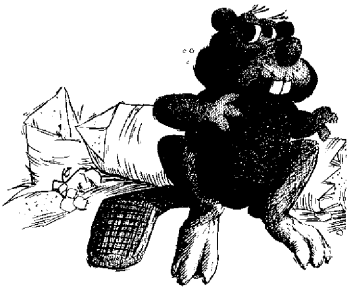
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT



Ambiguity and Decision Problems Concerning Number Systems

Karel Culik II

Arto Salomaa

CS-82-14

May, 1982

UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT

AMBIGUITY AND DECISION PROBLEMS CONCERNING NUMBER SYSTEMS*

Karel Culik II

Arto Salomaa†

Department of Computer Science

University of Waterloo

Waterloo, Ontario, Canada N2L 3G1

ABSTRACT

The representation of integers in arbitrary number systems is considered. The main emphasis is on problems concerning ambiguity, completeness and equivalence. We develop a rather general automata-theoretic method for solving such, in essence, purely number-theoretic problems. The method seems to be applicable in a variety of different situations.

1. INTRODUCTION

Recent work in the theory of codes (see, for instance Maurer et al. (1982)), as well as in cryptography, has led to problems dealing with the representation of positive integers in arbitrary number systems. Here "arbitrary" means that the digits may be larger than the base and that some integers may have several

* This work was supported by Natural Sciences and Engineering Research Council of Canada, Grant Nos. A7403 and A1617.

† On a leave of absence from the University of Turku, Turku, Finland.

representations or none at all. Typical questions arising are: Do the sets of numbers represented by two given number systems coincide? Is the representation of numbers according to a given number system ambiguous or unambiguous?

Very little is known about the solution of such problems in spite of their fundamental number-theoretic nature and also in spite of the fact that the representation of integers is fundamental also in the theory of computing. Moreover, such problems seem to be closely connected with the theory of arithmetical codes, in particular with the work of P. Elias (see Jelinek (1968), also Jürgensen and Kunze (1980).) Unfortunately there fails to be a general framework or theory for dealing with such problems although there are some scattered results such as the one by Honkala (1982).

The purpose of the present paper is to lay the foundations for such a theory by discussing the basic notions in a systematic way and introducing a technique for solving decision problems. It is interesting to note that this technique is based on results in automata theory, and we do not know any other way of solving the purely number-theoretic problems we are dealing with! Of course, the constructions can be "translated" into a language which does not use automata theory but they may then become very complicated.

A brief outline of the contents of the paper follows. Section 2

introduces the basic notions and gives also some examples used in the later theoretical discussions. After some preliminary lemmas given in Section 3, we establish our basic tool (the "translation lemma") in Section 4. The translation lemma has several direct corollaries and some of its modifications yield the main decidability results obtained in Sections 5 and 6. Section 5 gives a solution for the equivalence problem and also discusses some problems dealing with the characterization of representable sets. Section 6 deals with ambiguity and inherent ambiguity. For instance another proof for the result due to Honkala (1982) is obtained by our translation lemma. Finally, Section 7 makes an excursion into L systems, pointing out how our results can be interpreted in terms of L systems.

The paper is self-contained: only the basics about regular languages and gsm mappings, Salomaa (1973), are required on the part of the reader. Maurer et al. (1982) have more motivation for the study of number systems, and Rozenberg and Salomaa (1980) information about the notions and results referred to in Section 7.

2. DEFINITIONS AND EXAMPLES

We begin by defining the fundamental notions of this paper.

A *number system* is a $(v+1)$-tuple

$$N = (n, m_1, \ldots, m_v)$$

of positive integers such that $v \geq 1$, $n \geq 2$ and $1 \leq m_1 < m_2 < \cdots < m_v$. The number n is referred to as the *base* and the numbers m_i as *digits*.

A nonempty word

$$m_{i_k} m_{i_{k-1}} \cdots m_{i_1} m_{i_0} \quad , \quad 1 \leq i_j \leq v \ ,$$

over the alphabet $\{m_1,\ldots,m_v\}$ is said to *represent* the integer

$$[m_{i_k} \cdots m_{i_0}] = m_{i_0} + m_{i_1} \cdot n + m_{i_2} \cdot n^2 + \cdots + m_{i_k} \cdot n^k \ .$$

The set of all represented integers is denoted by $S(N)$. A set of positive integers is said to be *representable by a number system*, shortly *RNS*, if it equals the set $S(N)$, *for some number system* N.

Two number systems N_1 and N_2 are called *equivalent* if $S(N_1) = S(N_2)$. A number system N is called *complete* if $S(N)$ equals the set of all positive integers. It is called *almost complete* if there are only finitely many positive integers not belonging to $S(N)$.

A number system N is termed *ambiguous* if there are two distinct words w_1 and w_2 over the alphabet $\{m_1,\ldots,m_v\}$ representing the same integer: $[w_1] = [w_2]$. Otherwise, N is termed *unambiguous*.

An *RNS* set is termed *unambiguous* if it equals $S(N)$, for some unambiguous number system N. Otherwise, it is termed *inherently ambiguous*. (Thus, an *RNS* set S being inherently

ambiguous means that whenever $S = S(N)$ then N is ambiguous.)

Example 2.1: For each $n \geq 2$, the number system

$$N = (n, 1, 2, \ldots, n)$$

is complete and unambiguous. Consequently, for different values of n we get equivalent systems. Representation according to N is customarily referred to as the *n-adic* representation of integers.

Remark Many of our results remain valid also in the case where zero and even negative numbers are allowed to be among the digits. The presence of 0, however, immediately induces ambiguity and, furthermore, the applications seem to motivate the definition given above. The reader is referred to Salomaa (1973) for a discussion about the differences between n-adic and n-ary (i.e., including 0 among the digits) representations.

Example 2.2: Consider the number system $N = (2,2,3,4)$. We claim that $S(N)$ consists of all positive integers that are not of the form $2^k - 3$, for some $k = 2,3,4, \cdots$. Thus, 1, 5, 13, 29, 61 are the first few numbers missed.

In fact, no number of the form $2^k - 3$ can be in $S(N)$ because if $x = 2^k - 3$ is the smallest such number in $S(N)$, we consider the representation $[a_1 \cdots a_m] = x$. Here obviously $m \geq 2$ and $a_m = 3$ (because otherwise the represented number

is even). But now $[a_1 \cdots a_{m-1}] = 2^{k-1}-3$, contradicting the choice of x.

On the other hand, for any $k \geq 1$, an arbitrary integer x satisfying $2^{k+1}-2 \leq x \leq 2^{k+2}-4$ is represented by some word of length k over the digit alphabet. This can be easily established by induction on k. Observe that

$$[2^k] = 2^{k+1}-2 \quad \text{and} \quad [4^k] = 2^{k+2}-4.$$

Hence, our claim concerning $S(N)$ follows. Note also that N is ambiguous, 8 being the smallest number with two representations. We'll see in Section 6 that $S(N)$ is, in fact, inherently ambiguous. In the "dyadic" number system $N_2 = (2,1,2)$, $S(N)$ is represented by all words over $\{1,2\}$ that are not of the form 2^i1, for some $i \geq 0$. Thus, a regular expression can be given for the set of words representing $S(N)$ in dyadic notation. An analogous result for an arbitrary base will be shown in Section 4.

Example 2.3: The number system $N = (2,1,4)$ is unambiguous. This is easy to verify directly.

We claim that $S(N)$ equals the set of numbers incongruent to 2 modulo 3. We show first that all numbers in $S(N)$ are of this type. This is clearly true of numbers represented by words of length 1 over the digit alphabet. On the other hand, whenever x is congruent to 0 (resp. 1) modulo 3, then both $2x+1$ and $2x+4$ are congruent to 1 (resp. 0) modulo 3. Hence, induction on the

length of the representing words shows that every number in $S(N)$ is incongruent to 2 modulo 3.

Conversely, to show that all such numbers are in $S(N)$, we assume the contrary. Let x be the smallest number incongruent to 2 (mod 3) which is not in $S(N)$. Hence, $x = 3k$ or $x = 3k+1$, for some k. Assume first that $x = 3k$. Then if k is odd (resp. even), the last digit in the representation of x must be 1 (resp. 4) and the number $(3k-1)/2 = 3(k-1)/2 + 1$ (resp. $(3k-4)/2$) is congruent to 1 modulo 3 and not in $S(N)$ (because, otherwise, x would be in $S(N)$). This contradicts the choice of x. Assume, secondly, that $x = 3k+1$. A similar contradiction now arises by considering the number $(x-1)/2$ or $(x-4)/2$, depending whether k is even or odd.

Observe, finally, that in unary notation the set $S(N)$ of Example 2.2 is non-regular, whereas the set $S(N)$ of Example 2.3 is regular.

Example 2.4: This example is a more general one. Consider, for $k \geq 3$, the number system $N(k) = (2,2,k)$. When is $N(k)$ unambiguous? Clearly, $N(k)$ is unambiguous if k is odd. Thus, assume that $k = 2m$. It is easy to see that if m is even, then $N(k)$ is unambiguous. The first odd values of m yielding an unambiguous $N(k)$ are: 11, 19, 23, 27, 35, 37, 39, 43, 45, 47, 51, 53, 55, 59, 67, 69, 71, 75, 77, 79, 83, 87, 89, 91, 93, 95, 99. The reader is referred to Maurer et al. (1982) for more information as regards

this example.

3. PRELIMINARY LEMMAS

This section contains lemmas dealing with the transition from one base to another, ambiguity, and the construction of some classes of RNS sets. Note that all our constructions are effective, although this is not explicitly stated.

Lemma 3.1: For every number system $N = (n, m_1, \ldots, m_v)$ and every integer $k \geq 2$, there is a number system $N_1 = (n^k, \alpha_1, \ldots, \alpha_t)$ equivalent to N. Moreover, if N is unambiguous then so is N_1.

Proof. We choofe the digits α to consist of all integers of the form

$$m_{i_1} + m_{i_2} n + \cdots + m_{i_k} n^{k-1}, \quad 1 \leq i_j \leq v . \tag{1}$$

Then clearly $S(N_1) = S(N)$. The second sentence follows because an ambiguity according to N_1 can be converted into an ambiguity according to N: if there are two words representing the same number according to N_1, there are also two words representing the same number according to N.

■

Applying Lemma 3.1 to Example 2.2, we see that the same set $S(N)$ is represented also by the number system (4, 2, 3, 4, 6, 7, 8,

9, 10, 11, 12). A more detailed analysis shows that actually the system (4, 2, 3, 4, 6, 7, 8, 9) is sufficient for this purpose. This is basically due to the ambiguity of the system N. Observe, however, that in the general case all of the digits given by (1) are needed.

The following lemma is from Maurer et al. (1982) but it is also easy enough to establish directly.

Lemma 3.2: A number system $N = (n, m_1, \ldots, m_v)$ is ambiguous if $v > n$. N is unambiguous if the digits m_j lie in different residue classes modulo n.

Lemma 3.3: No finite set is RNS, whereas every cofinite set is RNS.

Proof: Our definition of a number system guarantees that there is always a base $n \geq 2$ and at least one digit. Hence, the first part of the assertion is obvious. To establish the second part, consider an arbitrary cofinite set S. Without loss of generality, we assume that the complement of S is nonempty. Let k be the greatest number in the complement.

We now define a number system N as follows. The base n equals $k+1$. The digits consist of all numbers $< n$ that are in S, as well as of all numbers i such that

$$n \leq i \leq n^2+n-1 \ . \qquad (2)$$

(Obviously all numbers given by (2) are in S.) Then clearly $S(N) \subseteq S$. To establish the reverse inclusion, it suffices to show that every number $\geq n$ is in $S(N)$. But this follows by an easy induction, using the digits (2).

∎

The range of i in (2) is the best possible in the general case. For instance, consider the set S missing only the number 1. Then the base of N will be 2, and Example 2.2 shows that we need all of the digits given by (2): The digits 2, 3, 4 are insufficient.

Lemma 3.4: Let $n \geq 2$ be arbitrary. Then every (nonempty) union of some residue classes modulo n is RNS. Consequently, both odd and even numbers form an RNS set.

Proof: Given n and some (at least one!) residue classes modulo n, we define a number system N as follows. The base equals n. The set of digits consists of all integers that are in one of the given residue classes and are also $\leq n^2-1$.

Clearly, every number in $S(N)$ is in one of the given residue classes. We still have to prove that every number in the given residue classes is in $S(N)$. Assume the contrary, and let $q = kn+i$ be the smallest number that is in one of the given residue classes but not in $S(N)$. Then our choice of the digits guarantees that $k \geq n$. Hence, there exists a number i_1, with the properties

$$0 \leq i_1 \leq n-1 \ , \quad k-i_1 > 0 \ , \quad k-i_1 \equiv i \pmod{n} \ .$$

We now write q in the form

$$q = (k-i_1)n + (i_1 n + i) \ ,$$

where $i_1 n + i \leq n^2-1$ and, thus, $i_1 n + i$ is among our digits. Now $k-i_1$ cannot belong to $S(N)$ because, otherwise, q would be in $S(N)$. Since $k-i_1 \equiv i(\mathrm{n})$ and, hence, $k-i_1$ is in one of the given residue classes, this contradicts the choice of q.

∎

The proof of Lemma 3.4 shows that the set $S(N)$ of Example 2.3 is represented also by the number system (3, 1, 3, 4, 6, 7). In some sense, this number system is more "natural" for the set consisting of all integers incongruent to 2 modulo 3. However, it is ambiguous, whereas the set of Example 2.3 is unambiguous!

It is easy to see that the bound n^2-1 given in the proof of Lemma 3.4 is the best possible in the general case.

4. TRANSLATION LEMMA AND COROLLARIES

We shall now introduce the method which will be basic for our decidability results. It consists of representing the sets $S(N)$ as regular languages. The following result is referred to as the "translation lemma".

Lemma 4.1: For every number system $N = (n, m_1, \ldots, m_v)$, one can construct a regular expression $\rho(N)$ over the alphabet

$\{1,\ldots,n\}$ such that the set of words in the language denoted by $p(N)$, when these words are viewed as n-adic numbers, equals the set $S(N)$.

Proof: We construct a generalized sequential machine M translating words over the alphabet $\{m_1,\ldots,m_v\}$ into equivalent (i.e., representing the same number "over" the base n) words over the alphabet $\{1,\ldots,n\}$. The construction is based on the fact that the "carry" will always be bounded in such computations. The input and output format of M will be explained below.

The state set of M consists of the states $q_0,q_1,\ldots,q_{2t}$, where $t = \max(n,m_v)$, and of a special final state q. The input alphabet is $\{m_1,\ldots,m_v,\#\}$, and the output alphabet $\{1,\ldots,n\}$. q_0 is the initial state and q the only final state. The behavior of M is specified as follows. Intuitively, being in the state q_i means that there is a carry i in the computation so far.

Thus, when reading the letter j in the state q_i, M produces the output letter j' and goes to the state $q_{i'}$, where i' and j' are unique integers satisfying

$$i + j = j' + i'n \quad , \quad 1 \leq j' \leq n \quad .$$

It is easy to verify inductively that, in this procedure, i' never becomes greater than $2t$, so M has the required state $q_{i'}$. Finally, when reading the letter $\#$ in the state q_i, M produces the output i in reverse n-adic notation (i.e., the rightmost digit

represents the highest power) and goes to the state q. Thus, proper translations are obtained only for words from $\{m_1,\ldots,m_v\}^*\#$. Moreover, M translates such words, viewed as numbers represented according to N in reverse notation and provided with the boundary marker $\#$, into words over $\{1,\ldots,n\}$, representing the same number in reverse n-adic notation. Consequently, the mirror image of the language $M(\{m_1,\ldots,m_v\}^+\#)$ represents the set $S(N)$ in n-adic notation, and clearly a regular expression $\rho(N)$ can be constructed as required.

As an example, consider the number system $N = (2,a,b)$ where $a = 13$ and $b = 22$. The computation of M for the input $aba\#$ is given below:

STATE:	q_0	q_6	q_{13}	q_{12}
INPUT:	a	b	a	$\#$
OUTPUT:	1	2	2	212
NEW STATE:	q_6	q_{13}	q_{12}	

Clearly, the dyadic word 212221 represents the same number (109) as the word aba (which happens to be its own mirror image) according to N. Of course, the reason for the reverse n-adic notation and mirror images in the proof of Lemma 4.1 is merely the fact that we have followed the customary operational mode of gsm's : inputs are read from left to right. On the other hand, in number system notation, the digits representing the highest numbers are customarily on the left.

Observe also that the sequential machine M in the proof of Lemma 4.1 is deterministic. It is also almost a Mealy machine: the only time it may produce more than one output letter (or none at all) is the end of the computation when it scans $\#$.

The following two results are now immediate corollaries of Lemma 4.1.

Theorem 4.2: It is decidable whether or not a given number system is almost complete.

Proof: Given N, we have to find out whether $S(N)$ is cofinite. By Lemma 4.1, this amounts to deciding the cofiniteness of a regular language.

∎

Deciding the completeness of a giving number system is trivial: a number system is complete if and only if every number less than or equal to the base is among the digits.

Theorem 4.3: The equivalence problem is decidable for number systems with the same base.

Proof: By Lemma 4.1, since the n-adic representation is unambiguous, we only have to decide the equivalence of two regular expressions.

∎

By Lemma 3.1, Theorem 4.3 can be extended to concern the case where the bases of the number systems considered are powers of the same number. However, in the next section we shall establish decidability in general.

We mention, finally, that the translation lemma can be extended to the case where the digits are arbitrary integers: a gsm mapping from the number system notation to the n-adic notation can be constructed also in this case. Hence, Theorems 4.2 and 4.3 remain valid in this more general set-up.

5. EQUIVALENCE AND CHARACTERIZATION

We shall establish in this section the decidability of the equivalence of two number systems in the general case. We shall also consider some problems dealing with the characterization of RNS sets. We begin with a lemma useful in many considerations involving number systems. The lemma resembles some fixed-point results in language theory.

Lemma 5.1: Let $n \geq 2$ and $1 \leq m_1 < \cdots < m_v$ $(v \geq 1)$ be given integers. Consider the number system $N = (n, m_1, \ldots, m_v)$. Then the set $X = S(N)$ satisfies the equation

$$X = \{nx + m_j \mid x \in X, 1 \leq j \leq v\} \cup \{m_1, \ldots, m_v\} \qquad (3)$$

and, moreover, $S(N)$ is the only set of positive integers satisfying (3).

Proof: Clearly $S(N)$ is a solution of (3): the first term on the right side represents the operation of adding one digit to the right. To show that the solution of (3) is unique, we let X be an arbitrary solution. Then clearly m_j is in X, for $1 \leqq j \leqq v$. Thus, all one-digit numbers (we are considering the representation according to N) are in X. We make the following inductive hypothesis: all k-digit numbers are in X. But now the first term of the union shows that all $(k+1)$-digit numbers are in X. Consequently, $S(N)$ is included in X.

To establish the reverse inclusion $X \subseteq S(N)$, we assume the contrary, and let x be the smallest number in $X - S(N)$. Since x is in X, it must belong to one of the terms of the union on the right side of (3). Because x is not in $S(N)$, it cannot be in $\{m_1,\ldots,m_v\}$. Consequently, there are numbers $x_1 \in X$ and j such that $x = nx_1 + m_j$. Here x_1 cannot belong to $S(N)$ because, otherwise, x is in $S(N)$, a contradiction. Consequently, x_1 is in $X - S(N)$. Because clearly $x_1 < x$, this contradicts the choice of x. This implies that $X = S(N)$.

■

We introduce the following notation in order to avoid confusion and cumbersome terminology. Consider a word w over the alphabet $\{1,\ldots,n\}$. We denote by $\nu(w)$ the integer denoted by w in n-adic notation. Thus, if $n = 2$ then $\nu(212221) = 109$.

Lemma 5.2: Let A be a regular language over the alphabet $\{1,\ldots,n\}$, $n \geq 2$, and let a and b be positive integers. Then the language

$$L(A,a,b) = \{ w \in \{1,\ldots,n\}^+ \mid \nu(w) = a \cdot \nu(x) + b \text{ ,for some } x \text{ in } A \}$$

is also regular.

Proof: The proof is analogous to that of Lemma 4.1. We construct a gsm M that receives as its inputs words x over the alphabet $\{1,\ldots,n\}$, multiplies the word (viewed in the reverse n-adic notation) by a and adds b to the result. The output is produced, letter by letter, also in reverse n-adic notation. The carry will be bounded and can, thus, be included in the states of M. Our lemma now follows because gsm mappings preserve regularity.

■

Theorem 5.3: The equation $S(N) = \{\nu(w) \mid w \in A\}$ is decidable for a given regular language A over the alphabet $\{1,\ldots,m\}$, $m \geq 2$, and for a given number system $N = (n,m_1,\ldots,m_\nu)$.

Proof: By Lemma 5.1, $S(N) = \{\nu(w) \mid w \in A\}$ holds if and only if

$$A = \bigcup_{1 \leq i \leq \nu} L(A,n,m_i) \cup \{\nu^{-1}(m_1),\ldots,\nu^{-1}(m_\nu)\} \qquad (4)$$

(Clearly, if A and B are regular languages over the alphabet $\{1,\ldots,m\}$, then $A = B$ holds if and only if

$\{\nu(w) \mid w \in A\} = \{\nu(w) \mid w \in B\}$ holds. This follows by the unambiguity of the m-adic representation.)

By Lemma 5.2, the equation (4) is decidable.

■

Theorem 5.4: It is decidable whether or not two given number systems are equivalent.

Proof: The theorem is a direct consequence of Lemma 4.1 and Theorem 5.3.

■

Lemma 4.1 associates a regular expression to each *RNS* set. Conversely, not every regular expression gives rise (in this sense) to an *RNS* set (see, for instance, Lemma 3.3). The following *characterization problem* is open: characterize the regular expressions giving rise to *RNS* sets. Indeed, we do not even know the solution for the following decision problem: decide of a regular expression whether or not it gives rise to an *RNS* set.

On the other hand, the following result is immediate by Rice's theorem (see Rogers (1967)), since the property of being an *RNS* set, or an *RNS* set to a given base, is a nontrivial one.

Theorem 5.5: It is undecidable whether or not a given recursively enumerable set is *RNS*. It is also undecidable, given a recursively enumerable set S and an integer $m \geq 2$, whether

or not there is a number system $N = (n, m_1, \ldots, m_v)$ such that $S = S(N)$.

We mention, finally, that the decidability of the following problem is open: given a number system N and an integer $m \geq 2$, is there a number system N_1 with base m such that $S(N) = S(N_1)$? A solution for the case where N_1 is assumed to be unambiguous is given in the next section.

6. AMBIGUITY AND INHERENT AMBIGUITY

We begin this section by giving a different proof to the result established by Honkala (1982) - the proof given by Honkala does not use automata theory.

Theorem 6.1: It is decidable whether or not a given number system $N = (n, m_1, \ldots, m_v)$ is ambiguous.

Proof: Consider the generalized sequential machine M constructed in the proof of Lemma 4.1. Clearly, N is unambiguous if and only if M gives rise to an injective gsm mapping.

Whether or not M gives rise to an injective gsm mapping can be tested by the following procedure. Let u be the square of the number of states in M. We claim that if the gsm mapping is not injective then M maps two such words $w_1\#$ and $w_2\#$ into the same word such that at least one of w_1 and w_2 is of length

$\leq u$. Clearly, this condition is testable.

To establish our claim, we let $w_1\#$ and $w_2\#$ be shortest (in the sense that the sum of their lengths is smallest) words mapped into the same word by M. Let $p_1,p_2,\cdots$ (resp. $q_1,q_2,\cdots$) be the sequence of states entered by M when reading letter by letter the word w_1 (resp. w_2). (Recall that M is a Mealy machine when reading w_1 and w_2.) If both w_1 and w_2 are of length $> u$, then there are i and j, $i < j$, such that $(p_i, q_i) = (p_j, q_j)$. But this means that we can remove from w_1 and w_2 every letter between and including the $(i+1)$st and jth letter, and the resulting words $w'_1\#$ and $w'_2\#$ are still mapped into the same word by M. This, however, contradicts the choice of the pair (w_1, w_2). Hence, our claim follows.

∎

The following theorem gives still another proof for the decidability of the ambiguity of a given number system. We use somewhat stronger tools than in the proof of Theorem 6.1.

Theorem 6.2: Let $N = (n, m_1, \ldots, m_v)$ be a number system and A the corresponding regular language according to Lemma 4.1. Let $L(A,a,b)$ be the regular language considered in Lemma 5.2. Then N is unambiguous if and only if, for all distinct i and j,

$$L(A,n,m_i) \cap L(A,n,m_j) = \phi .$$

Proof: The intersection being nonempty means that there two words, one ending with m_i and the other with m_j, over the digit alphabet representing the same number according to N. Hence, the intersection being empty is a necessary and sufficient condition for the unambiguity.

∎

An analogous argument can be used also to solve a problem related to the one mentioned at the end of Section 5.

Theorem 6.3: It is decidable of a given number system N and an integer $m \geq 2$ whether or not there exists an unambiguous number system N_1 with base m satisfying $S(N) = S(N_1)$.

Proof: Let n be the base of N and let A be the regular language representing $S(N)$ in n-adic notation according to Lemma 4.1. Clearly, we can effectively list A in an order which is strictly increasing with respect to the mapping ν. Let $x_1, x_2, \cdots$ be this listing.

We now test whether such a system N_1 as required exists. This is done by a method of successive approximations. Obviously $\nu(x_1)$ must be one of the digits. Consider the number system $K_1 = (m, \nu(x_1))$. We may test by Theorem 5.4 whether or not $S(N) = S(K_1)$. If the answer is "yes", we have found N_1 as required. Otherwise, we find the first element x_j in the x_i-sequence which is not "covered" by K_1 in the sense that $\nu(x_j)$ is

not in $S(K_1)$. (This is clearly effective. If there are no such elements, which can also be found out effectively, we terminate with a negative answer.) If a system N_1 as required exists, $\nu(x_j)$ must be among the digits.

We denote $K_2 = (m, \nu(x_1), \nu(x_j))$ and repeat the above considerations for K_2. If this does not lead to termination, we consider the resulting system K_3, and so forth. If we have not terminated after considering K_m, we may terminate with a negative answer, by Lemma 3.2. Observe that, because of the unambiguity of N_1, numbers "covered" by previous systems are not candidates for digits.

■

The technique of the previous proof, i.e. testing the listing of A step by step, seems to be applicable in a number of situations.

The decidability of the inherent ambiguity of a given *RNS* set is open. By Theorem 6.3, it suffices to find an effective bound for the base.

One can also introduce for number systems in a natural fashion (as for context-free grammars, see Salomaa (1973)), the notions of the degree of ambiguity and the degree of inherent ambiguity. The resulting existence and decision problems are all open.

In the remainder of this section, we consider some special cases of ambiguity and inherent ambiguity.

Theorem 6.4: The set of even numbers is unambiguous. The set of odd numbers is inherently ambiguous.

Proof: Consider the number system $N = (2,2,4)$. Clearly, for any $m \geq 1$, there is a one-to-one correspondence between dyadic representations of m and representations of $2m$ according to N. (The former are obtained from the latter through division by 2.) Since the dyadic representation is unambiguous and complete, the first sentence follows. (In fact, one can prove in the same way that, for every $k \geq 1$, the set

$$\{ik \mid i \geq 1\}$$

is unambiguous.)

To prove the second sentence, consider any number system N representing the set of odd numbers. Clearly, all the digits must be odd numbers and, hence, the base n must be even. This implies that all odd numbers i such that $1 \leq i \leq n-1$ must be among the digits. On the other hand if, for some odd i with the property $1 \leq i \leq n-1$, neither $n+i$ nor $2n+i$ are among the digits, then $2n+i$ is not in $S(N)$. (This follows because the numbers n^2 and jn with $j \geq 3$ are too large to take part in a representation of $2n+i$.) But if $n+i$ is among the digits, then $n+i$ itself has two representations. Consequently, we may assume that, for every odd i with the property $1 \leq i \leq n-1$, $2n+i$ is among the digits. But now, for $n \geq 4$, the equation

$$3 \cdot n+1 = n+(2n+1)$$

shows that $3n+1$ has two repreentations.

Consider, finally, the case $n = 2$. We have shown that 1 and 5 must be among the digits. But now the number 7 has two representations: $2+5 = 2^2+2+1$. Hence, Theorem 6.4 follows by Lemma 3.4.

■

We now return to Example 2.2 and show that the set $S(N)$ is inherently ambiguous. Consider any number system N_1 with base n, representing $S(N)$. Since the number 1 is not of the proper form, it cannot appear as a digit. Consequently, all numbers of the proper form among the numbers $1,\ldots,2n+1$ must appear as digits. (Observe that $2n+2$ is the smallest number we can get using the first power of n.) We now choose i in such a way that $2 \leqq i \leqq 3$ and $n+i$ is of the proper form. (This is possible because no two consecutive numbers are of the improper form.) Since 2 and 3 are among the digits, the equation

$$2n+(n+i) = 3n+i$$

shows that N_1 is ambiguous. Hence, $S(N)$ is inherently ambiguous. An alternative argument can be based on Lemma 3.2. One can also show that, in any number system N_1 representing the set $S(N)$, the base must be a power of 2.

As indicated already in Example 2.2, a regular expression can be given for the set $S(N)$. This regular expression corresponds to

the n-adic notation, for a suitably chosen n. However, in unary notation, $S(N)$ is nonregular. In fact, it is an open problem to decide of a given N whether $S(N)$ is regular or nonregular in unary notation.

The last result in this section shows that in some cases ambiguity depends essentially on the base.

Theorem 6.5: There is an *RNS* set S possessing representations with different bases m and n such that it has an unambiguous representation with the base m, whereas every representation of S with base n is ambiguous.

Proof: The set $S = S(N)$ in Example 2.3 satisfies the required conditions. Lemma 3.2 shows that the number system (2, 1, 4) is unambiguous. Hence, we may choose $m = 2$. Lemma 3.4 shows that $S = S(N_1)$ where $N_1 = (3,1,3,4,6,7)$. To complete the proof, we show that every representation of S with the base 3 is ambiguous. Indeed, 1 and 3 must be digits because, otherwise, they are not represented. Also 4 must be among the digits because, otherwise, 7 is not represented. But 4 being a digit implies that it has two representations, which shows ambiguity.

■

7. APPLICATIONS TO L SYSTEMS

For a reader interested in L systems, we now show how our results can be "translated" into results concerning L systems. The interconnection with L codes was investigated by Maurer et al. (1982). However, there are also some more direct interconnections. The reader is referred to Rozenberg and Salomaa (1980) for all unexplained notions.

Lemma 7.1: For every number system $N = (n, m_1, \ldots, m_v)$, there is a OL system G with the alphabet $\{S, a\}$ such that

$$L(G) = \{S\} \cup \{Sa^i \mid i \in N(S)\} . \tag{5}$$

Proof: The axiom of G is S, and the productions are:

$$a \to a^n , \quad S \to Sa^{m_i} \quad \text{for } i = 1, \ldots, v .$$

■

Lemma 7.2: For every number system N, there is a $DTOL$ system G such that (5) holds.

Proof: The system G of Lemma 7.1 can immediately be transformed into a $DTOL$ system by pairing each of the S-productions with the a-production.

■

A 2×2 matrix representation, analogous to growth function representations, is also possible.

By Lemma 7.1 and 7.2, our decidability and ambiguity results can be expressed in terms of the corresponding L systems. For instance, the equivalence problem is decidable for OL systems of the form of Lemma 7.1. It is well-known that the equivalence problem is undecidable for OL systems in general, whereas it is decidable for unary OL systems. The systems of Lemma 7.1 are slightly different from the unary systems.

We want to emphasize that the decidability results obtained from the results of this paper by Lemmas 7.1 and 7.2 are new in the theory of L systems. Indeed, in spite of the direct interconnection, we could not use the theory of L systems to settle the decision problems considered in this paper.

8. CONCLUSION

We have studied in a systematic way representability, ambiguity and decision problems dealing with number systems. Some important questions still remain open, as pointed out above. Another open problem area is to develop some useful "normal forms" for number systems. Also open is to what extent our results carry over to number systems having arbitrary integers as digits. A rather surprising fact is that our main results are based entirely on automata theory. This can be viewed as another indication of the diverse applicability of automata theory! It remains to be seen whether these results can be obtained also by

purely number-theoretic arguments.

REFERENCES

Honkala, J. (1982) Unique representation in number systems and L codes. *Discrete Applied Mathematics*, 4, 229-232.

Jelinek, F. (1968) *Probabilistic Information Theory*. McGraw-Hill, New York.

Jürgensen, H. and Kunze, M. (1980) Redundanzfreie Codes als Kryptocodes. Technical Report TI 8/80, Darmstadt Technical University.

Maurer, H. A., Salomaa, A. and Wood, D. (1982) L codes and number systems. *Theoretical Computer Science*, to appear.

Rogers, Hartley Jr. (1967) *Theory of Recursive Functions and Effective Computability*. McGraw-Hill, New York.

Rozenberg, G. and Salomaa, A. (1980) *The Mathematical Theory of L Systems*. Academic Press, New York.

Salomaa, A. (1973) *Formal Languages*. Academic Press, New York.