $u <_p v$ or $u <_s v$. Suppose $u <_p v$ holds, that is, $v = uy$ for some $y \in X^+$. Consider $x \in [X^+]^n$ with $\operatorname{cont} x = \{u, v\}$. Then $u^m v \in (\operatorname{cont} x)^{m+1} \cap X^+ (\operatorname{cont} x)^m X^+$. Thus $x \in \varrho_{n,m}$. This implies $L \notin \mathcal{I}_{n,m}$. This shows the inclusion $\mathcal{I}_{n,m} \subseteq \mathcal{C}_b$, hence also the inclusion $\mathcal{I}_{n,\infty} \subseteq \mathcal{C}_b$.

The inclusions $\mathcal{I}_m \subseteq \mathcal{I}_{n,m} \subseteq \mathcal{I}_{n,\infty}$ and $\mathcal{I}_m \subseteq \mathcal{I} \subseteq \mathcal{I}_{n,\infty}$ are immediate consequences of the definition. The inequalities $\mathcal{I}_{n,m} \neq \mathcal{I}_{n,\infty}$, $\mathcal{I}_m \neq \mathcal{I}$, and $\mathcal{I} \neq \mathcal{I}_{n,\infty}$ will follow from statements (4), (7), and (5), respectively. Here we need to prove only the inequality $\mathcal{I}_{n,\infty} \neq \mathcal{C}_b$. Suppose $a \in X$. Then $\{a^2\}$ is a bifix code, but $\{a^2\} \notin \mathcal{I}_{n,\infty}$ for $n \in \mathbb{N}$.

By [Shy7] one has $\mathcal{I}_{2m+1,m} = \mathcal{I}_m$. By definition, $\mathcal{I}_{n+1,m} \subseteq \mathcal{I}_{n,m}$. Using the fact that $\mathcal{I}_m = \bigcap_{n \in \mathbb{N}} \mathcal{I}_{n,m}$ one obtains statement (2).

To prove statement (3), we give examples of languages $L_n \in \mathcal{I}_{n,m} \setminus \mathcal{I}_{n+1,m}$.

Let $X = \{a, b\}$ and $w_i = a^i b^i a^i$ for $i \in \mathbb{N}$. Now define L_n as

$$\{w_2, w_3, \ldots, w_{n+1}, w_1 w_2^{m-n+1} w_3 \cdots w_{n+2}\}$$

if $n \leq m$, as

$$\{w_i w_{i+1} \mid i = 1, 2, \ldots, n-1\} \cup \{w_{n+1}, w_n w_{n+1}^{m-(n-1)/2} w_{n+2}\}$$

if $m < n < 2m$ and n is odd, as

$$\{w_i w_{i+1} \mid i = 1, 2, \ldots, n-2\} \cup \{w_n, w_{n+1}, w_{n-1} w_n w_{n+1}^{m-n/2} w_{n+2}\}$$

if $m < n < 2m$ and n is even, and as

$$\{w_i w_{i+1} \mid i = 1, 2, \ldots, n+1\}$$

if $n = 2m$. One then verifies that L_n has the required properties.

To prove statement (4) one observes that the inclusion $\mathcal{I}_{n,m} \subseteq \mathcal{I}_{n,m+1}$ is an immediate consequence of the definition. Moreover, with L_n as above, one has $L_{n-1} \in \mathcal{I}_{n,m+1} \setminus \mathcal{I}_{n,m}$.

For (5), the inclusion $\mathcal{I}_{n+1,\infty} \subseteq \mathcal{I}_{n,\infty}$ is an immediate consequence of the definition. To prove the inequality, we give examples of languages $M_n \in \mathcal{I}_{n,\infty} \setminus \mathcal{I}_{n+1,\infty}$. Moreover, statements (6) and (8) together will imply the inequality for $n = 1$. Hence, here we consider the case of $n \geq 2$ only. Define M_n as

$$\{w_i w_{i+1} \mid i = 1, 2, \ldots, n\} \cup \{w_{n+1} w_1\}.$$

One verifies that M_n has the required properties.

For (6), from [Ito2] one has $\mathcal{Q} \subseteq \mathcal{C}_2$. If a, b are distinct elements of X, then $\{a, b, ab\} \in \mathcal{Q} \setminus \mathcal{C}_n$ for $n > 2$. Therefore, we only need to prove that $\mathcal{I}_{1,m} = \mathcal{Q}$ for all m, or equivalently, that

$$w^{m+1} \notin X^+ w^m X^+ \iff w \in Q$$

holds true for all m. If $w^{m+1} \in X^+w^mX^+$ then $w = xy = yx$ for some $x, y \in X^+$, that is, $w \notin Q$. Conversely, if $w \notin Q$, that is, $w = x^i$ for some $x \in X^+$ and $i > 1$ then

$$w^{m+1} = x^{i(m+1)} \in X^+w^mX^+$$

for $m \geq 1$.

Statement (7) is proved in [Shy7]. Finally, in (8) the inclusion $\mathcal{I}_{2,\infty} \subseteq \mathcal{Q} \cap \mathcal{C}_b$ is obvious by (1) and (6). On the other hand, the language $\{a, bab\}$ is a bifix code and in $\mathcal{Q}$, but not in $\mathcal{I}_{2,\infty}$. □

Corollary 5.36 *Let $k, n \in \mathbb{N}$ and $m \in \mathbb{N} \cup \{\infty\}$ with $n \leq 2m + 1$. If $k < n$ and $\omega \subseteq [X^+]^k$ then $\mathcal{I}_{n,m} \neq \mathcal{L}_\omega$. Moreover, if ω is any finitary relation on X^+ then $\mathcal{I} \neq \mathcal{L}_\omega$.*

Proof: One applies Theorem 5.27. □

Note that the inclusion $\mathcal{I}_{n,m} \subseteq \mathcal{I}_{n+1,m+1}$ which Figure 3 in Section 1.1 seems to suggest is not true in general. For instance, the language

$$\{abcd, bc, dabcdebf, eb, fbce\}$$

is in $\mathcal{I}_{4,3}$, but not in $\mathcal{I}_{5,4}$. This example uses an alphabet with six letters. However, using the encoding of the letters $a, b, \ldots, e$ by $aba, a^2b^2a^2, \ldots, a^6b^6a^6$ one obtains a language in $\mathcal{I}_{4,3} \setminus \mathcal{I}_{5,4}$ which uses only two letters.

The hierarchy of n-ps-languages is obtained in a slightly different fashion [Ito3]. Recall that a language is called an n-ps-language if every subset of at most n elements is a prefix code or a suffix code. The hierarchy of n-ps-codes is known to collapse at $n = 4$. For $n \in \mathbb{N}$, $n \geq 2$, define the n-ary relation $\omega_{(n,p,s)}$ on X^+ by

$$x \in \omega_{(n,p,s)} \iff \text{cont}\, x \notin \mathcal{C}_p \cup \mathcal{C}_s$$

for $x \in [X^+]^n$ where $\mathcal{C}_p$ and $\mathcal{C}_s$ are the classes of prefix and of suffix codes, respectively. Then $\mathcal{L}_{\omega_{(n,p,s)}} = \mathcal{PS}_n$ where, as before, $\mathcal{PS}_n$ denotes the family of n-ps-codes.

We also consider the ternary relation $\omega_{(g3,p,s)}$ defined as the upward symmetric closure of the set of triples $x = (x_1, x_2, x_3) \in [X^+]^3$ satisfying the following three conditions:

(1) $|\text{cont}\, x| > 1$.

(2) If $|\text{cont}\, x| = 3$ then $x_1 \leq_p x_2 \wedge x_1 \leq_s x_3$.

(3) If $|\text{cont}\, x| = 2$ then $x_1 = x_2 \wedge x_1 \leq_p x_3 \wedge x_1 \leq_s x_3$.

The $\omega_{(g3,p,s)}$-independent languages are the ***generalized* 3-*ps-codes*** of [It3]. Their family is denoted by $\mathcal{GPS}$.

Theorem 5.37 *The following properties obtain for n-ps-codes:*

(1) The hierarchy of n-ps-codes is finite, that is,

$$\mathcal{C}_p \cup \mathcal{C}_s = \ldots = \mathcal{PS}_5 = \mathcal{PS}_4 \subsetneq \mathcal{PS}_3 \subsetneq \mathcal{GPS} \subsetneq \mathcal{PS}_2 = \mathcal{L}_{<_d} \subsetneq \mathcal{C}_2 = \mathcal{L}_{\omega_C}.$$

(2) There is no binary relation ω such that $\mathcal{PS}_n = \mathcal{L}_\omega$ for $n = 3, 4$.

(3) There is no binary relation ω such that $\mathcal{GPS} = \mathcal{L}_\omega$.

(4) There is no ternary relation ω such that $\mathcal{PS}_4 = \mathcal{L}_\omega$.

Proof: Statement (1) is proved in [Ito3]. Statements (2)–(4) are consequences of Theorem 5.27. □

CHAPTER 6

Closing Remarks

The aim of this thesis is to develop a universal mechanism which provides a tool for constructing codes and extracting common properties of codes. The earlier work was mainly based on strict binary relations: several classes of codes were expressed as the independent sets with respect to certain strict binary relations. For convenience, we call this technique the strict-binary mechanism. The binary relations $\leq_{pi}$ and ω_{sh_n} studied in this thesis serve to classify the classes of p-infix codes and n-shuffle codes, respectively. The constructions of solid codes and intercodes indicate that certain natural classes of codes cannot be expressed as the independent sets with respect to strict binary relations. Hence, it is necessary to modify the strict-binary mechanism.

By providing an appropriate definition, we generalize the strict-binary mechanism to an n-ary mechanism and derive several useful results. It should be noted that th original definitions and results for binary relations can be reformulated and preserved in the new framework.

The two most important results are the gap theorem and the inclusion theorem. The gap theorem provides a powerful tool to determine the impossibility of characterizing certain classes of languages by n-ary relations. Using the gap theorem, we can give a very easy proof of that there is no n-ary relation such that the independent sets with respect to this n-ary relation is exactly the class of all codes. The inclusion theorem provides a tool to exhibit hierarchies of classes of codes by means of hierarchies of finitary relations.

Besides the investigation of the mechanism to classify classes of codes and n-codes, we also construct two new classes of codes, the solid codes and the intercodes. Both are related to the class of comma-free codes. It must be mentioned that comma-free codes have a very good decoding property, that is, their synchronous decoding delay is equal to 1. This property is very useful in the design of circuits of coders and decoders. The solid codes form a proper subfamily of the class of comma-free codes and have some interesting additional properties with respect to decoding. The decipherability of coded messages is further studied in [Yu2].

The intercodes are a generalization of comma-free codes with the feature of being synchronously decipherable, limited and circular. The property that a language is an intercode of index m if and only if it is a $2m+1$-intercode of index m provides a strong connection between codes and n-codes. This result also provides a useful example for the generalization of the binary-relation mechanism.

According to matroid theory, relations and their independent sets are related to some kind of dependence systems. However, the connection is not as obvious as it looks at first glance. Moreover, the studies of this thesis indicate that the hierarchy constructions of n-codes and n-ps-codes are just special cases of a general technique for constructing hierarchies. These two issues are discussed in [Jür6].

The studies of this thesis provide a uniform and powerful mechanism to characterize certain classes of languages and to extract the general properties of relations and languages, and also contribute considerably to the understanding of the crucial structural properties of the relationship between classes of codes and n-codes.

REFERENCES

[Ber1] Berlekamp, E.R. *Algebraic coding theory*, McGraw-Hill, New York, 1968.

[Ber2] Berstel, J. and Perrin D. *Theory of Codes*, Academic Press, Inc., Orlando; Toronto, 1985.

[Cha1] Cha, Y.B. and Shyr, H.J. *Some Algebraic Properties of Prefix Codes*, Nanta Mathematica, Singapore, **VI** (1973), no.2, 60–64.

[Cri1] Crik, H.C., Griffith, J.S. and Orgel, L.E. *Codes without Commas*, Proc. Nat. Acad. Sci., **43** (1957), 416–421.

[Day1] Day, P.H. and Shyr, H.J. *Languages defined by some partial orders*, Soochow Journal of Mathematics, **9** (1983), 53–62.

[Gin1] Ginsburg, S. *Algebraic and Automata-Theoretic Properties of Formal Languages*, Fundamental Studies in Computer Science, **2**, North-Holland Publishing Company, Amsterdam, Oxford, 1975.

[Gol1] Golomb, S.W. and Gordon, B. *Codes with Bounded Synchronization Delay*, Information and Control, **8** (1965), 355–372.

[Gol2] Golomb, S.W., Gordon, B. and Welch, L.R. *Comma-Free Codes*, Canadian Journal of Mathematics, **10** (1958), 202–209.

[Grä1] Grätzer, G. *Universal Algebra*, D. van Nostrand Co., Princeton, 1968.

[Guo1] Guo, Y.Q., Shyr, H.J. and Thierrin, G. *QF-Disjunctive Languages*, Papers on Automata and Languages, **VII** (1985), 1–28.

[Hai1] Haines, L.H. *On Free Monoids Partially Ordered by Embedding*, J. Combinatorial Theory, **6** (1969), 94–98.

[Hig1] Higman, G. *Ordering by divisibility in abstract algebras*, Proc. London Math. Soc. **3** (1952), no.2, 326–336.

[Hop1] Hopcroft, J.E. and Ullman, J.D. *Formal Languages and Their Relation to Automata*, Addison-Wesley, Massachusetts, 1976.

[Hsi1] Hsieh, C.Y., Hsu, S.C. and Shyr, H.J. *Some Algebraic Properties of Comma-Free Codes,* RIMS Kenkyuroku, **697**, Japan, (1989), 57–66.

[Ito1] Ito, M. *Strongly Infix Codes,* (in preparation).

[Ito2] Ito, M., Jürgensen, H., Shyr, H.J. and Thierrin, G. *Anti-Commutative Languages and n-Codes,* Discrete Applied Mathematics, **24** (1989), 187–196.

[Ito3] Ito, M., Jürgensen, H., Shyr, H.J. and Thierrin, G. *n-Prefix-Suffix Languages,* International Journal of Computer Mathematics, **30** (1989), 37–56.

[Ito4] Ito, M., Jürgensen, H., Shyr, H.J. and Thierrin, G. *Outfix and Infix Codes and Related Classes of Languages,* International Journal of Computer Mathematics, to appear.

[Jig1] Jiggs, B.H. *Recent Results in Comma-Free Codes*, Canadian Journal of Mathematics, **15** (1963), 178–187.

[Jul1] Jullien, M.P. *Sur un Théorème d'extension dans la Théorie des Mots* C. R. Acad. Sci. Paris Sér., **A266** (1968), 651–654.

[Jür1] Jürgensen, H., Thierrin, G. and Yu, S.S. *r-Shifting Languages,* in preparation.

[Jür2] Jürgensen, H. and Yu, S.S. *Relations on Free Monoids, Their Independent Sets, and Codes*, Technical Report, Department of Computer Science, The University of Western Ontario, **257** (1989).

[Jür3] Jürgensen, H. and Yu, S.S. *Solid Codes,* EIK, to appear.

[Jür4] Jürgensen, H. and Thierrin, G. *Infix Codes*, Topics in the Theoretical Bases and Applications of Computer Science, Proc. Fourth. Hungarian Comp. Sci. Conf., Györ, 1985, edited by M. Arató, I. Kátai, L. Varga. Akadémiai Kiadó, Budapest, 1986, 25–29.

[Jür5] Jürgensen, H., Shyr, H.J. and Thierrin, G. *Codes and Compatible Partial Orders,* S. Wolfenstein (ed.), Algebra and Order, Proc. 1st Internat. Sympos. on Order Algebraic Structures, Heldermann-Verlag, Berlin, (1986), 323–333.

[Jür6] Jürgensen, H. and Yu, S.S. ***Dependence Systems and Hierarchies of Families of Languages***, in preparation.

[Lal1] Lallement, G. *Semigroups and Combinatorial Applications*, John Wiley and Sons, New York, 1979.

[Lyn1] Lyndon, R.C. and Schützenberger, M.P. ***On the Equation $a^M = b^N c^P$ in a Free Group***, Michigan Math. J. **9** (1962), 289–298.

[Per1] Perrin, D. and Perrot, J.F. ***Sur les Codes Préfixes Complets Finis***, C. R. Acad. Sci., Paris, **269** (1969), 1116–1118.

[Pet1] Petrich, M. and Thierrin, G. ***The Syntactic Monoid of an Infix Code***, to appear.

[Rei1] Reis, C.M. ***F-Disjunctive Congruences and a Generalization of Monoids with Length***, Semigroup Forum, **41** (1990), 291–306.

[Sal1] Salomaa, A. *Jewels of Formal Language Theory*, Rockville Md.: Computer Science Press, 1981.

[Sch1] Schein, B.M. ***Homomorphisms of Subdirect Decompositions of Semigroups***, Pacific Journal of Mathematics, **17** (1966), 529–547.

[Sch2] Schützenberger, M.P. ***On an Application of Semigroup Methods to Some Problems in Coding***, IRE Trans. Information Theory, IT-2 (1956), 47–60.

[Sch3] Schützenberger, M.P. ***Sur une question concernant certains sour-monoides libres***, C. R. Acad. Sci. Paris, **261** (1965), 2419–2420.

[Sch4] Schützenberger, M.P. ***On the Synchronizing Properties of Certain Prefix Codes***, Information and Control, **7** (1964), 23–36.

[Sch5] Schützenberger, M.P. ***On a Family of Submonoids***, Publ. Math. Inst. Hunger. Acad. Sci. Ser. *A* **VI** (1961), 381–391.

[Shy1] Shyr, H.J. *Free Monoids and Languages*, Lecture Notes, Department of Mathematics, Soochow University, Taipei, Taiwan, 1979.

[Shy2] Shyr, H.J. ***Ordered Catenation and Regular Free Disjunctive Languages,*** Information and Control, **46** (1980), no.3, 257–269.

[Shy3] Shyr, H.J. ***Some Free Submonoids of the Free Monoid of Prefix Codes,*** Semigroup Forum, **11** (1975) 22–29.

[Shy4] Shyr, H.J. and Thierrin, G. ***Codes and Binary Relations,*** Lecture Notes in Mathematics, Springer-Verlag, Séminare d'Algèbre, Paul Dubreil, Paris **586** (1975-1976), 180–188.

[Shy5] Shyr, H.J. and Thierrin, G. ***Hypercodes,*** Information and Control, **24** (1974), $N^0$1, 45–54.

[Shy6] Shyr, H.J. and Yu, S.S. ***Solid Codes and Disjunctive Domains,*** Semigroup Forum, **41** (1990), no.1, 23–37.

[Shy7] Shyr, H.J. and Yu, S.S. ***Intercodes and Some Related Properties,*** Soochow Journal of Mathematics, **16** (1990), no.1, 95–107.

[Szi1] Szilard, A.L. *On Closure Properties of Language Operations*, Thesis (M.A.), University of Western Ontario, 1968.

[Thi1] Thierrin, G. ***Sur la Structure des Demigroupes***, Alger Math. **3** (1956), 161–171.

[Thi2] Thierrin, G. ***The Syntactic Monoid of a Hypercode***, Semigroup Forum, **6** (1973), 227–231.

[Thi3] Thierrin, G. and Yu, S.S. ***Shuffle Relations and Codes,*** submitted for publication.

[Val1] Valkema, Erich ***Syntaktische Monoide und Hypercodes,*** Semigroup Forum, **13** (1976/77), no.2, 119–126.

[Woo1] Wood, D. *Theory of Computation*, Harper & Row, Publishers, Inc. New York, 1987.

[Yu1] Yu, S.S. ***A Characterization of Intercodes,*** International Journal of Computer Mathematics, **36**, (1990), 39–45.

[Yu2] Yu, S.S. ***The Decomposition of Infinite Coded Message***, Manuscript, 1990.