Discrete Applied Mathematics 32 (1991) 183–193
North-Holland

Power languages and density

Valtteri Niemi

Mathematics Department, University of Turku, SF-20500 Turku, Finland

Received 10 April 1989
Revised 16 October 1989

Abstract

Niemi, V., Power languages and density, Discrete Applied Mathematics 32 (1991) 183–193.

The class of all languages can be seen as a distributive lattice with respect to a preorder defined by letter-to-letter morphisms. Maximal dense intervals in the lattice are investigated. The results are based on a construction that builds a new language, so-called power language, from subsets of a given language. Applications to grammar form theory and graph theory are also presented.

1. Introduction

The notion of a morphism plays a central role in the theory of languages. In form theory, our attention is focused especially on *length preserving*, i.e., *letter-to-letter* morphisms. We can compare two languages by asking whether the first one can be mapped into the other by a letter-to-letter morphism. Essentially, this relation defines a *preorder* in the whole class of languages. In fact, we obtain a well-organized structure: the class of languages is a *distributive lattice* with respect to the preorder.

The structure in the lattice of languages is very rich: for instance, it contains *dense intervals*. Two characterizations for such intervals were given in [1] and [3]. In this paper we study the possibility of enlarging an interval while retaining density. The major problem is whether there are exact limits for such extension processes. In other words, we seek *maximal* dense intervals.

It turns out that there always exists an exact bound for all extensions directing upwards. It means that every dense interval is contained in another one which is maximal from above. Unfortunately, there are no respective lower bounds in general. In [3] it is shown that a dense interval is maximal from below iff its lower limit is a so-called *nonlooping* language. Therefore, the existence problem of maximal density reduces to search of a dense interval reaching down to some nonlooping language. All nonlooping languages are finite and quite "low" in the lattice order but, however, such intervals do exist. An explicit example is given in [4].

Our upper bound deals with an analogue of the power set construction. Let us consider a language L as a set. Then the power set of L can also be interpreted as a language, *if* all words in L are of equal length. Namely, the alphabet of the new language is the power set of the old alphabet and, for instance, the first letter in a word corresponding to a subset of L is simply the set of first letters in that subset. An interesting feature in this interpretation is the fact that the mapping from subsets to words is not necessarily injective; several subsets of L may correspond to the same word.

To generalize this construction to the case where the language L contains words of distinct length we qualify only those subsets that consist of uniform length words. Again, we interpret each uniform length subset as a word over the power set alphabet. In this way we define, for every language L, its *power language* $L+$.

Our main result implies that, if a dense interval contains a given language L, we may extend the interval to reach up to $L+$ but not higher without losing density.

The ordering of languages is totally independent of grammatical hierarchies. However, when grammars are considered as finite languages the preorder relation may be applied to them, leading us to *grammar form theory*. Our results have several consequences on grammar forms. The existence problem of maximal dense intervals of grammar forms (see [2]) is settled. Moreover, we prove that it is *decidable* whether or not two given context-free grammars G_1 and G_2 make up a maximal dense interval $(\mathscr{L}(G_1), \mathscr{L}(G_2))$.

Another application is on *graph theory*. Directed graphs may be interpreted as languages consisting solely of two-letter words. Then the preorder relation between languages corresponds to general coloring of a digraph by another one. Our results suit directly to this case.

The paper is structured as follows. Some preliminaries are presented in Section 2 while Section 3 contains the main theorem with its proof. The existence of maximal dense intervals is shown in Section 4 and consequences in grammar forms and digraphs are presented in Sections 5 and 6. For the end we briefly discuss some open questions in Section 7.

2. Preliminaries

Throughout the paper we use the following convention, customary in form theory. (The empty word is denoted by λ.)

The λ-convention: Given two languages L_1 and L_2 we say that they are *equal* (modulo λ) if $L_1 - \{\lambda\} = L_2 - \{\lambda\}$. Similarly, we say two language families $\mathscr{L}_1$ and $\mathscr{L}_2$ are *equal* (modulo λ and $\emptyset$) if for every $L_1 - \{\lambda\} \neq \emptyset$ in $\mathscr{L}_1$ there is an L_2 in $\mathscr{L}_2$ such that $L_1 - \{\lambda\} = L_2 - \{\lambda\}$ and vice versa.

Essentially this means that we ignore the empty set in language families and the empty word in languages. Consult [3] for justification of this convention.

A language form is defined as follows. Consider an arbitrary language $L \subseteq \Sigma^*$

over a finite alphabet Σ. A language L' over a finite alphabet Σ' is an *interpretation* of L, in symbols $L' \leq L$, if there exists a letter-to-letter morphism $h : \Sigma' \to \Sigma$ such that

$$h(L') \subseteq L.$$

The morphism h is called an *interpretation morphism*. Throughout the paper all morphisms considered are assumed to be letter-to-letter.

The *linguistical family* of a language form L is defined by

$$\mathscr{L}(L) = \{L' \mid L' \leq L\}.$$

Two languages are *equivalent*, denoted by $L_1 \sim L_2$, if $\mathscr{L}(L_1) = \mathscr{L}(L_2)$.

Clearly the relation $\leq$ is reflexive and transitive, hence $\mathscr{L}(L_1) \subseteq \mathscr{L}(L_2)$ iff $L_1 \leq L_2$. Consequently, $L_1 \sim L_2$ iff $L_1 \leq L_2$ and $L_2 \leq L_1$. If $L_1 \leq L_2$ but not $L_2 \leq L_1$, then we say L_1 is a *proper* interpretation of L_2, written $L_1 < L_2$. A language form L is *minimal* if there is no language form $L' \subset L$ such that $L' \sim L$.

We say that (L_1, L_2) denotes an *interval*, if $L_1 < L_2$, and hence $\mathscr{L}(L_1) \subset \mathscr{L}(L_2)$. The language L_1 (respectively, L_2) is called the *lower* (respectively, *upper*) *limit* of the interval. The interval is *dense*, if for all languages L_3 and L_4 such that $L_1 \leq L_3 < L_4 \leq L_2$ there exists a language L_5 with $L_3 < L_5 < L_4$.

A dense interval (L_1, L_2) is *maximal*, if there are no languages L_3 and L_4 such that

(i) $L_3 \leq L_1 < L_2 \leq L_4$,
(ii) $L_3 < L_1$ or $L_2 < L_4$ (or both),
(iii) (L_3, L_4) is dense.

In the sequel we denote the phrase "*maximal dense interval*" shortly by "*MDI*".

A language L is *looping*, if either L contains a word with at least two occurrences of the same letter or there exist distinct words $w_1, \ldots, w_m$ in L ($m \geq 2$) and distinct letters $a_1, \ldots, a_m$ in alph(L) such that a_i and a_{i+1} occur in w_i, $1 \leq i \leq m-1$, while a_m and a_1 occur in w_m. If L is not looping we say it is *nonlooping*. If L is not equivalent to any nonlooping language it is said to be *inherently looping*.

Let us denote the family of all nonlooping languages by $\mathscr{L}$(NL). We say that two languages L_1 and L_2 are *nonlooping equivalent*, in symbols, $L_1 \sim_N L_2$, if $\mathscr{L}(L_1) \cap \mathscr{L}(\mathrm{NL}) = \mathscr{L}(L_2) \cap \mathscr{L}(\mathrm{NL})$. Similarly, we write $L_1 \leq_N L_2$ to mean $\mathscr{L}(L_1) \cap \mathscr{L}(\mathrm{NL}) \subseteq \mathscr{L}(L_2) \cap \mathscr{L}(\mathrm{NL})$. Now $L_1 \sim_N L_2$ iff $L_1 \leq_N L_2$ and $L_2 \leq_N L_1$.

As regards definition of a grammar form, we refer to [6].

Let G_1 and G_2 be two grammar forms such that $\mathscr{L}(G_1) \subset \mathscr{L}(G_2)$. Then they form an *interval* $(\mathscr{L}(G_1), \mathscr{L}(G_2))$. The interval is *dense* if for any two families $\mathscr{L}(G_3)$ and $\mathscr{L}(G_4)$ in the interval the following implication is valid:

if $\mathscr{L}(G_3) \subset \mathscr{L}(G_4)$ *then* there exists a grammar form G_5 such that

$$\mathscr{L}(G_3) \subset \mathscr{L}(G_5) \subset \mathscr{L}(G_4).$$

The concept of *maximal denseness* is defined for *grammatical families* (i.e.

language families defined by grammar forms) analogously as for linguistical families above.

Next we recall some theorems of [3].

Proposition 2.1. *Given two languages L_1 and L_2 with $L_1 < L_2$, the interval (L_1, L_2) is dense iff $L_1 \sim_N L_2$. (This result is originally from* [1].)

Proposition 2.2. *Let (L_1, L_2) be a dense interval with L_1 inherently looping. Then the interval is not maximal dense.*

Proposition 2.3. *The collection of all linguistical families*

$$\mathfrak{L} = \{\mathscr{L}(L) \mid L \text{ is a language}\}$$

is a distributive lattice with respect to the containment.

According to Proposition 2.1, every dense interval (L_1, L_2) is a convex subset of some equivalence class $[L]$ defined by the equivalence relation $\sim_N$. The whole equivalence class $[L]$ is always a sublattice of $\mathfrak{L}$, as easily seen by [3, Theorem 2.9]. Consider now an arbitrary MDI within an equivalence class $[L]$. Its upper (respectively, lower) limit must be a maximal (respectively, minimal) element in $[L]$. Since $[L]$ is a lattice, a maximal (respectively, minimal) element is necessarily greatest (respectively, smallest) element of $[L]$. This means, in particular, that every language belongs to at most one MDI.

3. Main theorem

In this section we first give the detailed construction of the *power language* $L+$. Then we present and prove the main theorem of the paper.

We use the following notations. Let L be a language over a (finite) alphabet Σ, and $k \geq 1$. Let us denote

$$L(k) = _{\text{df}} L \cap \Sigma^k = \{w \in L \mid |w| = k\}.$$

Let then w be a word and $i \geq 1$. We denote by $w_{(i)}$ the ith letter in the word w. The latter denotation is also generalized to cover the case of *languages*:

$$L_{(i)} = _{\text{df}} \{w_{(i)} \mid w \in L\} \ (\subseteq \Sigma).$$

Now we are ready to begin the construction of $L+$. The first task is to define the alphabet used (we denote it by $\Sigma+$). The cardinality of $\Sigma+$ is equal to the number of nonempty subsets of Σ. In fact, we could use the subsets themselves as elements of $\Sigma+$, but for the sake of clearness we only index the elements by subsets. Thus, letters in the words of $L+$ are of the form $\alpha(\Sigma')$ where $\emptyset \neq \Sigma' \subseteq \Sigma$. For instance, we have letters $\alpha(a, b, c)$, $\alpha(a, b)$, $\alpha(a, c)$, $\alpha(b, c)$, $\alpha(a)$, $\alpha(b)$ and $\alpha(c)$ in case $\Sigma = \{a, b, c\}$.

Next we construct words of the sublanguage $L+(k)$ for each $k \geq 1$. The language $L+$ is determined completely, since $L+ = \bigcup_{k \geq 1} L+(k)$. Consider the sublanguage $L(k)$ of L. Similarly as in the construction of the alphabet, we first index the words of $L+(k)$ by nonempty subsets of $L(k)$. Thus, words in the language $L+(k)$ are of the form $\omega(W)$ where $\emptyset \neq W \subseteq L(k)$. For instance, we might have words $\omega(abc, bba, ccc)$, $\omega(aa)$, $\omega(a, c)$ etc. in $L+$.

The letters of the words $\omega(W)$ are obtained by the following rule:
(Let $\emptyset \neq W \subseteq L(k)$ and $1 \leq i \leq k$.)

$$\omega(W)_{(i)} = \alpha(W_{(i)}). \tag{R}$$

In other words, the ith letter in the word indexed by words $w_1, \ldots, w_n$ is itself indexed by the set of ith letters in the same words.

It is worth noting that the function ω is not necessarily injective, while the function α is injective *per definitionem*.

The construction is now completed. It is easy to see that, if L is finite and effectively constructable, then $L+$ is an effectively constructable (finite) language. Moreover, if L is recursive (respectively, recursively enumerable), then $L+$ is also recursive (respectively, recursively enumerable).

We may now establish our main result.

Theorem 3.1. *Let L be an arbitrary language and $L+$ its power language. Then*
(A) *L can be embedded in $L+$, i.e. there is a language L_0 such that $L \cong L_0 \subseteq L+$;*
(B) *$L+ \sim_N L$;*
(C) *for every language L' the following holds: if $L' \leq_N L$ then $L' \leq L+$.*

Proof. We must show that the power language $L+$ satisfies the conditions (A)–(C).

(A) We see easily that the language L is isomorphic to the sublanguage of $L+$ indexed by singleton sets. More accurately: Let $a \in \Sigma$. Consider the morphism h obtained by restricting the function α to singleton sets, i.e. $h : \Sigma \to \Sigma+$, $a \mapsto \alpha(a)$. Let now $w \in L$. The ith letter of the word $h(w)$ in $(\Sigma+)^*$ is $\alpha(w_{(i)})$. On the other hand, by the rule (R), the ith letter of the word $\omega(w)$ is also $\alpha(w_{(i)})$. Since $|h(w)| = |w| = |\omega(w)|$, we have $h(w) = \omega(w)$. The morphism h is clearly injective (since α is), and we have

$$L \cong h(L) = \{\omega(w) \mid w \in L\} \subseteq L+.$$

(B) We begin by observing that the relation $L \leq_N L+$ follows directly from the condition (A). Thus, we have to prove the reverse relation. For that purpose we recall Theorem 3.4 from [3]. The theorem is crucial in our subsequent reasoning, and we have slightly strengthened it from [3]. However, the stronger version is quite easily deduced from the original one.

Proposition 3.2. *Let L_1 and L_2 be arbitrary languages. Denote* $\mathrm{alph}(L_i) = \Sigma_i$ *for*

$i=1,2$. Then $L_1 \leq_N L_2$ iff there exists a finite-letter substitution $\delta: \Sigma_1 \to \Sigma_2$ such that the following two conditions hold:

(i) *$\delta(a)$ is nonempty for every $a \in \Sigma_1$.*

(ii) *The inclusion*

$$\delta(a) \subseteq \{b \in \Sigma_2 \mid (\forall x_1 \dots x_{i-1} a x_{i+1} \dots x_t \in L_1) \\ (\exists y_1 \dots y_{i-1} b y_{i+1} \dots y_t \in L_2)\,(\forall k = 1, \dots, t)\, y_k \in \delta(x_k)\}$$

is valid.

(Here $x_i = a$ and $y_i = b$. Moreover, $x_k \in \Sigma_1$ and $y_k \in \Sigma_2$ for $k = 1, \dots, t$, and it is assumed that $1 \leq i \leq t$. In case $i = 1$ we make the convention that the notation $x_1 \dots x_{i-1}$ means the empty word. Similarly, $x_{i+1} \dots x_t$ is empty, if $i = t$.)

Proof of Theorem 3.1 (continued). Based on Proposition 3.2, we show that $L+ \leq_N L$ by defining a finite-letter substitution $\delta: \Sigma+ \to \Sigma$ and verifying that δ fulfils conditions (i) and (ii). The substitution δ is defined by

$$\delta: \alpha(\Sigma') \mapsto \Sigma' \quad \text{for each } \emptyset \neq \Sigma' \subseteq \Sigma.$$

The condition (i) is clear, since Σ' is always nonempty. To prove that also (ii) holds let us assume that $\alpha(\Sigma')$ is an arbitrary element of $\Sigma+$. Furthermore, let us choose an arbitrary element $b \in \delta(\alpha(\Sigma')) = \Sigma'$.

We have to show that the quantified sentence in the set definition of (ii) is true for $a = \alpha(\Sigma')$ and b. Hence, assume $x_1 \dots x_{i-1} \alpha(\Sigma') x_{i+1} \dots x_t \in L+$. This means that $x_1 \dots x_{i-1} \alpha(\Sigma') x_{i+1} \dots x_t = \omega(W)$ for some $W \subseteq L(t)$. Furthermore, $\omega(W)_{(i)} = \alpha(\Sigma')$.

On the other hand, the rule (R) implies that $\omega(W)_{(i)} = \alpha(W_{(i)})$. Hence, $\Sigma' = W_{(i)}$ and, in particular, $b \in W_{(i)}$. It follows that there exists a word $w \in W \subseteq L(t)$ such that $b = w_{(i)}$, in other words, $w = y_1 \dots y_{i-1} b y_{i+1} \dots y_t \in L$.

We still have to demonstrate that $y_k \in \delta(x_k)$ for each $k = 1, \dots, t$. Thus, let k be fixed. Clearly, $x_k = \omega(W)_{(k)}$. On the other hand, by (R), $\omega(W)_{(k)} = \alpha(W_{(k)})$. Hence, $x_k = \alpha(W_{(k)})$, from which it follows that $\delta(x_k) = \delta(\alpha(W_{(k)})) = W_{(k)}$, by the definition of δ. Since $w \in W$ and $y_k = w_{(k)}$, we may conclude that $y_k \in W_{(k)} = \delta(x_k)$.

(C) Let L' be a language such that $L' \leq_N L$. Then it follows from Proposition 3.2 that there exists a finite-letter substitution $\delta: \mathrm{alph}(L') \to \Sigma$ that satisfies conditions (i) and (ii).

Let us define a (letter-to-letter) morphism $h: L' \to (\Sigma+)^*$ by

$$h: a \mapsto \alpha(\delta(a)) \quad \text{for each } a \in \mathrm{alph}(L').$$

We have to prove that $h(L') \subseteq L+$. Let us choose an arbitrary $v = a_1 \dots a_t$ from L'. We make use of the following

Claim 3.3. *$\delta(a_i) = (\delta(v) \cap L)_{(i)}$ for each $i = 1, \dots, t$.*

Proof. The inclusion of the latter set in the former one is clear, since $\delta(a_i) =$

$\delta(v)_{(i)}$. To prove the opposite inclusion let us first fix i and then choose an arbitrary $b \in \delta(a_i)$. Now (ii) implies that there exists a word $y_1 \dots y_{i-1} b y_{i+1} \dots y_t \in L$ such that $y_k \in \delta(a_k)$ for each $k = 1, \dots, t$. In other words, $y_1 \dots y_{i-1} b y_{i+1} \dots y_t \in \delta(v) \cap L$.

In particular, $b \in (\delta(v) \cap L)_{(i)}$. □

Proof of Theorem 3.1 (continued). Let us denote $W = \delta(v) \cap L$. It follows from our claim and the condition (i) of Proposition 3.2 that $\emptyset \neq W = \delta(v) \cap L \subseteq L(t)$. Now the claim and the rule (R) together imply that $\alpha(\delta(a_i)) = \alpha(W_{(i)}) = \omega(W)_{(i)}$ for each $i = 1, \dots, t$. Hence, $h(v) = \alpha(\delta(a_1 \dots a_t)) = \alpha(\delta(a_1)) \dots \alpha(\delta(a_t)) = \omega(W)_{(1)} \dots \omega(W)_{(t)} = \omega(W) \in L+$.

Thus, $L' \leq L+$ and our proof is completed. □

We conclude this section by an example of the power language construction.

Example. Let $L = \{a^n b^n \mid n \geq 1\} \cup \{b^n a^n \mid n \geq 1\}$, $\Sigma = \{a, b\}$. Now $\Sigma+ = \{\alpha(a), \alpha(b), \alpha(a, b)\}$. For the sake of (notational) convenience, we omit the α-notation and use brackets instead. Hence, $\Sigma+ = \{[a], [b], [a, b]\}$. For each $n = 1, 2, \dots$, the sublanguage $L+(2n)$ consists of words $\omega(a^n b^n)$, $\omega(b^n a^n)$ and $\omega(a^n b^n, b^n a^n)$. All sets of the form $L+(2n-1)$ are empty. Moreover, $\omega(a^n b^n) = [a]^n [b]^n$, $\omega(b^n a^n) = [b]^n [a]^n$ and $\omega(a^n b^n, b^n a^n) = [a, b]^n [b, a]^n = [a, b]^{2n}$.

Thus, $L+ = \{[a]^n [b]^n \mid n \geq 1\} \cup \{[b]^n [a]^n \mid n \geq 1\} \cup \{[a, b]^{2n} \mid n \geq 1\}$.

We see that, indeed, the condition (A) is satisfied, since L is isomorphic to the union of first two parts of $L+$.

It is easy to see that $L+$ is, in fact, equivalent to the language $\{a^2\}^*$. Therefore, by condition (B), $L \sim_N \{a^2\}^*$. Now it follows from Proposition 2.1 that the interval $(L, \{a^2\}^*)$ is dense. On the other hand, L is inherently looping, hence Proposition 2.2 implies that the interval is not a MDI.

4. Maximal dense intervals do exist

This section shows how Theorem 3.1 can be used to settle the existence of MDI. The following result is crucial for this purpose.

Theorem 4.1. *Let N be an arbitrary nonlooping language.*

- *If $N+ \not\leq N$ (that means $N < N+$) then $(N, N+)$ is a MDI.*
- *If $N+ \leq N$ (that means $N \sim N+$) then N does not belong to any dense interval.*

Proof. Assume firstly that $N < N+$. Then $(N, N+)$ is an interval. The condition (B) implies that $N+ \sim_N N$, hence it follows from Proposition 2.1 that $(N, N+)$ is dense.

Let L' be a language with $L' \sim_N N$ ($\sim_N N+$). Then, in particular, $N \leq L'$. On the

other hand, it follows from the condition (C) that $L' \leq N+$. Thus, $L' \in (N, N+)$ and, by Proposition 2.1, it can be seen that $(N, N+)$ is a MDI.

Assume secondly that $N \sim N+$ and N belongs to a dense interval (L_1, L_2). Hence, Proposition 2.1 implies that $L_1 \sim_N L_2$. Since N is nonlooping and $N \leq L_2$, it follows that $N \leq L_1$. On the other hand, the fact $N \in (L_1, L_2)$ implies that $N \sim_N L_2$, hence $L_2 \leq_N N$. Now it follows from (C) that $L_2 \leq N+$, and by our assumption, $L_2 \leq N+ \leq N$. Thus, $L_2 \leq N \leq L_1$, hence (L_1, L_2) cannot be an interval at all. □

Corollary 4.2. *Every* MDI *can be represented in the form* $(N, N+)$, *where* N *is a nonlooping language. If* N *is demanded to be minimal, then the representation is unique within isomorphism (as easily derived from the fact that two minimal and equivalent language forms are also necessarily isomorphic).*

Given an arbitrarily chosen nonlooping language N, it seems to be very probable that the equivalence relation $N \sim N+$ holds, but there are also exceptions. As already mentioned in the Introduction, one such exception is presented in [4]. In this paper we skip the example and simply state

Theorem 4.3. *There exist MDI's of linguistical families.*

After gaining a positive answer to an existence problem it is natural to consider *decidability* problems. Proposition 2.2 and Theorem 4.1 together give us the following result.

Theorem 4.4. *Given a context-free grammar* G *it is decidable whether or not* $L = L(G)$ *is a lower limit of any MDI.*

Moreover, in the positive case we are able to construct the MDI in question. (Here the definite article "the" is justified by the observations made in the end of Section 2.)

Proof. The decision algorithm runs as follows. First check whether L is inherently looping. This is decidable, since G is context-free. If the answer is negative, construct the power language $L+$. This step is effective, since L is necessarily finite. For the end check whether $L+ \leq L$. If, again, the answer is negative, L is the lower limit of the MDI $(L, L+)$. Otherwise, L is not a lower limit of any MDI. □

5. Applications to grammar forms

To transfer Theorem 4.3 to the case of grammar forms and grammatical families, we first need the following fact from [3].

Proposition 5.1. *Let* $(\mathscr{L}(G_1), \mathscr{L}(G_2))$ *be a dense interval such that the language* $L(G_1)$ *is inherently looping. Then the interval is not maximal dense.*

Proposition 5.1 means that all (so far hypothetical!) MDI's of grammatical families lie on the range of finite languages. More specifically, every grammatical family in a MDI consists solely of finite languages. On the other hand, in the case where only finite languages are involved, linguistical families and grammatical families coincide. In particular, the equality $\mathscr{L}(L(G)) = \mathscr{L}(G)$ holds, whenever $L(G)$ is finite. Moreover, the collection of language families

$$\{\mathscr{L}(G) \mid L(G) \text{ is finite}\} = \{\mathscr{L}(L) \mid L \text{ is finite}\}$$

is a sublattice of the collection $\mathfrak{L}$ referred to in Proposition 2.3.

For these reasons our results for language forms can be carried to cover also the case of grammar forms. In particular, by Theorem 4.3, we obtain

Theorem 5.2. *There exist MDI's of grammatical families.*

From Theorem 4.4 it is possible to derive the following decidability result.

Theorem 5.3. *Given two context-free grammars G_1 and G_2 it is decidable whether or not $(\mathscr{L}(G_1), \mathscr{L}(G_2))$ is a MDI.*

Proof. By the observations made before Theorem 5.2, $(\mathscr{L}(G_1), \mathscr{L}(G_2))$ is a MDI iff $(L(G_1), L(G_2))$ is a MDI. Check first whether $L(G_1)$ is a lower limit of some MDI via the algorithm of Theorem 4.4. If the answer is positive, construct $L(G_1)+$ and check whether relations $L(G_1) < L(G_1)+ \sim L(G_2)$ hold. The last step is effective, since G_2 is context-free. □

6. Applications to digraphs

As well known, graphs and directed graphs can be interpreted as (finite) languages with only two-letter words. The alphabet of the language is equal to the node set of the (di)graph. A word ab in the language corresponds to an arc from the node a to the node b in the digraph. Similarly, an edge between a and b in the graph corresponds to the pair of words $\{ab, ba\}$.

A coloring of a (di)graph by another (di)graph corresponds to an interpretation morphism from a language to another.

An (undirected) graph may be viewed as a special type of a digraph in which all arcs are two-way. On the other hand, for each digraph there exists an underlying graph that is obtained by replacing all arcs by edges.

In the case of graphs the question of density is clear: all intervals are dense (see [5]). Digraph intervals are, instead, not always dense. In fact, a digraph has a *predecessor* (with respect to the interpretation relation as a preorder) iff the underlying graph is a tree. However, it is decidable, due to Proposition 3.2, whether or not a given interval of digraphs is dense.

A new feature is the possibility to apply the power language construction in order to maximize any dense interval upwards. In the opposite direction the situation is different: for instance, the dense interval from the digraph $\{ab, ba\}$ to the digraph $\{aa\}$ cannot be maximized downwards. (It is instructive to notice that this interval contains all nontrivial *undirected* graphs. Hence, $(\{ab, ba\}, \{aa\})$ is the only MDI of graphs.)

On the other hand, there exist dense intervals that are maximal also from below. All what is needed to establish an example of the latter case is a digraph D such that its underlying graph is a tree and it isn't itself equivalent to the digraph $D+$. (Then the interval $(D, D+)$ is a MDI.) An idea for construction of such a digraph is presented in [4], that means there exist also MDI's of digraphs.

7. Open problems

An open problem that is most closely related to our discussions is the following:

> Is it decidable whether a given language L belongs to some MDI?

The difficult case is a finite, inherently looping L. Whether L belongs to some MDI, depends on whether there is some nonlooping language N nonlooping equivalent to L. Both positive and negative answers are possible: An example of negative case is the language $\{aa\}$ which is not nonlooping equivalent to any nonlooping language (as may be proved by Proposition 3.2).

One crucial idea of this paper is the observation that all dense intervals can be maximized upwards. This is true for *linguistical* families in general, but holds provably for *grammatical* families only in the finitary case. If $L(G)$ is infinite, we cannot be sure that there exists a *grammar*, say $G+$, such that $L(G)+ = L(G+)$ and $(\mathscr{L}(G), \mathscr{L}(G+))$ is maximal from above. Therefore, we may ask:

> Are there *grammatical* intervals $(\mathscr{L}(G_1), \mathscr{L}(G_2))$ that cannot be maximized upwards?

Such nonlooping languages that are lower limits of MDI's are quite rare. Thus, it might be possible to find some nice characterization for them.

The notion of a power language is probably worth of interest by itself. Put some restrictions on L and see how they reflect in $L+$.

In addition to the aforementioned problems there are interesting possibilities to generalize the theory. For instance, replace all letter-to-letter morphisms by nonerasing ones in the basic definitions. Do we still obtain a distributive lattice of languages, (maximal) dense intervals, etc.?

References

[1] H.A. Maurer, A. Salomaa, E. Welzl and D. Wood, Denseness, maximality and decidability of grammatical families, Ann. Acad. Sci. Fenn. 11 (1986) 167–178.

[2] H.A. Maurer, A. Salomaa and D. Wood, Dense hierarchies of grammatical families, J. ACM 29 (1982) 118–126.

[3] V. Niemi, Density of grammar forms (Parts I and II), Internat. J. Comput. Math. 20 (1986) 3–21 and 91–114.

[4] V. Niemi, Maximal dense intervals of grammar forms, in: Proceedings ICALP '88, Lecture Notes in Computer Science 317 (Springer, Berlin, 1988) 424–438.

[5] E. Welzl, Color-families are dense, Theoret. Comput. Sci. 17 (1982) 29–42.

[6] D. Wood, Grammar and L Forms: An Introduction, Lecture Notes in Computer Science 91 (Springer, Berlin, 1980).