LOGICAL DEFINABILITY OF SOME RATIONAL TRACE LANGUAGES *

Choffrut Christian and Leucio Guerra

Université Paris 7, LITP

Tour 55-56, 1er étage, 75 251 Paris Cedex

May 10, 2011

Abstract

Trace monoids are obtained from free monoids by defining a subset I of pairs of letters that are allowed to commute. Most of the work of this theory is an attempt to relate the properties of these monoids to the properties of I. Following the work initiated by Büchi we show that when the reflexive closure of I is transitive (the trace monoid is then a free product of free commutative monoids) it is possible to define a second-order logic whose models are the traces viewed as dependence graphs and which characterizes exactly the sets of traces that are rational. This logic essentially utilizes a predicate based on the partial ordering defined by the dependence graph and a predicate related to a restricted use of the comparison of cardinality.

1 Introduction

Given a finite set Σ (an *alphabet*) we denote by Σ^* the free monoid it generates and we call *words* its elements. A relation of *partial commutation* or *independence* is a symmetric and irreflexive relation $I \subseteq \Sigma \times \Sigma$. The *trace monoid* defined by this relation is the monoid $M(\Sigma, I)$ presented by $< \Sigma;\ ab = ba$ for all $(a, b) \in I >$, i.e., the quotient $M(\Sigma, I) = \Sigma^*/\underset{I}{\sim}$, where $\underset{I}{\sim}$ is the congruence over Σ^* generated by the relation $\{ab = ba \mid (a, b) \in I\}$.

Given an arbitrary monoid M, two basic families of subsets of M are usually defined, namely the family Rat$M$ of *rational* subsets obtained from the empty

*this research was supported by the PRC Mathématiques et Informatique

set and the singletons by using the rational operations of set union, product and star only, and the family Rec$M$ of *recognizable* subsets defined as the subsets that are unions of classes of a congruence of finite index on M. It is well-known that Rec$M$ is included in Rat$M$ whenever M is finitely generated. In the case of trace monoids this inclusion is strict if at least two letters commute (e.g., for the presentation $M =< a, b; ab = ba >$ the subset $(ab)^*$ is not recognizable since its inverse image in the free monoid $\{a, b\}^*$ is the so-called Dyck language D_1^* (cf. [**?**], p. 35) which is known not to be rational).

The connection between logic and finite automata is due to Büchi (and later to Elgot) who was concerned with deciding the weak second-order monadic theoriy of successor and who showed how automata naturally come up when defining a normal form for their closed formulas [**?**] and [**?**]). This enabled him to reduce the decidability of the logic to the decidability of the emptiness problem for finite automata. These automata recognized infinite words and were only considered as a tool for solving problems in mathematical logic. The other direction, namely using logic to classify languages of (actually finite) words or equivalently considering words as models of a certain theory, can be credited to McNaughton and Papert with their characterization of the "star-free" events by means of first order logic [**?**]. Further refinements of this class of languages (the "dot-depth" hierarchy) as well as an extension of the notion of "star-freeness" to infinite words followed (cf., e.g., [**?**] and [**?**]).

Perhaps the most natural way of introducing partial commutations in free monoids is obtained by passing from single words to pairs of words. Multiplying pairs componentwise reduces to dealing with direct products of free monoids. If say, the two components are written on disjoint alphabets, this leads to considering that these two alphabets commute. We briefly discuss below the difficulties encountered when trying to characterize the rational subsets of these monoids.

The special case of the subfamily of recognizable subsets of a trace monoid is more amenable than the case of the family of rational subsets, and the characterization can be achieved with no restriction on the relation of partial commutation. More precisely, considering the elements of an arbitrary trace monoid as "dependence graphs" (which in a sense is the extension of the linear structure of words), it has been shown that the family of recognizable subsets is logically characterizable [**?**]. In a sense, our result is a continuation of the latter in that it views traces as graphs, but it is concerned with rational rather than recognizable subsets. However, some extra restrictions are necessary, and we focus our attention on trace monoids where the reflexive closure of the relation I is transitive (cf. Example 2 in section **??**) in which case the monoid $M(\Sigma, I)$ is a free product of free commutative monoids. This class of trace monoids enjoys remarkable properties (cf., e.g., [**?**], [**?**] and [**?**]). The dependence graph of a trace t is a finite acyclic graph labelled by the letters of the alphabet where each node is associated with exactly one occurrence of a letter in t. Roughly speaking the edges between the nodes record the non-commutations between the corresponding occurrences of the letters. We enrich the second-order logic of [**?**]

with a second-order predicate that says, with some restriction, that two subsets have the same cardinality, yielding thus what we call the restricted cardinality logic. Our main result then states the following:

Theorem 1.1 *Let $M(\Sigma, I)$ be a trace monoid such that the reflexive closure of the commutation relation I is transitive. Then a subset R of $M(\Sigma, I)$ is rational if and only if there exists a closed formula of the restricted cardinality logic such that a trace of $M(\Sigma, I)$ belongs to R if and only if it is a model of this formula.*

A related result, with the hypothesis that all letters commute, may be found in the literature. Indeed, using Presburger arithmetic, i.e., first-order logic of addition, Ginsburg and Spanier were able to logically characterize the rational trace languages in the commutative case (cf. [**?**], Theorem 1.3). It is a different approach from ours since traces are identified with n-tuples of integers. The language of traces associated with a formula having free individual variables $x_1, x_2, \ldots, x_n$ is defined as the set of n-tuples for which the formula holds true.

Now we want to argue that our result is the most general possible if we do not restrict the use of connectives. Indeed, when our hypothesis on I is violated, then Σ contains 3 different letters a, b, c where a commutes with b and c and no other commutation between them holds. Then the set of rational subsets is not closed under intersection ($(ab)^*c^* \cap b^*(ac)^*$ is not rational). Worse, when Σ contains 4 different letters a, b, c, d where a and b commute with c and d and no other commutation between them holds, then the emptiness problem for the intersection is not decidable, since the Post Correspondence Problem can be reduced to it (cf., e.g., [**?**], Theorem III 8.4). This does not rule out a possibility for a logical characterization but it is an indication that if such a characterization exists it is probably more involved.

Section **??** presents the preliminaries, i.e., it recalls the basic notions on trace monoids, rational and recognizable subsets, free commutative monoids etc.., and exposes our second-order logic. In section **??** we show how to assign a closed formula of the restricted cardinality logic to each rational language by first considering the special case of the free commutative monoids. The converse, i.e., how to associate a rational language to each closed formula is done in the last section. It uses a structural induction on the formulas and takes crucial advantage of the specific properties of the rational subsets when the reflexive closure of the commutation relation is transitive.

2 Preliminaries

2.1 Trace monoids

The theory of trace monoids has been given much attention in these last ten years as witnessed by the lenghty bibliography of the survey article [**?**]. We content ourselves with recalling what is essential to our purpose.

Given a finite set Σ (an *alphabet*) we denote by Σ^* the free monoid it generates. The elements of Σ^* are called *words* and the *empty* word is denoted by 1. Intuitively a trace is a word that is defined up to the commutations of certain pairs of letters. More precisely, we define a relation of *partial commutation* or *independence* as a symmetric and irreflexive relation $I \subseteq \Sigma \times \Sigma$. The *trace monoid* defined by this relation is the monoid $M(\Sigma, I)$, or simply $M(\Sigma)$ when I is understood, presented by $< \Sigma; ab = ba$ for all $(a, b) \in I >$. Equivalently, it is the quotient $M(\Sigma, I) = \Sigma^*/\underset{I}{\sim}$, where $\underset{I}{\sim}$ is the congruence generated by the relation $\{ab = ba \mid \text{for all } (a, b) \in I\}$. When no confusion may arise, we drop the subscript I and write $\sim$ instead of $\underset{I}{\sim}$. The *empty* trace, denoted by 1, is the class of the empty word.

Example 1. — Consider $\Sigma = \{a, b, c\}$ and assume $I = \{(a, b), (b, a), (c, b), (b, c)\}$. Then $a\underline{ba}cbc \sim aab\underline{cb}c \sim aabbcc$ etc...

In particular when $I = \emptyset$, then we obtain the free monoid and when $I = \Sigma\times\Sigma - \{(a, a) \mid a \in \Sigma\}$ then all letters commute and we get the free commutative monoid generated by Σ, also denoted by $\Sigma^{\oplus}$. We are concerned here with a combination of these two cases where the reflexive closure of the relation I is transitive, i.e., $M(\Sigma, I)$ is a free product of free commutative monoids as formally explained in subsection **??**. Then there exists a natural partition of the alphabet Σ into maximal subalphabets $\Sigma_i, i = 1, \cdots, n$ of pairwise commuting letters. In other words, for a given word in Σ^*, the commutations may only occur inside the maximal factors that are written over some fixed subalphabet Σ_i.

Example 2. — Consider $\Sigma = \{a, b, c, d\}$ and let $I = \{(a, b), (b, a), (c, d), (d, c)\}$. The relation of commutation determines the two subalphabets $\Sigma_1 = \{a, b\}, \Sigma_2 = \{c, d\}$. If we start with the word $w = abadccbabcdd$ then the unique commutations occur inside the maximal factors $w_1 = aba, w_2 = dcc, w_3 = bab, w_4 = cdd$. E.g., $a\underline{ba}dccbabcdd \sim aabdccbab\underline{cd}d \sim aab\underline{dc}cbabdcd \sim aabcdcbabdcd$ etc...

2.2 Rational and recognizable subsets

We refer to the handbook of J. Berstel listed in the reference section and especially to its Chapter III for any further reading on the topic. Here, we recall what is necessary for our purpose.

Given an arbitrary monoid M, two basic families of subsets of M are usually defined, namely the family of *rational* and the family of *recognizable* subsets. The first one, denoted by $\mathrm{Rat}M$, is the least family $\mathcal{F}$ of subsets of M containing the empty set and the singletons and closed under *set union* $X \cup Y$ more usually denoted by $X + Y$, *product* $X.Y$ also denoted by XY and defined as $\{xy \in M \mid x \in X \text{ and } y \in Y\}$ and *star* X^* where X^* stands for the submonoid generated by X :

(i) $\emptyset \in \mathcal{F}$ and for all $m \in M$, we have $\{m\} \in \mathcal{F}$.

(ii) for all $X, Y \in \mathcal{F}$, we have $X + Y, XY, X^* \in \mathcal{F}$.

It is folklore that rational subsets of arbitrary monoids may be defined by means of finite automata. Numerous equivalent definitions may be found in the literature but for our purpose the following will suffice. A *finite automaton* over an arbitrary monoid M is a quadruple $\mathcal{A} = (Q, q_-, Q_+, T)$ where Q is a finite set of *states*, $q_- \in Q$ is the *initial state*, $Q_+ \subseteq Q$ is the set of *final states* and $T \subseteq Q \times M \times Q$ is a finite set of *transitions.* A *path* is a sequence

$$c : q_0 \overset{x_1}{\to} q_1 \overset{x_2}{\to} \cdots q_{n-1} \overset{x_n}{\to} q_n$$

where $n \geq 0$ and $(q_{i-1}, x_i, q_i) \in T$ for $i = 1, \cdots, n$. The path c is *successful* if its first state is the initial state and its last state is a final state: $q_0 = q_-$ and $q_n \in Q_+$. The *label* of the path c is the element $x_1 x_2 \cdots x_n \in M$. By convention, the label is the unit of the monoid whenever $n = 0$. The subset of M *accepted* by the automaton $\mathcal{A}$ is the set of labels of all successful paths. Then a subset $R \subseteq M$ is rational if and only if it is accepted by some finite automaton. Also, it is an easy exercise to verify that if M has no divisor of the unit, i.e., $xy = 1$ implies $x = y = 1$, and if $1 \notin R$ then R can be accepted by a *normalized* automaton in the sense of ([**?**], p. 138), i.e, the set of final states is reduced to a unique final state $q_+ \neq q_-$, there is no incoming transition in q_- and no outgoing transition in q_+. This family enjoys important properties such as being closed under morphisms or more generally under rational substitutions. We recall that given two monoids M and N a *rational substitution* of M into N is a mapping h that associates with every element of M a rational subset of N in such a way that $h(m_1 m_2) = h(m_1)h(m_2)$ holds for all $m_1, m_2 \in M$. Then we have (cf., e.g., [**?**] Proposition I.4.2):

Proposition 2.1 *Let M and N be two monoids and let R be a rational subset of M. If h is a rational substitution of M into N then $h(R) = \bigcup_{m \in R} h(m)$ is a rational subset of N.*

The family Rat$M$ is not necessarily closed under the Boolean operations, but when M is a trace monoids, the closure holds if and only if the reflexive closure of the reflexive closure of the relation of independence is a transitive (cf., e.g., [**?**] and [**?**] Theorem 3.6)

Theorem 2.2 *Let $M(\Sigma, I)$ be a trace monoid. Then the family of all rational subsets of $M(\Sigma, I)$ is closed under set union and complementation if and only if for all $(a, b), (b, c) \in I$, $a \neq c$ implies $(a, c) \in I$.*

The second family of subsets of M, denoted by Rec$M$ contains all subsets that are saturated by a congruence of finite index on M. In other words, $R \in \mathrm{Rec}M$ if and only if there exists a congruence $\sim$ of finite index on M such that: for all $x, y \in M, x \in R$ and $x \sim y$ implies $y \in R$. It is well-known that Rec$M$ is included in Rat$M$ whenever M is finitely generated (cf., [**?**] or [**?**], Proposition III.2.4). In the case of trace monoids this inclusion is strict as soon as some pair of letters commute, e.g., for the presentation $M =< a, b;\ ab = ba >$ the subset $(ab)^*$ is not recognizable since its inverse image in the free monoid $\{a, b\}^*$ is not rational. In [**?**] the family of recognizable trace languages is characterized by means of a second-order logic. Due to the non closure of rational trace languages under intersection there is no hope we can extend this result by just adding some new predicate. Rather, we will restrict ourselves to the case where this family is a Boolean algebra, i.e., where the reflexive closure of the commutation relation is transitive.

2.3 Dependence graphs

The second-order language that we are interested in will be interpreted on traces considered as *dependence graphs*. We thus start with recalling this notion (cf. [**?**]). Let I be a commutation relation on the alphabet Σ, let $w = a_1 a_2 \cdots a_n \in \Sigma^*$ (with $a_i \in \Sigma$ for $i = 1, \cdots, n$) be a word and let t be the trace it represents. The *dependence graph* associated with t is a finite directed acyclic graph whose nodes are labelled by the letters of the alphabet and which intuitively, records the "minimal" precedence between the letters, i.e., those that imply all the other non-commutations by transitivity. Formally, it is a triple $G = (V, E, \lambda)$ where each vertex $v_i \in V$ is identified with the i-th occurrence in w, where the label of v_i is a_i and where (v_i, v_j) is an edge of E if $i < j$, $(a_i, a_j) \notin I$ and for no $i < k < j$, the condition $(a_i, a_k), (a_k, a_j) \notin I$ holds (cf. [**?**], p. 13). It is not difficult to see that this definition up to a renaming of the vertices, only depends on the equivalence class of w, i.e., on the trace t. We also write $v \leq v'$ (resp. and $v < v'$) when there is a path (resp. a non trivial path) from v to v' in the dependence graph. This relation defines a partial ordering on the set of vertices. By a *subgraph* G' of G we mean a triple (V', E', λ') where $V' \subseteq V$, $E' = E \cap V' \times V'$ and λ' is the restriction of λ to V'.

Example 2. (continued). — Consider the word *abadccbabcdd* of example 2 and mark the positions of each occurrence of some letter: $a_1 b_2 a_3 d_4 c_5 c_6 b_7 a_8 b_9 c_{10} d_{11}$ d_{12}. Then the dependence graph, shown in Fig. 1, has 12 vertices and e.g., there is an edge between 4 and 7 because $(d, b) \notin I$ holds and the occurrences c_5 and c_6 inbetween commute with d_4. Also we have the relation $3 < 8$ because there is a non trivial path, i.e., a path of length greater than 0, from the vertex 3 to the vertex 8. The reader may check that all the equivalent words have the same graph up to the renaming of the occurrences.

Fig. 1 The dependence graph associated with the trace of Example 2

2.4 The restricted cardinality logic

The logic that we introduce to characterize the rational trace languages consists of

- a denumerable collection of individual variables $x, y, \cdots$
- a denumerable collection of set variables $X, Y, \cdots$
- a monadic predicate symbol Q_a for each letter $a \in \Sigma$
- a binary relation symbol $\prec$ between individual variables
- a binary relation symbol Card_r of *restricted cardinality* between set variables.

It is to be interpreted in dependence graphs of a fixed trace monoid $M(\Sigma, I)$, individual and set variables refering to the nodes of a given graph $G = (V, E, \lambda)$. The predicate $Q_a(x)$ is interpreted as asserting the property that the node x of G is labelled by a. The relation $x \prec z$ is a refinement of the relation "there exists a path in the dependence graph", i.e., of the relation $<$ defined over the nodes of the dependence graph: $x \prec z$ is true if and only if there exists a path from x to z for which there exists no letter $a \in \Sigma$ such that all the nodes including ends on this path are labelled by a. The relation on the set of nodes thus defined is a partial strict ordering and can be expressed in terms of the relation $\leq$, e.g.:

$$x \prec z \equiv \exists y \quad (x \leq y \wedge y \leq z \wedge \bigvee_{a \neq b} (Q_a(y) \wedge (Q_b(x) \vee Q_b(z)))$$

We set $x \preceq y$ whenever $x \prec y$ or $x = y$. As a special case, if all letters commute, then any two distinct elements of the dependence graph are incomparable with respect to $\prec$. We discuss at the beginning of section **??** the reason why we chose this relation instead of the more usual relation $<$ on the dependence graph. Finally, the binary relation $\mathrm{Card}_r(X, Y)$ is interpreted as follows:

the predicate $\mathrm{Card}_r(X, Y)$ is true if and only if X and Y have the same cardinality and for all $x \in X$ and $y \in Y$, x and y are incomparable with respect to $\prec$.

As a curiosity, let us quote [**?**] where a different restriction of the set cardinality equality is used.

Example 2. (continued) . — The Hasse diagram of the relation $\prec$ associated with the trace of Example 2, is as shown in Fig.2.

Fig. 2 The Hasse diagram of the relation $\prec$

There are five *atomic formulas*:

$$(i)\ x = y \quad (ii)\ x \prec y \quad (iii)\ Q_a(x) \quad (iv)\ \mathrm{Card}_r(X, Y) \quad \text{and} \quad (v)\ x \in X$$

E.g., in the previous example, $2 \prec 9$ and $1 \prec 7$ hold but $1 \prec 3$ does not. Also, $\mathrm{Card}_r(\{4\}, \{6\})$ holds, but $\mathrm{Card}_r(\{1, 3\}, \{2\})$ and $\mathrm{Card}_r(\{4\}, \{9\})$ do not. Finally, we have $Q_c(5)$ is true but $Q_a(6)$ is false. The *formulas* are built up of atomic formulas by the Boolean connectives $\neg$, $\wedge$ and the quantifier $\exists$ for individual and subset variables.

The *restricted cardinality logic*, $\mathcal{RCL}$ for short, is the set of all sentences (i.e., of all the closed formulas, or all the formulas with no free variables). With every sentence ϕ, we associate the set L_ϕ of all traces $t \in M(\Sigma, I)$ that satisfy ϕ:

$$L_\phi = \{t \in M(\Sigma, I) \mid t \models \phi\}$$

Our main result is the following logical characterization of the rational trace languages:

Theorem 2.3 *Let $M(\Sigma, I)$ be a trace monoid such that the commutation relation I is transitive. Then a subset R of $M(\Sigma, I)$ is rational if and only if there exists a closed formula ϕ of the restricted cardinality logic such that $R = L_\phi$ holds.*

Example 3. — Consider $\Sigma = \{a, b, c\}$ where all letters commute. We want to define the subset of all traces of the form $a^n b^n c^n$ for all $n \geq 0$. We use 3 set variables X_a, X_b, X_c standing for the positions of the occurrences of a, b and

c respectively. We need to say (i) that the 3 subsets are disjoint, (ii) that all occurrences belong to the union of these 3 subsets, (iii) that the 3 subsets are of the same cardinality and (iv) that all elements of X_a (resp. X_b, X_c) are labelled by a (resp. b, c). Actually, because of the interpretation of Q_a this last condition implies the first one. This leads to the following sentence that is not strictly speaking consistent with the previous definition because of the use of the standard shorthands, but that can easily be seen to be equivalent:

$$\begin{array}{ll}\exists X_a & \exists X_b \exists X_c \\ & [\forall x(x \in X_a \vee x \in X_b \vee x \in X_c) \wedge (\mathrm{Card}_r(X_a, X_b) \wedge \mathrm{Card}_r(X_b, X_c)) \\ & \wedge \forall x((x \in X_a \rightarrow Q_a(x)) \wedge (x \in X_b \rightarrow Q_b(x)) \wedge (x \in X_c \rightarrow Q_c(x))]\end{array}$$

3 From trace languages to sentences

The purpose of this section is to show how to assign a sentence of the restricted cardinality logic to each rational trace language, provided the relation I is transitive. As we have already mentioned, such trace monoids are obtained from free commutative monoids via the operation of free product, so we briefly recall this classical notion.

3.1 Free product of monoids

Given n disjoint monoids $M_i, i = 1 \cdots, n$ and $e_i, i = 1 \cdots, n$ their units, the free product of the M_i's, denoted by $M_1 * \cdots * M_n$, consists, as a set, of a new element e not belonging to $\bigcup_{1 \leq i \leq n} M_i$ and of all finite sequences $x_1, x_2, \cdots, x_p$ of non unit elements alternatively belonging to different monoids:

> if $p = 0$ then the sequence reduces to e, otherwise for all $i = 1, \cdots, p-1$, we have $x_i \in M_j - \{e_j\}$ and $x_{i+1} \in M_k - \{e_k\}$, for some $1 \leq j, k \leq n, j \neq k$

In order to provide this set with a binary operation, let $x = x_1, x_2, ..., x_p$ and $y = y_1, y_2, ..., y_q$ be two elements. The element e is defined as the unit so we may assume that $p > 0$ and $q > 0$ holds. The product $x.y$ is recursively defined as follows:
(1)

$$x.y = \begin{cases} \text{(i) } x_1, x_2, \ldots, x_p, y_1, y_2, \ldots, y_q \text{ if } x_p \text{ and } y_1 \text{ belong to different monoids} \\ \text{(ii) } x_1, x_2, \ldots, x_p y_1, y_2, \ldots, y_q \text{ if for some } 1 \leq i \leq n, x_p, y_1 \in M_i \\ \quad \text{and } x_p y_1 \neq e_i \text{ holds} \\ \text{(iii) } x'.y' \text{ where } x' = x_1, x_2 \ldots, x_{p-1} \text{ and } y' = y_2, \ldots, y_q \text{ if } x_p, y_1 \in M_i \\ \quad \text{for some } 1 \leq i \leq n \text{ and } x_p y_1 = e_i \end{cases}$$

When no confusion may arise we will drop the dot between x and y. Also, the sequence $x = x_1, x_2, ..., x_p$ will simply be denoted by $x_1 x_2 \cdots x_p$ and referred to as the *canonical factorization* of x and $x_1, x_2, ..., x_p$ as the *canonical factors*.

Let us apply this notion to trace monoids. Consider the graph $\#I$ of the relation I whose set of vertices is Σ and whose edges connect a and b for all $(a,b) \in I$. Suppose that for some decomposition $\Sigma = \bigcup_{1 \leq i \leq n} \Sigma_i$, the graph $\#I$ is the union of its restrictions to the subalphabets Σ_i:

$$\#I = \bigcup_{1 \leq i \leq n} \#I_i \quad \text{where } I_i = I \cap \Sigma_i \times \Sigma_i$$

Then $M(\Sigma, I)$ is the free product of the submonoids generated by the subalphabets, i.e.:

$$M(\Sigma, I) = M(\Sigma_1, I_1) * \cdots * M(\Sigma_n, I_n) \tag{2}$$

Indeed, set $R_i = \{(ab, ba) \in \Sigma_i \times \Sigma_i \mid (a,b) \in I_i\}$ for every $i = 1, \ldots, n$. Then $M(\Sigma_i, I_i)$ is presented by $< \Sigma_i; R_i >$ for every $i = 1, \ldots, n$ and $M(\Sigma, I)$ is presented by $< \bigcup_{1 \leq i \leq n} \Sigma_i; \bigcup_{1 \leq i \leq n} R_i >$. In particular, every trace in $t \in M(\Sigma, I)$ has a canonical factorization $t = t_1 t_2 \ldots t_p$. Let $G = (V, E, \lambda)$ be the dependence graph of t. Then the subgraphs $G_1 = (V_1, E_1, \lambda_1), \ldots, G_p = (V_p, E_p, \lambda_p)$ associated with the canonical factors are called the *canonical subgraphs*. The dependence graph G of t is thus the product of the dependence graphs $G_1, G_2, ..., G_p$ (the composition in the sense of [**?**], Definition 1.11), which are thus linearly ordered. In particular, for all $1 \leq i < j \leq p$, there exists a path from each vertex x of V_i to each vertex y of V_j. By definition of the relation $\prec$, this implies in particular $x \prec y$.

We are interested in the case of trace monoids $M(\Sigma, I)$ where the relation I is transitive. Let $\Sigma_1, \cdots, \Sigma_n$ be the maximal subalphabets consisting of letters that commute pairwise: for all $i = 1, \ldots, n$, and all $a \neq b \in \Sigma_i$, $(a,b) \in I$ holds. Then Σ_i, $i = 1, \ldots, n$, generates a submonoid that is isomorphic to the free commutative monoid $\Sigma_i^{\oplus}$. Thus the monoid $M(\Sigma, I)$ is isomorphic to $\Sigma_1^{\oplus} * \cdots * \Sigma_n^{\oplus}$. Observe, as a special case, that the free monoid Σ^* is the free product of the free commutative submonoids generated by each of the n letters: $\Sigma^* = a_1^* * \cdots * a_n^*$.

Example 1.— With the commutation relation of Example 2. in section **??**, the trace $t = abadccbabcdd$ is factored into $t = t_1 t_2 t_3 t_4$ where $t_1 = a^2 b \in \{a,b\}^{\oplus}$, $t_2 = c^2 d \in \{c,d\}^{\oplus}$, $t_3 = ab^2 \in \{a,b\}^{\oplus}$ and $t_4 = cd^2 \in \{c,d\}^{\oplus}$.

Example 2.— With the hypotheses of the previous example, consider $t = abaabcdcdd$. Then $t = t_1 t_2$, where $t_1 = a^3 b^2$, $t_2 = c^2 d^3$. Furthermore, $V = \{i \mid 1 \leq i \leq 10\}$, $V_1 = \{i \mid 1 \leq i \leq 5\}$ and $V_2 = \{i \mid 6 \leq i \leq 10\}$, see Fig. 3.

Fig. 3 A trace having two canonical factors

3.2 Associating a sentence with a rational subset: the commutative case

In this section all letters commute, i.e., the trace monoid is $M = \Sigma^{\oplus}$. Then it is an elementary result that the rational languages are *semilinear*, i.e., they are finite unions of *linear* sets, each of which is of the form (cf., e.g., [**?**]):

(3) $\quad uv_1^* v_2^* \cdots v_n^* \quad$ for some integer $n \geq 0$ and some $u, v_1, \ldots, v_n \in M$

We will freely use the ordinary abbreviations, $\forall x$, $\forall X$, $X = Y$, $X = \emptyset$, $X \cap Y$, $X \subseteq Y$, $X = \bigcup_{1 \leq i \leq k} X_i$ etc... We also introduce some other abbreviations that help reading the formulas. The first one expresses the fact that, for a fixed integer k, a subset has cardinality k:

$$|X| = k \equiv (\exists x_1 \cdots, x_k \in X \bigwedge_{1 \leq i < j \leq k} x_i \neq x_j) \wedge (\forall x_1 \cdots x_{k+1} \in X \bigvee_{1 \leq i < j \leq k+1} x_i = x_j)$$

The second says that for a fixed integer k, the set X has k times as many elements as the set Y. This is translated by saying that there exist k disjoint subsets $X_1, \cdots, X_k$ of the same cardinality as Y, the union of which equals X. Observe that in the present case the predicate Card_r can be used without any restriction to express the fact that two subsets have the same cardinality, since any two distinct elements are incomparable:

$$|X| = k|Y| \equiv \exists X_1 \cdots \exists X_k (X = \bigcup_{1 \leq i \leq k} X_i) \wedge (\bigwedge_{1 \leq i < j \leq k} X_i \cap X_j = \emptyset) \wedge (\bigwedge_{1 \leq i \leq k} \mathrm{Card}_r(X_i, Y))$$

Let us verify that all rational trace languages R in the free commutative monoid can be defined by some closed formula of the language $\mathcal{RCL}$. Observe first that it clearly suffices to show that linear sets as (3) are definable. We start with the two special cases where $R = a^r$ for some $a \in \Sigma$ and $R = v^*$ for some $v \in \Sigma^{\oplus}$ and then we prove that the products of languages of this type are still definable.

If $R = a^r$ for some $a \in \Sigma$, then we have: $R = L_\phi$ where

$$\phi \equiv \exists X (|X| = r) \wedge \neg \exists X (|X| = r + 1) \wedge \forall x Q_a(x)$$

Now assume $v = a_1^{r_1} a_2^{r_2} \cdots a_p^{r_p}$ and $R = v^*$. Then $R = L_\phi$ where

$$\phi \equiv \exists X \exists X_1 \cdots \exists X_p (\forall x (x \in \bigcup_{1 \leq i \leq p} X_i)) \wedge (\bigwedge_{1 \leq i \leq p} |X_i| = r_i |X|) \wedge (\forall x \bigwedge_{1 \leq i \leq p} (x \in X_i \to Q_{a_i}(x)))$$

Now assume we have $R_1 = L_{\phi_1}$ and $R_2 = L_{\phi_2}$. In order to prove the last properties we resort to the standard method of relativization. Consider two different variables X_1 and X_2 and transform ϕ_1 into ϕ'_1 by replacing all occurrences of the quantifiers $\exists X$ and $\exists x$ by $\exists X(X \subseteq X_1)$ and $\exists x(x \in X_1)$ respectively. Formally, for all formulas ψ we define recursively the formula ψ' as follows: if $\psi = \chi \wedge \sigma$ then $\psi' = \chi' \wedge \sigma'$, if $\psi = \neg\chi$ then $\psi' = \neg\chi'$, if $\psi = \exists X\chi$ then $\psi' = \exists X(X \subseteq X_1 \wedge \chi')$ and finally if $\psi = \exists x\chi$ then $\psi' = \exists x(x \in X_1 \wedge \chi')$. Similarly, transform ϕ_2 into ϕ''_2 by replacing all occurrences of the quantifiers $\exists X$ and $\exists x$ by $\exists X(X \subseteq X_2)$ and $\exists x(x \in X_2)$ respectively. Then the product R_1R_2 is defined by the formula ϕ where:

$$\phi \equiv \exists X_1 \exists X_2 (X_1 \cap X_2 = \emptyset) \wedge (\forall x(x \in X_1 \cup X_2)) \wedge (\phi'_1 \wedge \phi''_2)$$

3.3 Associating a sentence with a rational subset: the general case

In order to easily express in logical terms the fact that a trace belongs to a certain rational subset we need some preparatory work consisting essentially of shorthands. Before tackling the logical aspect of the problem we first express the property of a trace to be recognized by a finite automaton in a suitable way. Though the constructions are general, we expose them in the context of free products of free commutative monoids. Let R be a rational subset of $\Sigma_1^{\oplus} * \cdots * \Sigma_n^{\oplus}$ and let $\mathcal{A}$ be a finite automaton that recognizes it. Then each $t \in R$ can be factored into canonical factors (cf. section **??**) whose occurrences belong to some subalphabet Σ_i, the occurrences of two consecutive factors belonging to two different subalphabets: $t = t_1 \cdots t_p$ where $t_k \in \Sigma_{i_k}^{\oplus}$ for $k = 1, \cdots, p$ and $i_k \neq i_{k+1}$ for $k = 1, \cdots, p-1$, cf. Examples 2 1 and 2 of section **??**.

Now in a path labelled by t in the automaton $\mathcal{A}$ we may group all the transitions that involve a specific subalphabet. This naturally leads us to extend the ordinary definition of finite automaton by allowing transitions to be defined by arbitrary rational subsets (not just letters) in a specific subalphabet. The following technical Lemma allows to logically express the fact that a trace is recognized by some finite automaton.

Lemma 3.1 *Let R be a rational subset of $M = \Sigma_1^{\oplus} * \cdots * \Sigma_n^{\oplus}$. Then there exists a finite set (of states) Q, two specific elements q_-, $q_+ \in Q$ and for each $q,q' \in Q$ and each $i = 1, \ldots, n$, a rational subset $X^i_{q,q'} \in \mathrm{Rat}\Sigma_i^{\oplus}$ such that the following holds:*

for all $t \in M - \{1\}$ and its canonical factorization $t_1 \cdots t_p$, we have $t \in R$ if and only if there exists a sequence $q_0, q_1, \ldots, q_p$ of elements such that:

$q_0 = q_-, q_p = q_+$, and for each $i = 1, \ldots, p$ there exists $1 \leq j \leq n$ such that $t_i \in X^j_{q_{i-1},q_i}$

Proof .– Let Q be the set of states of a finite automaton $\mathcal{A}$ recognizing R. As we saw in II. 2, we may assume without loss of generality that it is normalized, i.e., in particular, that it has a unique initial state q_- and a unique terminal state q_+. With all pairs of states (q, q') and all subalphabets $\Sigma_i, i = 1, \cdots, n$ we associate the rational set $X^i_{q,q'}$ of all traces in $\Sigma_i^{\oplus}$ that are the label of a path from q to q'. Clearly, if a trace t is recognized by $\mathcal{A}$ then decomposing t into canonical factors $t_1 \cdots t_p$ whose letters belong to different alphabets, we define a path:
(4)

$$\begin{cases} c: \quad q_0 \overset{t_1}{\to} q_1 \overset{t_2}{\to} q_2 \cdots \overset{t_p}{\to} q_p \text{ where } q_0 = q_-, q_p = q_+ \text{ and} \\ \qquad \text{for each } i = 1, \ldots, p \text{ there exists } j = 1, \ldots, n \text{ such that } t_i \in X^j_{q_{i-1}, q_i} \end{cases}$$

Conversely, if the condition of the lemma holds then there exists a path as in (4) which completes the verification.

With the help of the previous statement we may now enquire how to logically translate the fact that a trace t belongs to a certain rational subset. We make the hypothesis that the empty trace does not belong to the rational subset. Indeed, it is clear that a subset R is rational if and only if $R - \{1\}$ is itself rational. Since the empty trace can be expressed by $\neg\exists X(X \neq \emptyset)$, then if 1 belongs to R and if $R - \{1\} = L_\phi$, it follows that $R = L_{\phi \vee \neg\exists X(X \neq \emptyset)}$ holds. Now let $\mathcal{A}$ be an automaton accepting R and let $t = t_1 \cdots t_p$ be the factorization of a trace t into the canonical factors in the subalphabets $\Sigma_1, \cdots, \Sigma_n$. Then t is accepted by $\mathcal{A}$ if and only if for each factor t_i the following holds:

> either t_i is the unique canonical factor of t, its alphabet is Σ_j, it takes the initial state q_- to the final state q_+ and it belongs to $X^j_{q_-, q_+}$
>
> or t_i is the first factor but not the last factor, its alphabet is Σ_j, it takes the initial state to the state q and it belongs to $X^j_{q_-, q}$
>
> or t_i is the last factor but not the first factor, its alphabet is Σ_j, it takes the state q to the final state q_+ and it belongs to X^j_{q, q_+}
>
> or t_i is not the first or the last factor, its alphabet is Σ_j, it takes the state q to the state q' and it belongs to $X^j_{q,q'}$.

The informal description of a formula stating that t belongs to R makes a decisive use of the notion of canonical factor. In terms of dependence graph such a factor is characterized by the fact that two elements x, y are incomparable with respect to the relation $\prec$, and that it is maximal for this property:

$$\begin{aligned} Fact(X) \equiv \forall x \forall y \quad & (x \in X \wedge y \in X) \to (\neg(x \prec y) \wedge \neg(y \prec x)) \\ \wedge \forall Y \quad & (X \subset Y) \to \exists x \exists y \;\; (x \in Y \wedge y \in Y \wedge (x \prec y \vee y \prec x)) \end{aligned}$$

We now turn to translate the notion of transition. In the case of free monoids the state is read for each occurrence of a letter or, in terms of dependence graph, is attached to the current node. Thus "the current state is q" is translated to "the current node belongs to S_q" where the subset S_q corresponds to q. Here we have grouped all the letters belonging to a given subalphabet into factors. Therefore, "the current state is q" is translated by "the nodes of the current canonical factor are included in S_q" where the subset S_q corresponds to q.

Now we need to express the transitions between canonical factors as is done between letters in the free monoids. Since the canonical factors are linearly ordered it is most natural to try defining the successor Y of a subset X:

$$Succ(Y,X) \equiv \exists x \exists y \ \ (x \in X) \wedge (y \in Y) \wedge (x \prec y) \wedge \forall z \ \ (x \preceq z \preceq y) \rightarrow (x = z \vee y = z)$$

Dually, we have

$$Pred(Y,X) \equiv Succ(X,Y)$$

Observe that $Succ(X,Y)$ and $Succ(Y,X)$ may hold at the same time but we will use this predicate for canonical factors, in which case they are exclusive of one another.

Let us now refine further the formula assigned to the automaton $\mathcal{A}$. The previous discussion showed that we need two types of set variables. One for the states of the automaton: $S_0, S_1, \cdots, S_m$ where the states are renumbered $0, \cdots, m$, (0 stands for the initial state q_- and m stands for the final state q_+), the other one for the n subalphabets $\Sigma_i : T_1, \cdots, T_n$. Because of the previous section, after the renaming of the states, we may asssume that the formulas $\phi^k_{i,j}$ associated with the rational subsets $X^k_{i,j}$ that define the set of traces over the commutative free monoids $\Sigma_k^{\oplus}$ that take the state i to the state j, are given. In other terms: $X^k_{i,j} = L_{\phi^k_{i,j}}$. These observations lead to the following new formulation of the acceptance of a trace:

there exist subsets of nodes $S_0, S_1, \cdots, S_m$ each corresponding to some state $0, \cdots, m$ of the automaton there exist subsets of nodes $T_1, \cdots, T_n$ each corresponding to some subalphabet Σ_i for $i = 1, \cdots, n$

> such that the subsets $S_0, S_1, \cdots, S_m$ are pairwise disjoint and each T_i corresponds to the nodes labelled by letters in the alphabet Σ_i and for each canonical factor X we have
>
> > if X has no successor and predecessor then
> >
> > > for some subalphabet Σ_k, its alphabet is included in Σ_k and its current state is the final state m and it satisfies $\phi^k_{0,m}$
> >
> > or if X has no successor but has a predecessor then
> >
> > > if Y is a predecessor that is a canonical factor then for some state i and some subalphabet Σ_k , the state of Y is i, the state of X is m and the alphabet of X is included in Σ_k and X satisfies satisfies $\phi^k_{i,m}$

or if X has no predecessor but has a successor then
for some state j and some subalphabet Σ_k, the state of X is j and its alphabet is included in Σ_k and it satisfies $\phi^k_{0,j}$

or if X has a successor and a predecessor Y then
if Y is a predecessor that is a canonical factor then for some states i and j and some subalphabet Σ_k , the state of Y is i and the state of X is j and the alphabet of X is included in Σ_k and X satisfies $\phi^k_{i,j}$

In our logic this gives the final expression where for all formulas θ and all subset variables X, the formula $\theta(X)$ applies to the structure whose set of vertices is X. This is achieved as in the previous subsection by recursively replacing each occurrence of $\exists z$ and $\exists Z$, with z and Z individual and subset variables, by $\exists z(z \in X)$ and $\exists Z(Z \subseteq X)$ respectively:

$$\exists S_0 \exists S_1 \cdots \exists S_m \exists T_1 \cdots \exists T_n$$
$$\Big(\bigwedge_{0\le i<j\le m} S_i \cap S_j = \emptyset\Big) \wedge \Big(\bigwedge_{1\le i\le n} x \in T_i \leftrightarrow \big(\bigvee_{a\in\Sigma_i} Q_a(x)\big)\Big) \wedge (\forall X(\mathit{Fact}(X) \to \psi))$$

where $\psi \equiv \psi_1 \vee \psi_2 \vee \psi_3 \vee \psi_4$ with

$$\begin{aligned}\psi_1 \equiv\ & \neg\exists Y(\mathit{Pred}(Y,X) \vee \mathit{Succ}(Y,X)) \\ & \wedge \bigvee_{1\le k\le n} (X \subseteq T_k \wedge X \subseteq S_m \wedge \phi^k_{0,m}(X))\end{aligned}$$

$$\begin{aligned}\psi_2 \equiv\ & (\neg\exists Y\ \ \mathit{Succ}(Y,X)) \wedge (\exists Y\ \ \mathit{Pred}(Y,X)) \\ & \wedge\exists Y \bigvee_{\substack{0\le i\le m\\ 1\le k\le n}} (\mathit{Fact}(Y) \wedge \mathit{Pred}(Y,X)) \wedge Y \subseteq S_i \wedge X \subseteq S_m \wedge X \subseteq T_k \wedge \phi^k_{i,m}(X))\end{aligned}$$

$$\begin{aligned}\psi_3 \equiv\ & (\neg\exists Y\ \ \mathit{Pred}(Y,X)) \wedge (\exists Y\ \ \mathit{Succ}(Y,X)) \\ & \wedge \bigvee_{\substack{0\le j\le m\\ 1\le k\le n}} (X \subseteq S_j \wedge X \subseteq T_k \wedge \phi^k_{0,j}(X))\end{aligned}$$

$$\begin{aligned}\psi_4 \equiv\ & (\exists Y\ \ \mathit{Pred}(Y,X)) \wedge (\exists Y\ \ \mathit{Succ}(Y,X)) \\ & \wedge\exists Y (\bigvee_{\substack{0\le i,j\le m\\ 1\le k\le n}} (\mathit{Fact}(Y) \wedge \mathit{Pred}(Y,X) \wedge Y \subseteq S_i \wedge X \subseteq S_j \wedge X \subseteq T_k \wedge \phi^k_{i,j}(X)))\end{aligned}$$

4 From sentences to trace languages

In this section we prove, under the hypothesis that the relation I is transitive, that the set of traces satisfying a given sentence of the restricted cardinality logic is rational.

We first start with a technical, yet standard observation that simplifies the verification of the statements. In order to stay with set variables only and not to have to care about individual variables, we note that the same expressive power is obtained by introducing the monadic relation Singleton(X) on set variables and by using as atomic formulas the following:

(i) $X \subseteq Y$ (ii) Singleton(X) (iii) $X \subseteq Q_a$ (iv) $\text{Card}_r(X,Y)$ and (v) $X \prec Y$

where $X \subseteq Q_a$ stands for: $\forall x(x \in X \to Q_a(x))$ and where $X \prec Y$ stands for

$$\forall x \forall y (x \in X \wedge y \in Y) \to x \prec y$$

We also assume that the denumerable collection of set variables is numbered in some way: $X_1, X_2, \ldots, X_i, \ldots$.

Let us briefly comment on the proof technique and justify the use of the relation $\prec$ instead of $<$ as a predicate over the vertices for the model of a dependence graph. We prove this direction of the Theorem by induction on the number of connectives involved in the definition of a given formula. However, this requires to assign a truth value to all subformulas, and those are not closed in general. We thus need to find a way to extend the induction hypothesis to all formulas. Let ϕ be a formula whose free variables are $X_{i_1}, \ldots, X_{i_p}$. The idea is to consider an interpretation of each free variable X_{i_k} for $k = 1, \ldots, p$, and to mark all the nodes that belong to the subset associated with X_{i_k} in the interpretation. In other words, the free variables of the formula are encoded into the label of each node by augmenting it with as many new components as there are free variables. The k-th new component corresponding to the k-th free variable X_{i_k} in the formula, records whether or not the node is in the subset of V associated with X_{i_k}, each letter in Σ giving thus rise to 2^p different letters. The commutation relation J on the augmented alphabet inherited from the initial relation I is no longer transitive except in trivial cases. The intermediate language associated with a formula ϕ is defined on the set of traces over this new alphabet. Because J is no longer transitive, the intermediate languages are not rational in general (e.g., in the case of a single letter a, the intermediate language associated with the formula $\text{Card}_r(X,Y)$ is the set of words having equal number of occurrences of two distinct letters, say a_0 and a_1). This leads to apply a morphism that makes some more letters commute (a_0 and a_1 would commute, which yields a rational language in the free commutative monoid generated by a_0 and a_1). However, in the structural induction step dealing with the conjunction for example, one is faced with the problem of taking the morphic image of the intersection of two subsets. It is not true in general that this is the image of the intersection, except if the two subsets are saturated by the kernel congruence defined by the morphism. It just happens that considering the relation $\prec$, the intermediate languages are necessarily saturated, which would not be the case with $<$. The second notion of augmented language associated with ϕ is then obtained by applying this "commuting" morphism. In the end, and this justifies our construction, the augmented language associated with a closed formula coincides with the set of traces that satisfy this formula.

4.1 Intermediate languages

Assume the set of free variables of a formula ϕ is $X_{i_1}, X_{i_2}, \dots, X_{i_p}$ with $i_1 < \dots < i_p$. We define the *augmented* alphabet associated with ϕ as comprising all pairs consisting of a letter of the alphabet Σ and a sequence $\alpha = \alpha_1, \alpha_2, \dots, \alpha_p$ of 0's and 1's:

$$\Delta = \{(a, \alpha) \mid a \in \Sigma, \alpha_i \in \{0, 1\}, i = 1, \dots, p\}$$

Let $\pi : \Delta \to \Sigma$ be the application that selects the first component

$$\pi(a, \alpha) = a$$

Consider the commutation relation on Δ inherited from Σ

(1) $$J = \pi^{-1}(I) = \{((a, \alpha), (b, \beta)) \mid (a, b) \in I\}$$

Observe that J is no longer transitive, i.e., $M(\Delta, J)$ is not a free product of free commutative monoids when $p > 0$: it is a free product of direct products of free monoids. Indeed, for $i = 1, \dots, n$ set

(2) $$\Delta_i = \pi^{-1}(\Sigma_i) \quad J_i = J \cap (\Delta_i \times \Delta_i)$$

and for all $a \in \Sigma$ set

(3) $$\Delta_a = \pi^{-1}(a) \quad J_a = J \cap (\Delta_a \times \Delta_a)$$

Observe that no commutation occurs between the letters of Δ_a, i.e., $J_a = \emptyset$, which means that the submonoid generated by a is isomorphic to the free monoid Δ_a^*: $M(\Delta_a, J_a) \cong \Delta_a^*$. Since the subalphabets Δ_i's are disjoint and no commutation occurs between letters belonging to different subalphabets, we have, via (1) in subsection **??**

(4) $$M(\Sigma, J) = M(\Sigma_1, J_1) * \cdots * M(\Sigma_n, J_n)$$

Now for all $i = 1, \dots, n$, two subalphabets Δ_a and Δ_b, $a \neq b \in \Sigma_i$ commute globally which means that we have the direct product of free monoids

(5) $$M(\Sigma_i, J_i) = \prod_{a \in \Sigma_i} M(\Delta_a, J_a)$$

as claimed.

As a consequence, all $t \in M(\Delta, J)$ have a canonical factorization $t = t_1 \cdots t_p$ where for each $i = 1, \cdots, p$ all the nodes of t_i are labelled by the letters of the subalphabet Δ_j for some $j = 1, \cdots, n$, two consecutive factors corresponding to two different subalphabets Δ_j. Because of (1), π extends uniquely to a morphism of $M(\Delta, J)$ that we still denote by the symbol π

$$\pi : M(\Delta, J) \to M(\Sigma, I)$$

Now we interpret each set variable as a set of positions in the dependence graph.

Definition 1. — The *intermediate language* associated with ϕ is the subset $F_\phi \subseteq M(\Delta, J)$ consisting of all traces $t \in M(\Delta, J)$ such that $\pi(t)$ satisfies ϕ when each subset variable $X_{i_k}, k = 1, \cdots, p$ is interpreted as being the subset of nodes labelled by $(a, \alpha_1, \ldots, \alpha_p)$ such that $\alpha_k = 1$ holds.

Example 1.— Consider the formula $\phi \equiv X_1 \subseteq X_2$ and the commutation relation $I = \{(a, b), (b, a), (c, d), (d, c)\}$. We have $\Sigma_1 = \{a, b\}$ and $\Sigma_2 = \{c, d\}$. Then the augmented alphabet Δ possesses 3 components, one for Σ, one for X_1 and one for X_2. The dependence graph in Fig. 4 represents a trace in $M(\Delta, J)$.

Fig. 4 A trace of $M(\Delta, J)$ that belongs to F_ϕ

Furthermore this trace belongs to F_ϕ since $X_1 \subseteq X_2$ holds in the dependence graph of $\pi(t)$ (cf. Fig. 3) with X_1 and X_2 interpreted as $\{3, 5, 7, 8, 9\}$ and $\{1, 3, 5, 6, 7, 8, 9, 10\}$ respectively.

4.2 Augmented languages

The idea behind the augmented language is to make all the letters in Δ_a, for each $a \in \Sigma$, commute. We define the commutation relation K on Δ:

$$K = J \cup \{((a, \alpha), (a, \beta)) \mid \alpha \neq \beta\} \tag{6}$$

Then K is transitive, i.e., $M(\Delta, K)$ is a free product of free commutative monoids. Furthermore $J \subseteq K$ holds and therefore there exists a canonical morphism

$$\gamma : M(\Delta, J) \to M(\Delta, K) \tag{7}$$

and we denote by $\underset{\gamma}{\sim}$ the kernel congruence it generates.

Definition 2. — The *augmented language* associated with ϕ is the morphic image $L_\phi = \gamma(F_\phi)$.

Example 1.— (continued) The dependence graph obtained by letting all the letters of the augmented alphabet inside a given subalphabet, commute, is shown in Fig. 5. It belongs to L_ϕ:

Fig. 5 A trace with the commutation relation Δ

The following notion is crucial for the proof of the Theorem.

DEFINITION 3.— A subset F of $M(\Delta, J)$ is *saturated* by $\underset{\gamma}{\sim}$ if $F = \gamma^{-1}\gamma(F)$ holds, i.e., if for all $u, u' \in M(\Delta, J)$, $u \in F$ and $\gamma(u) = \gamma(u')$ implies $u' \in F$.

Observe that saturation can be expressed in terms of dependence graphs. Indeed, F is saturated if and only if for all dependence graphs in F, and all pairs of consecutive nodes v, v' with respect to the ordering $<$ on this graph, whose labels have the same first component, the dependence graph obtained by exchanging the labels of v and v' represents an element of F (e.g., Fig. 4, vertices 1 and 3 could exchange their labels).

We make a final remark before tackling the proof. Our two definitions do not really rely on the specific numbering of the variables. What follows shows that we can work "up to a renaming of the variables". Indeed, let ϕ be a formula and $X_{i_1}, \ldots, X_{i_p}$, $i_1 < \ldots < i_p$ its collection of free variables. Consider a renaming of the variables, ϕ' the formula ϕ after the renaming and Δ' the augmented alphabet of ϕ'. Let $X_{j_1}, \ldots, X_{j_p}$, $j_1 < \ldots < j_p$ be the new collection of the free variables. Denote by σ the one-to-one correspondence on the indices such that the initial free variable X_{i_k} corresponds to the new variable $X_{i_{\sigma(k)}}$ and let h be the one-to-one mapping of Δ onto Δ' that associates with the letter $(a, \alpha_1, \ldots, \alpha_p) \in \Delta$ the letter $(a, \alpha_{\sigma(1)}, \ldots, \alpha_{\sigma(p)} \in \Delta'$. Then $F_{\phi'} = h(F_\phi)$ (resp. $L_{\phi'} = h(L_\phi)$ holds where, by abuse of notation h stands for the automorphism of $M(\Sigma, J)$ (resp. $M(\Sigma, K)$ that extends uniquely the mapping of Δ onto Δ'.

4.3 Induction

We are now in a position to state the induction hypothesis precisely:

(i) the intermediate language F_ϕ is saturated by the congruence $\underset{\gamma}{\sim}$, i.e., $F_\phi = \gamma^{-1}\gamma(F_\phi) = \gamma^{-1}(L_\phi)$ holds

and

(ii) the augmented language L_ϕ is rational in $M(\Delta, K)$.

Observe that the intermediate language of each of the atomic formulas is saturated by the congruence $\underset{\gamma}{\sim}$. Indeed, the truth value of these formulas is invariant under commutation of the nodes having the same first components of the labels. Also, when F_ϕ is rational, which as we will see, is the case for the first three atomic formulas, then so is L_ϕ by homomorphism, which means that assertion (ii) will only have to be verified for the remaining cases. We now turn to the actual verification and recall that the relations J and K are those defined by equations (1) and (6) respectively. By abuse of notation, given any subset $\Delta' \subseteq \Delta$, we will denote by $M(\Delta', J)$ (respectively $M(\Delta', K)$), the submonoid of $M(\Delta, J)$ (respectively $M(\Delta', K)$), generated by Δ', which rigorously speaking should be written as $M(\Delta', J \cap \Delta' \times \Delta')$ (respectively $M(\Delta', K \cap \Delta' \times \Delta')$), .

(i) $\phi \equiv X \subseteq Y$. Up to a renaming of the variables, the formula is equivalent to $X_1 \subseteq X_2$ and the augmented alphabet

$$\Delta = \{(a, \alpha_1, \alpha_2) \mid a \in \Sigma, \alpha_1, \alpha_2 \in \{0, 1\}\}$$

has three components, one for Σ, one for X_1 and one for X_2 respectively. Set

$$\Delta' = \{(a, \alpha_1, \alpha_2) \mid a \in \Sigma \text{ and } 0 \leq \alpha_1 \leq \alpha_2 \leq 1\}$$

Then we get

$$F_\phi = M(\Delta', J) \qquad \text{and} \qquad L_\phi = \gamma(F_\phi)$$

Thus F_ϕ is rational in $M(\Delta, J)$, so L_ϕ is rational in $M(\Delta, K)$.

(ii) $\phi \equiv \text{Singleton}(X)$. The augmented alphabet Δ has two components, one for Σ and one for X respectively. We set $\Delta_0 = \{(a, 0) \mid a \in \Sigma\}$ and $\Delta_1 = \{(a, 1) \mid a \in \Sigma\}$. Then we obtain

$$F_\phi = M(\Delta_0, J)\Delta_1 M(\Delta_0, J) \qquad \text{and} \qquad L_\phi = \gamma(F_\phi) = M(\Delta_0, K)\gamma(\Delta_1)M(\Delta_0, K)$$

Thus F_ϕ is rational in $M(\Delta, J)$, so L_ϕ is rational in $M(\Delta, K)$.

(iii) $\phi \equiv X \subseteq Q_a$. Again we consider the augmented alphabet Δ with two components, one for Σ and one for X respectively and we set $\Delta' = \{(a,1)\} \cup \{(b,0) \mid b \in \Sigma\}$. Then

$$F_\phi = M(\Delta', J) \qquad \text{and} \qquad L_\phi = \gamma(F_\phi)$$

Thus F_ϕ is rational in $M(\Delta, J)$, so L_ϕ is rational in $M(\Delta, K)$.

(iv) $\phi \equiv \mathrm{Card}_r(X, Y)$. Up to a renaming of the variables, the formula is equivalent to $\mathrm{Card}_r(X_1, X_2)$ and the augmented alphabet Δ has three components, one for Σ, one for X_1 and one for X_2 respectively. For $i = 1, \cdots, n$, let F_i and L_i be the intermediate and augmented languages associated with the formula $\phi \equiv \mathrm{Card}_r(X, Y)$ interpreted in the commutative submonoid generated by Σ_i. Let $t \in M(\Delta, J)$ be a trace and let $t = t_1 \cdots t_p$ be its canonical factorization. Then t belongs to F_ϕ if and only if exactly one of its factors, say t_j, belongs to some F_i. By letting $\Delta' = \{(a,0,0) \mid a \in \Sigma\}$, we obtain

$$F_\phi = \bigcup_{1 \le i \le n} M(\Delta', J) \;\; F_i \;\; M(\Delta', J)$$

For $i = 1, \cdots, n$ set

$$\Delta'_i = \{(a,1,0) \mid a \in \Sigma_i\} \quad \Delta''_i = \{(a,0,1) \mid a \in \Sigma_i\} \text{ and } \Delta'''_i = \{(a,\alpha_1,\alpha_2) \mid a \in \Sigma_i \text{ and } \alpha_1 = \alpha_2\}$$

For all traces t, denote by $|t|_{\Delta'_i}$ and $|t|_{\Delta''_i}$ respectively, the number of occurrences of the letters of the subalphabets Δ'_i and Δ''_i in t. Then

$$F_i = M(\Delta_i, J) \cap \{t \mid |t|_{\Delta'_i} = |t|_{\Delta''_i}\}$$

Thus $L_i = \gamma(F_i) = M(\Delta'''_i, K)(\Delta'_i \Delta''_i)^*$ is rational in $M(\Delta, K)$ and therefore so is

$$L_\phi = \gamma(F_\phi) = \bigcup_{1 \le i \le n} M(\Delta', K) \;\; \gamma(F_i) \;\; M(\Delta', K)$$

(v) $\phi \equiv X \prec Y$. The formula is equivalent to $X_1 \prec X_2$ and the augmented alphabet Δ has three components, one for Σ, one for X_1 and one for X_2 respectively. For all $i = 1, \cdots, n$ we define:

$$\Delta'_i = \{(a,1,0) \mid a \in \Sigma_i\} \quad \Delta''_i = \{(a,0,1) \mid a \in \Sigma_i\} \quad \Delta_{\hat{i}} = \{(a,0,0) \mid a \notin \Sigma_i\}$$

We also set

$$\Delta_0 = \{(a,0,0) \mid a \in \Sigma\}$$

Since X_1 and X_2 must be disjoint, $F_\phi \subseteq M(\Delta', J)$ where

$$\Delta' = \{(a,\alpha_1,\alpha_2) \mid a \in \Sigma \text{ and } \alpha_1 + \alpha_2 < 2\}$$

Let us inquire about the conditions under which a trace $t \in M(\Delta', J)$ is not in

F_ϕ. It fails to be in F_ϕ if and only if the dependence graph of $\pi(t)$ satisfies one of the following conditions: (a) There exist some $x \in X$ and some $y \in Y$ that are incomparable with respect to the relation $\prec$. Then t contains a factor in

$$\Omega_1 = \bigcup_{1\le i\le n} \Delta'_i\Delta''_i$$

If the previous situation does not occur then there exists $x \in X$ and $y \in Y$ such that $y \prec x$. We may further assume that for all $y \prec z \prec x$, $z \notin X \cup Y$ holds. Then two cases must be considered:

(b) The labels of x and y are in two different subalphabets. Then t contains a factor in

$$\Omega_2 = \bigcup_{\substack{1\le i,j\le n\\ i\neq j}} \Delta''_j M(\Delta_0, J)\Delta'_i$$

(c) The labels of x and y are in the same subalphabet but the label of some intermediate node $y \prec z \prec x$ belongs to some other subalphabet. Then t contains a factor in

$$\Omega_3 = \bigcup_{1\le i\le n} \Delta''_i \; M(\Delta_0, J) \; \Delta_{\widehat{i}} \; M(\Delta_0, J) \; \Delta'_i$$

Setting:

$$\Omega = \Omega_1 \cup \Omega_2 \cup \Omega_3$$

we finally get:

$$F_\phi = M(\Delta', J) - M(\Delta', J) \;\; \Omega \;\; M(\Delta', J)$$

Now $M(\Delta', J)$ and $M(\Delta', J) \;\Omega\; M(\Delta', J)$ are saturated by the congruence $\underset{\gamma}{\sim}$ and furthermore they are rational in $M(\Delta, J)$. Then we obtain:

$$L_\phi = \gamma(F_\phi) = M(\Delta', K) - M(\Delta', K) \;\; \gamma(\Omega) \;\; M(\Delta', K)$$

i.e., L_ϕ is rational in $M(\Delta, K)$ because of Theorem **??**.

As a result, in all five cases the augmented trace languages associated with the atomic formulas are rational. We now prove that if the intermediate (resp. augmented) languages associated with ϕ, ϕ_1 and ϕ_2 are saturated by their respective congruences (resp. are rational) then so is the intermediate (resp. augmented) language associated with $\exists X\phi$, $\neg\phi$ and $\phi_1 \wedge \phi_2$.

(vi) $\exists X\phi$. We assume that the augmented alphabet Δ_p associated with ϕ has $(p + 1)$ components, one for Σ, and one for each free variable. We denote by J_p and K_p the commutation relations as defined in (1) and (6). Let F_ϕ and L_ϕ

be the intermediate and augmented languages of ϕ. Assume, without loss of generality, that $X = X_p$, i.e., X corresponds to the last component. Let Δ_{p-1} be the alphabet obtained from Δ_p by ignoring the last component:

$$\Delta_{p-1} = \{(a, \alpha_1, \cdots, \alpha_{p-1}) \mid (a, \alpha_1, \cdots, \alpha_p) \in \Delta_p \text{ for some } \alpha_p \in \{0, 1\}\}$$

Denote by χ the application of Δ_p onto Δ_{p-1} that maps $(a, \alpha_1, \cdots, \alpha_p)$ to $(a, \alpha_1, \cdots, \alpha_{p-1})$ and denote by J_{p-1} and K_{p-1} the commutation relations on Δ_{p-1} obtained by applying (1) and (6). Then χ naturally extends to a morphism:

$$\chi : M(\Delta_p, J_p) \to M(\Delta_{p-1}, J_{p-1})$$

and we have $F_{\exists X\phi} = \chi(F_\phi)$. Let us verify that the subset $F_{\exists X\phi}$ is saturated by the congruence γ_{p-1}. Indeed, following the observation following Definition 3, consider $u \in F_{\exists X\phi}$, and, in the dependence graph of u, two nodes v, v' that are consecutive with respect to the ordering $<$ and whose labels have the same first component. Consider a pre-image $t \in F_\phi$ of u (i.e., $u = \chi(t)$) and let $t' \in M(\Delta_p, J_p)$ be the trace obtained by exchanging the labels of v and v' in the dependence graph of t. Then $t' \in F_\phi$, by saturation of F_ϕ. Thus, $v' = \chi(t') \in F_{\exists X\phi}$ as claimed.

Consider the projection:

$$\chi' : M(\Delta_p, K_p) \to M(\Delta_{p-1}, K_{p-1})$$

and the morphisms

$$\gamma_p : M(\Delta_p, J_p) \to M(\Delta_p, K_p) \qquad \gamma_{p-1} : M(\Delta_{p-1}, J_{p-1}) \to M(\Delta_{p-1}, K_{p-1})$$

Then we have

$$\chi' \circ \gamma_p = \gamma_{p-1} \circ \chi$$

Indeed, since $M(\Delta_p, J_p)$ is the quotient of the free monoid Δ_p^* by the congruence generated by the relation J_p, it suffices to show that the images of the two-letter traces of the form $(a, \alpha_1, \ldots, \alpha_p)(b, \beta_1, \ldots, \beta_p)$ are equal in the two mappings. If $(a, b) \in I$ then these images are the commutation class of the element of $M(\Delta_{p-1}, K_{p-1})$ otherwise it is the element $(a, \alpha_1, \ldots, \alpha_{p-1})(b, \beta_1, \ldots, \beta_{p-1})$ of $M(\Delta_{p-1}, K_{p-1})$. Because of

$$L_{\exists X\phi} = \gamma_{p-1}(F_{\exists X\phi}) \text{ and } \gamma_p(F_\phi) = L_\phi$$

we get

$$\begin{aligned} L_{\exists X\phi} &= \gamma_{p-1}(\chi(F_\phi)) \\ &= \chi'(\gamma_p(F_\phi)) \\ &= \chi'(L_\phi) \end{aligned}$$

The closure property of rational subsets under homomorphic image (Proposition **??**) completes the verification in this case.

(vii) $\neg\phi$. Clearly, if Δ is the augmented alphabet then $F_{\neg\phi} = M(\Delta, J) - F_\phi$ is saturated by $\underset{\gamma}{\sim}$. Then we have: $L_{\neg\phi} = \gamma(F_{\neg\phi}) = M(\Delta, K) - \gamma(F_\phi) = M(\Delta, K) - L_\phi$ which completes the verification by Theorem **??**.

(viii) $\phi_1 \wedge \phi_2$. Without loss of generality we may assume that the collection of subset variables appearing in either ϕ_1 or ϕ_2 is $X_1, \ldots, X_p$ and denote by Δ the augmented alphabet with $p+1$ components and by J, K the corresponding commutation relations according to (1) and (6). Let $X_{\sigma(1)}, \ldots, X_{\sigma(r_1)}$ where σ is an increasing injection of $\{1, \ldots, r_1\}$ into $\{1, \ldots, p\}$, be the subset variables appearing in ϕ_1. Similarly, let $X_{\tau(1)}, \ldots, X_{\tau(r_2)}$ where τ is an increasing injection of $\{1, \ldots, r_2\}$ into $\{1, \ldots, p\}$, be the subset variables appearing in ϕ_2. We denote by Δ_1 and Δ_2 respectively, the augmented alphabets of the formulas ϕ_1 and ϕ_2 and by J_1, K_1 and J_2, K_2 the corresponding commutation relations as in (1) and (6). Consider the substitution h_1 of $M(\Delta_1, J_1)$ into $M(\Delta, J)$ that associates with every element $(a, \alpha_1, \ldots, \alpha_{r_1}) \in \Delta_1$ the set of elements:

$$\{(a, \alpha'_1, \ldots, \alpha'_p) \mid \alpha'_j \in \{0,1\} \text{ for } j = 1, \ldots, p \text{ and } \alpha'_{\sigma(i)} = \alpha_i \text{ for } i = 1, \ldots, r_1\}$$

Similarly, consider the substitution h_2 of $M(\Delta_2, J_2)$ into $M(\Delta, J)$ that associates with every element $(a, \alpha_1, \ldots, \alpha_{r_2}) \in \Delta_2$ the set of elements:

$$\{(a, \alpha'_1, \ldots, \alpha'_p) \mid \alpha'_j \in \{0,1\} \text{ for } j = 1, \ldots, p \text{ and } \alpha'_{\tau(i)} = \alpha_i \text{ for } i = 1, \ldots, r_2\}$$

Then we have

$$F_{\phi_1 \wedge \phi_2} = h_1(F_{\phi_1}) \cap h_2(F_{\phi_2})$$

Since F_{ϕ_1} and F_{ϕ_2} are saturated by $\underset{\gamma_1}{\sim}$ and $\underset{\gamma_2}{\sim}$ respectively, $F_{\phi_1 \wedge \phi_2}$ is saturated by the congruence $\underset{\gamma}{\sim}$ and furthermore

$$\begin{aligned} L_{\phi_1 \wedge \phi_2} &= \gamma(F_{\phi_1 \wedge \phi_2}) \\ &= \gamma(h_1(F_{\phi_1}) \cap h_2(F_{\phi_2})) \\ &= \gamma(h_1(F_{\phi_1})) \cap \gamma(h_2(F_{\phi_2})) \end{aligned}$$

Define the morphisms onto the corresponding monoids

$$\gamma_1 : M(\Delta_1, J_1) \to M(\Delta_1, K_1) \quad \gamma_2 : M(\Delta_2, J_2) \to M(\Delta_2, K_2)$$

and let $h'_1 : M(\Delta_1, K_1) \to M(\Delta, K)$ and $h'_2 : M(\Delta_2, K_2) \to M(\Delta, K)$ be the finite substitutions such that $h'_1 \circ \gamma_1 = \gamma \circ h_1$ and $h'_2 \circ \gamma_2 = \gamma \circ h_2$ holds. Then we obtain

$$\begin{aligned} L_{\phi_1 \wedge \phi_2} &= h'_1(\gamma_1(F_{\phi_1})) \cap h'_2(\gamma_2(F_{\phi_2})) \\ &= h'_1(L_{\phi_1}) \cap h'_2(L_{\phi_2}) \end{aligned}$$

which by closure of the rational sets under intersection (Theorem **??**) and finite substitution (Proposition **??**) completes the verification in this final case.